R

R&D Races

Jennifer F. Reinganum

Abstract

This article summarizes recent theoretical and empirical research on R&D races. Two canonical models of an R&D race are described, and their implications for the investment behaviour of incumbent leaders and potential entrants are discussed.

Empirical studies that attempt to verify or refute the implied patterns of investment are also discussed.

Keywords

R&D races; Innovation; Invention; Patents

JEL Classifications

O31

Models of R&D races, in which the winning firm receives all (or most) of the reward, rely on two basic paradigms. In a deterministic race (Barzel 1968), invention requires a known investment; the value of the patent grows over time and a potential inventor decides when to invest. Equivalently, the investment required to invent by time T is an increasing function $C(T)$. Let Πe^{-rT} denote the present value of a patent on an invention completed at T (assume that Π is also the social value). Among non-cooperative firms, the potential for pre-emption causes the date of invention to advance until $\Pi e^{-rT^0} - C\left(T^0\right) = 0$. Since T^0 precedes the socially optimal invention time, racing results in over-investment in R&D. This model assumes that only the winning firm actually invests its 'bid'.

The first fully game-theoretic model of a stochastic race is presented by Loury (1979); for a decision-theoretic antecedent, see Kamien and Schwartz (1974). Firm i's lump-sum investment x_i yields a random invention date τ_i with Pr $\{\tau_i \le t\} = 1 - \exp\{-h(x_i)t\}$; thus, firm i's hazard rate is $h(x_i)$. The firm is also uncertain about the success date of its rival, which invests simultaneously. If firm I succeeds at t, the chance that it is the first inventor is $1 - \exp\{-h(x_j)t\}$. Firm i's expected profit is $\Pi h(x_i)/[r + h(x_j) + h(x_i)] - x_i$. Lee and Wilde (1980) re−specified investment as a *rate* per unit of time. The resulting payoff for firm i is $[\Pi h(x_i) - x_i]/[r + h(x_j) + h(x_i)]$. In both versions, racing leads to over-investment. Malueg and Tsutsui (1997) incorporate uncertainty about the hazard rate; as time passes without success, firms become increasingly pessimistic and reduce their rates of investment.

Weeds (2002) combines aspects of the Barzel and the Loury models with the theory of real options. Two firms monitor the stochastic growth of the patent value and decide when to invest. The required investment is exogenous and lump-sum,

Macmillan Publishers Ltd (ed.), *The New Palgrave Dictionary of Economics*,
https://doi.org/10.1057/978-1-349-95189-5

and in exchange the firm receives an exponentially distributed time until success. Firms do not invest until certain threshold values of the patent are reached. In one type of equilibrium, a firm invests only because it fears pre-emption by its rival; the rival invests strictly later, and firms' profits are equal (and low). In the other type, firms engage in mutual forbearance, and invest only at the optimal joint-investment time. Profits are equal (and high); this equilibrium involves strategic *delay* in investment relative to the social optimum.

Action-Reaction or Increasing Dominance Within a Race

Although the exponential specification is gratifyingly simple, conditional on no success to date, the race looks exactly as it did at the beginning. If a firm could accumulate a 'lead', would a firm that is 'ahead' invest more than its rival? If so, then the race exhibits 'increasing dominance'; if not, then the lagging firm tends to catch up and the race exhibits 'action-reaction' (Vickers 1986). Of course, a lagging firm can invest more than the leader and still be less likely to win the race.

Harris and Vickers (1985) provide a deterministic racing model in which two firms alternate in making investments until one reaches the finish line. If one firm is sufficiently close to winning (in a region called its 'safety zone'), the other drops out and the leader proceeds unchallenged. In the limit (as the rate of alternation goes to infinity), the winning firm invests just enough to reach its safety zone, whereupon its rival drops out and the firm continues, at a more leisurely pace, to the finish line. If the firms are otherwise symmetric, the firm that begins the race closer to the finish line will win.

Doraszelski (2003) provides a stochastic model wherein a firm's hazard rate depends on its accumulated stock of knowledge; its investment rate can vary with its own progress and that of its rival. Firm i's hazard rate is $h^i = \lambda x_i + \gamma(z_i)^{\psi}$, where λ , γ , ψ are positive constants, x_i is firm i's rate of investment and z_i is firm i's accumulated stock of knowledge. Since x_i and z_i are substitutes in the hazard rate, a firm's equilibrium investment rate is decreasing in its knowledge stock, and the leading firm invests less than its rival; in this sense, the model exhibits action-reaction. Nevertheless, if one of the firms begins with a larger stock of knowledge, it remains *ex ante* more likely to win the race.

Progress can also be modelled by assuming that multiple stages must be completed in order to invent. Grossman and Shapiro (1987) provide a two-firm two-stage model; each stage involves a race of the Lee and Wilde form (see also Harris and Vickers, 1987, for a somewhat different model). They find that the leader invests more than the follower, but both increase their investments should the follower catch up. Thus, this literature suggests that a firm that is 'ahead' in the race is more likely to win.

Action-Reaction or Increasing Dominance Across Multiple Races

With a sequence of innovations, we ask whether a firm that is 'ahead' in the *market* (the one with a larger market share) would invest more in the next race than its rival. That is, will the industry's evolution exhibit increasing dominance (persistence of the market leader) or action-reaction (turnover of the market leader)? Consider an incumbent monopolist, with current flow profits of R, racing with a potential entrant for a cost-reducing invention. If the incumbent wins, it receives the present value of monopoly profits with the new technology, Π. If the entrant wins, the two firms compete in Cournot–Nash fashion, receiving the present value of profits π_I and π_E for the incumbent and entrant, respectively. A drastic invention is one for which $\pi_I = 0$ and $\pi_E = \Pi$ (that is, the incumbent can no longer compete when the entrant invents); if the innovation is non-drastic, then $\Pi > \pi_I + \pi_{E.}$

Gilbert and Newbery (1982) show that an incumbent monopolist would bid strictly more than a potential entrant for a non-drastic invention (and an equal amount for a drastic one); this is referred to as the 'efficiency effect'. Suppose an entrant would invent at time T^E, where $C\left(T^E\right) = \pi_E e^{-rT^E}$. The incumbent can permit this and receive $(R/r)\left(1-e^{-r^{T^E}}\right)+\pi I e^{-r^{T^E}}$, or it can pre-empt by bidding just over $C(T^E)$, and receive essentially $(R/r)\left(1-e^{-r^{T^E}}\right)+\Pi e^{-r^{T^E}}-C\left(T^E\right)$. Pre-emption is strictly preferred if $\Pi > \pi_I + \pi_E$; that is, if the invention is non-drastic. Vickers (1986) analyses a sequence of such races and finds that a sufficient condition for increasing dominance (action-reaction) is that each invention is sufficiently drastic (incremental).

Reinganum (1983) adapts the Lee and Wilde model to this scenario; the firms' pay-offs are now $[\Pi \mathrm{h}(x_I) + \pi_I \ \mathrm{h}(x_E) + R - x_I]/[r + h(x_E) + h(x_I)]$ for the incumbent and $[\pi_E h(x_E) - x_E]/[r + h(x_E) + h(x_I)]$ for the entrant. In equilibrium, the incumbent invests at a higher rate than the entrant, at least for innovations that are sufficiently drastic. This is a consequence of Arrow's (1962) 'replacement effect' (originally identified for a single inventor): the incumbent has a lower incentive to advance the invention date (and replace himself) than does the entrant. Reinganum (1985) extends this result to a sequence of drastic inventions. Thus, the stochastic model suggests action-reaction, at least for sufficiently drastic inventions; however, increasing dominance can occur for sufficiently incremental inventions.

Empirical Studies

Apparently, both models admit both investment patterns, but for opposite types (drastic versus incremental) of inventions. Since the nature of the inventive process (deterministic or stochastic) and that of the invention itself (drastic or incremental) are likely to vary across inventions and industries, the empirical pattern is also likely to vary. For a single race, the deterministic model has stark empirical implications: no real racing occurs beyond a possible initial burst. While the stochastic model implies investment by multiple firms, it is difficult to determine whether strategic effects play a significant role. The key indicator of strategic behaviour – the best response function – is an out-of-equilibrium phenomenon. Including rival investment in a regression explaining own investment should add no further information beyond that provided by the other explanatory variables. Using program-level data on pharmaceuticals R&D, Cockburn and Henderson (1994) detect no evidence of racing, but do find significant spillovers in output, calling into question the 'winner-take-all' assumption. They conclude that investment is driven by heterogeneous firm ability, adjustment costs, and technological opportunity. In a laboratory experiment based on Harris and Vickers's (1987) multi-stage model, Zizzo (2002) finds that (contrary to predictions) the investments of leaders and followers are not significantly different.

Lerner (1997) considers a sequence of races in the disk drive industry, with each invention serving to increase storage capacity. Firms that follow the

technological leaders are most likely to introduce improved drives and to make the greatest technological progress. His data also show that leaders in a given year had about a 40 per cent chance of remaining leaders the next year. Czarnitzki and Kraft (2004) employ data from a survey of German manufacturing firms, which also asks firms to state their motives in conducting R&D; thus they distinguish between (self-designated) potential entrants and incumbents. R&D expenditures per dollar of sales are significantly higher for potential entrants (but the sales normalization confounds effects). The results of the latter two studies seem consistent with the stochastic model for relatively drastic inventions, but also with the deterministic model for relatively incremental inventions.

See Also

▶ Patents

Bibliography

Arrow, K. 1962. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*. Princeton: Princeton University Press for the NBER.

Barzel, Y. 1968. Optimal timing of innovations. *Review of Economics and Statistics* 50: 348–355.

Czarnitzki, D., and K. Kraft. 2004. An empirical test of the asymmetric models on innovative activity: Who invests

more into R&D, the incumbent or the challenger? *Journal of Economic Behavior and Organization* 54: 153–173.
Cockburn, I., and R. Henderson. 1994. Racing to invest? The dynamics of competition in ethical drug discovery. *Journal of Economics and Management Strategy* 3: 481–520.
Doraszelski, U. 2003. An R&D race with knowledge accumulation. *RAND Journal of Economics* 34: 24–42.
Grossman, G., and C. Shapiro. 1987. Dynamic R&D competition. *Economic Journal* 97: 372–387.
Gilbert, R., and D. Newbery. 1982. Pre-emptive patenting and the persistence of monopoly. *American Economic Review* 72: 514–526.
Harris, C., and J. Vickers. 1985. Perfect equilibrium in a model of a race. *Review of Economic Studies* 52: 193–209.
Harris, C., and J. Vickers. 1987. Racing with uncertainty. *Review of Economic Studies* 54: 1–21.
Kamien, M., and N. Schwartz. 1974. Patent life and R&D rivalry. *American Economic Review* 64: 183–187.
Lee, T., and L. Wilde. 1980. Market structure and innovation: A reformulation. *Quarterly Journal of Economics* 94: 429–436.
Lerner, J. 1997. An empirical exploration of a technology race. *RAND Journal of Economics* 28: 228–247.
Loury, G. 1979. Market structure and innovation. *Quarterly Journal of Economics* 93: 395–436.
Malueg, D., and S. Tsutsui. 1997. Dynamic R&D competition with learning. *RAND Journal of Economics* 28: 751–772.
Reinganum, J. 1983. Uncertain innovation and the persistence of monopoly. *American Economic Review* 73: 741–748.
Reinganum, J. 1985. Innovation and industry evolution. *Quarterly Journal of Economics* 50: 81–99.
Vickers, J. 1986. The evolution of market structure when there is a sequence of innovations. *Journal of Industrial Economics* 35: 1–12.
Weeds, H. 2002. Strategic delay in a real options model of R&D competition. *Review of Economic Studies* 69: 729–747.
Zizzo, D. 2002. Racing with uncertainty: A patent race experiment. *International Journal of Industrial Organization* 20: 877–902.

Race and Economics

H. Stanback

The concept of race enters formal economic theory through a range of areas primarily within labour economics. These include discrimination, inequality, human capital, labour market competition and segmentation, and class relations. The first substantive attention by neoclassical theory to the economic problems posed by race began with the work of Gary Becker in 1957 which approaches the subject from the standpoint of discrimination. (Race is addressed from a structural standpoint by Marx in *Capital*, Volume I, through his analysis of the impact of slavery and the slave trade on the working class of the United States.)

Subsequent to Becker there have been numerous theoretical advances and approaches towards an understanding of the role of race in the economy. There are three alternative formulations of the problem: (1) employer or employee discrimination; (2) labour supply; (3) competition between capitals and between capital and labour.

Employer/Employee Discrimination

I. Becker presents race as a problem of ‘taste’ for discrimination or a ‘distaste’ for physical association with a particular race (his formulation would be equally applicable to any standard physical attribute such as sex). The taste for discrimination can come from the employer or employee. Becker’s employer distaste model assumes two societies; B, which is relatively labour-abundant, and W, which is relatively capital-abundant. These two societies engage in voluntary trade with each other but with the capital of W having a distaste for working in physical proximity to B labour. Given the assumptions of a pure theory of international trade and without such distaste, B and W exchange their respective relatively abundant quantities until the marginal product of each factor is equal in both societies. However, because of the distaste or subjective preference of W employers, their utility function must be augmented to include the number of B workers and $dU_w < dL_B < 0$. The capital exported by W must receive a money return greater than the return on capital domestically employed. The differences between the return on domestic vs. exported capital is the return or compensation W employers feel they must have for being physically close to B labour. Since such compensation must be positive, discrimination reduces the quantity of W capital exported and the quantity of B labour exported.

Becker's model concludes that as a result of discrimination, W labour's money income increases and W capital's net income falls because with the reduction in W capitals exported and B labour imported, W labour works with more capital and W capital works with less labour.

> Since $f_{LL} < 0$ and $f_{KL} < 0$, the f_L, the wage of labor, rises under the described conditions. Since $f_{KL} < 0$, the f_{KW}, the money return to domestic capital, falls as more capital is employed. (Becker 1957)

In the absence of monopoly, discrimination would end if one employer did not have or did not exercise his distaste. This employer would reap abnormal profits thus forcing other competitors to follow, assuming the drive for profits is stronger than racial distaste.

II. A more advanced version of the employer discrimination model are 'statistical discrimination models' (Reich 1981). Such models make discrimination by employers more 'rational' in their employee hirings than Becker's subjective preference criteria. Racial discrimination results from problems associated with the personnel costs of hiring, training and identifying productivity. Such costs give rise to discrimination in the normal pursuit of profits; consequently, discrimination persists.

Personnel costs affect racial employment in the following way. A Race A employer with few or no workers of Race B may want to employ some. However, the cost of hiring and training new workers would not be offset by Race B's lower wages. The costs are profit maximization considerations, therefore, racial inequality persists as an integral part of competition. Although marginal changes are made, the tendency is for no major overhaul in the racial composition of any employer's work force.

Prejudiced perceptions of Race B workers' productivity influence Race A employer's queueing or prioritizing workers for employment. Since it is costly to determine productivity prior to employment, employers of Race A presume that all workers of race B are less productive than Race A workers. This presumably protects the employer because it is much more costly to hire an inefficient worker than to pass over a productive one. As a result, Race B workers are not hired or hired at a lower wage rate. Prejudiced perceptions may emerge from a variety of sources. Regardless, in this model, they provide a low cost screen in the employee search process.

III. Employee discrimination models are based on perceived and competing economic and racial interests. Such models require the ability of workers of Race A to obtain cooperation from employers and other actors in Race A to form a 'cartel' to discriminate against workers of Race B. The following characteristics describe this cartel arrangement (see Krueger 1963; Bergmann 1971):

1. Race A capital and labour combine to discriminate against Race B's labour even though Race A labour gains while Race A capital loses (as per Becker).
2. Racial income differences can be accounted for through Race A's political control which limits inputs into Race B's schooling and, thus, skills.
3. Further discrimination against Race B takes place through the cartel arrangement by restricting hiring, occupational mobility, wage payments, access to capital markets, and through price discrimination.
4. Race B's labour is 'crowded' into certain lower paying occupations through racial discrimination. This results in a depressed marginal product because excess labour is employed and, hence, wage rates are depressed.

The results reflected in the combination/cartel arrangements require critical institutional mechanisms to assure enforcement of the 'rules' of the cartel.

Labour Supply

Race is a component of labour supply analysis primarily in relationship to human capital theory. Differences in the quantity/quality of human capital explains racial income differences. While the subjective demand for human capital by Race B may be the same as Race A's subjective demand for human capital, the objective capacity to invest is less due to lower initial endowments which may result from discrimination. Additionally the supply of human capital, e.g., education and health care, is likely to be less (Sowell 1975).

Labour supply analyses of racial inequality tend to locate the discrimination problem outside of the economic area, generally focusing on education as the critical form of human capital. Contemporary analyses of racial inequality have raised the question of the inequality of demand for human capital in the form of education (Sowell 1975). Accordingly, an emerging 'underclass', disproportionately of a particular race, has a low demand for education as a result of cultural variables which reject work at the prevailing wage rate. The important cultural variables can emerge from geographic dislocation such as migration from a rural to an urban setting, or from previous experience in the labour market.

Competition

All of the previously discussed approaches to race involve competition theories. Marxian economists have also approached race from the standpoint of either structural conflict between capital and labour, or between capitalists, or both. The Marxian concepts of accumulation and class struggle provide the foundation of this approach to an analysis of race. Race is examined not only to understand racial inequality as in previously discussed formulations, but also to understand competition between firms and capital labour conflict (Baron 1975; Harris 1972; Reich 1981; Sysmanski 1975).

Racial inequality, in this framework, is the product of the pursuit of profits. Particularly, the pursuit of cheap resource markets gives economic rationality to the use of race as a means of cheapening the cost of Race B's labour. Institutional arrangements as suggested in the 'cartel' approach, and/or direct discrimination as suggested in the employer discrimination approach, may be the explicit manifestation of the pursuit of profits.

Racial inequality may also be a product of capital–labour conflict due to deliberate manipulation of Race B's access to employment by either capital or labour from Race A. In such instances, race is utilized to strengthen or weaken a particular side of the conflict. Capital may utilize cheaper labour from Race B to reduce labour's bargaining strength. Race A labour may join with Race B labour to prevent such tactics and, thereby, strengthen Race A labour's posture.

Race, in this framework, both influences and is influenced by economic processes. Often in this framework, racial inequality and the manipulation of race in competition/conflict is considered endemic to the competitive profit pursuit.

In the past 30 years there has been a dramatic increase in the attention of economic theory to race and race related issues. As indicated above, these approaches either focus on market imperfections or on the structural character of capitalism as explanations of racial discrimination and/or inequality. Most structuralist analyses also utilized race as a means of explaining the dynamics of economic processes.

Bibliography

Baron, H. 1975. *The demand for Black labor*. Boston: New England.

Becker, G.S. 1957. *The economics of discrimination*, 2nd edn. Chicago: University of Chicago Press, 1971.

Bergmann, B. 1971. The effect on white incomes of discrimination in employment. *Journal of Political Economy* 71(2): 294–313.

Harris, D.J. 1972. The black ghetto as colony: A theoretical critique and alternative formulation. *Review of Black Political Economy* 2(4): 3–33.

Krueger, A. 1963. The economics of discrimination. *Journal of Political Economy* 71: 481–486.

Reich, M. 1981. *Racial inequality: A political economic analysis*. Princeton: Princeton University Press.

Sowell, T. 1975. *Race and economics*. New York: David McKay.
Sysmanski, A. 1975. Trends in economic discrimination against blacks in the U.S. working class. *Review of Radical Political Economics* 7(3): 1–21.

Racial Profiling

Petra E. Todd

Abstract
This article reviews the recent theoretical and empirical literature in economics that aims to establish empirically whether police engage in racially biased law enforcement practices. It considers different objective functions that might be posited for police officers and the tests that can be derived under these objectives. Assuming a hit rate objective function leads to a simple, empirically implementable outcomes-based test that can potentially explain an observed empirical regularity in many police data-sets whereby disparities in hit rates tend to be very small despite large disparities in search rates.

Keywords
Deterrence; Omitted-variable bias; Optimal auditing; Racial profiling; Rational expectations; Statistical discrimination

JEL Classifications
J7; K4

In recent years, numerous lawsuits have been brought against US city police departments alleging racially biased law enforcement practices. (Many of these lawsuits were initiated by the American Civil Liberties Union or the US Department of Justice.) As a result of closer scrutiny, many police departments now routinely collect administrative data on the characteristics of the individuals that they subject to stops and searches and on the outcomes of these encounters. It is a common finding in these data that African–Americans and Hispanic drivers are searched at a higher rate than white drivers. For example, African Americans represented 63 per cent of motorists searched by Maryland state police on the I-95 highway between 1995 and 1999, but only 18 per cent of motorists on the road. A refined version of this benchmark test for discrimination estimates the probability of being searched as a function of race and other observable characteristics thought to be related to criminal propensity. If race has explanatory power in the regression, then this is taken as evidence of discrimination (see, for example, Donohue 1999).

One drawback of benchmark tests is that they require data on the full set of characteristics that a police officer uses in deciding whether to search a motorist. (A training manual issued by the Illinois State Police highlights some indicators of criminal activity, such as tinted windows and leased vehicles.) If some characteristics are missing, then race could have explanatory power due to omitted-variable bias. Alternatively, if race is found to be insignificant, it is still possible that police target individuals with certain characteristics, because of their correlation with race and not because of their use in predicting criminality. Additionally, benchmark tests can reveal only whether a disparity exists and not the motivation for the disparity. In many racial profiling investigations, it is clear that a disparity exists and the key concern is whether the higher rates of stop and search among certain groups can be justified as an optimal monitoring response to higher rates of criminality. The judicial standpoint on racial profiling is not clear-cut. The dominant view is that race or ethnicity can be used as a factor in determining the likelihood that a person has committed a crime, so long as its use relates to law enforcement and is not a pretext for racial harassment. However, a significant dissenting view argues that race should not be used as a criterion, except in very limited cases, as when the race of a perpetrator of a particular crime is known; see Kennedy 1997. For detailed discussions of the legality of racial profiling practices, see, for example, Kennedy (1997), Harcourt (2004), Gross and Barnes (2002) and Persico and Castleman 2005.

This article reviews the recent theoretical and empirical literature in economics that aims to empirically establish from data on stops and searches of motor vehicles whether police behaviour is indicative of racial bias. The early literature on crime (for example, Becker 1968; Stigler 1970) examined citizens' incentives to misbehave under an exogenous probability of being monitored. The more recent literature assumes that police and citizens behave strategically, with police deciding on optimal search strategies and citizens deciding whether to break the law, given police search strategies. Unlike the criminology literature, where it is sometimes assumed that police can make citizens believe that the monitoring probabilities are higher than in reality (for example, Sherman 1990), the economics literature typically assumes rational expectations. (The recent economic literature is also related to the literature on optimal auditing, which mainly deals with income reporting and tax evasion; see Reinganum and Wilde 1986; Border and Sobel 1987; Scotchmer 1987.) An advance in the literature is a better understanding of the assumptions on police and motorist behaviour required to justify alternative tests for discrimination.

We next describe the frameworks that have been developed in the recent economic literature and the tests that have been derived from using these frameworks. Subsequently, we present a brief summary of some of the empirical evidence from police stop/search data-sets.

Theoretical Models of Police–Motorist Behaviour

Two leading paradigms are put forward in the economic literature. One assumes that police officers operate in a decentralized way, allocating their search activities so as to catch as many criminals as possible. In the context of motor vehicle searches, the goal is to maximize the number of successful searches given a cost of search, where a successful search is defined as one that uncovers some contraband. As noted in Persico (2002), an objective function that maximizes successful searches, or so-called hit rates, will in general not minimize the aggregate crime rate, because it does not give enough weight to the deterrent effects of policing: that is, it does not reward preventing a crime from being committed.

Nevertheless, in light of principal–agent problems in policing, a hit rate objective may still be a reasonable approximation to police behaviour. It is likely difficult for a police chief to verify that individual officers engage in search activities that deter crime, because the amount of crime deterred is usually not observed. How many criminals an officer apprehends is observed, providing a rationale for rewarding officers on that basis. A model where police act as independent agents trying to catch criminals can be viewed as a second-best objective that a police chief might reasonably adopt.

The other modelling framework examined in the literature is one in which a centralized police chief allocates resources so as to minimize the overall crime rate. We describe the theoretical and empirical results derived using these two different modelling frameworks, with particular emphasis on devising tests for racial bias.

Models of Hit Rate Maximization

The Model of KPT (2001)

Knowles et al. (2001) (KPT) develop a model of police–motorist behaviour that they use to study the implications of racial bias for equilibrium search outcomes. In the model, police officers decide which vehicles to subject to searches, and motorists decide whether to break the law by carrying contraband, such as drugs or illegal weapons, taking into account the probability of being searched.

In the absence of racial bias, each officer pursues a monitoring strategy that maximizes the number of successful search outcomes. Racial bias is introduced as a preference parameter that reduces the perceived cost of searching vehicles of black or Hispanic drivers relative to white drivers, which can lead to over-searching of these groups. An equilibrium implication of racially biased monitoring, shown in KPT and discussed further below, is that the hit rate (the rate at which contraband is seized) should be lower for the groups subject to bias. (The general idea that tastes for discrimination lead to lower profits for discriminators originated with Becker 1957. For further discussion of such tests in policing contexts, see Ayres 2002.)

Let $r \in \{A, W\}$ denote the race of the motorist (African-American or white), assumed to be observable by the police officer. (We assume two groups here, but the analysis extends straightforwardly to more groups.) Let c denote all characteristics other than race that are potentially used by the officer in the decision to search cars, which may be unobserved or only partially observed by the econometrician. For expositional ease, treat c as a one-dimensional variable (results extend to the multidimensional case), and denote the distribution of c in the white and African-American populations by $F(c|W)$ and $F(c|A)$.

It is assumed that an individual police officer allocates search efforts so as to maximize the number of convictions minus a cost of searching cars. Each officer can choose to search motorists of any type (c,r) at a marginal cost of t_r. Normalize the benefit of each arrest to equal 1, so that the cost is scaled as a fraction of the benefit (assume $t_w, t_A \in (0,1)$). In deciding whether to carry contraband, motorists consider the probability of being searched and the penalty if they were to be caught. If they do not carry, their payoff is zero whether or not the car is searched. If they carry, their payoff is $v(c, r) + x$ if not searched and $-j(c, r)$ if searched, where both $j(c, r)$ and $v(c, r)$ are positive. x represents private information of the motorists about their own benefit from carrying contraband. $j(c, r)$ can be interpreted as the cost of being convicted. (If there were discrimination in the court system leading to higher penalties for minority drivers, this could be thought of as operating through $j(c, r)$. KPT do not test for this type of discrimination.)

Denote by $\gamma(c, r)$ the probability that the police officer searches a motorist of type c, r. The expected payoff to a motorist from carrying contraband is

$$\gamma(c,r)[-j(c,r)] + [1 - \gamma(c,r)][v(c,r) + x]. \quad (1)$$

Given $\gamma(c, r)$, the motorist chooses to carry contraband if this expression is greater than zero. Motorists with a high realized value of x strictly prefer to carry drugs and those with small values strictly prefer not to carry. However, police search strategies can be conditioned only on c and r, because x is not directly observed by the police. Let G denote the event that the motorist searched is found guilty of carrying contraband, and denote the probability that a motorist of type c, r carries contraband by $P(G|c, r)$. (We do not allow for the possibility of false accusation by police or planting of evidence, as considered in Donohue and Levitt 2001.)

Assume that the police officer decides on the search probability $\gamma(c, r)$ (the probability of searching each motorist of type c, r) to maximize the number of successful searches, net of costs. He or she solves

$$\max_{\gamma(c,W),\gamma(c,A)} \sum_{r=W,A} \int [P(G|c,r) - t_r]\gamma(c,r) f(c|r)dc,$$

taking as given $P(G|c, r)$. The term $P(G|c, r) - t_r$ represents the expected profit from searching a motorist of type c, r. If $P(G|c, r) - t_r > 0$ then optimizing behaviour implies $\gamma(c, r) = 1$, that is, always search motorists of type c and r. If $P(G|c, r) = t_r$ then the police officer is willing to randomize over whether or not to search type c, r.

KPT introduce racial bias into this framework as a difference in the perceived cost of searching motorists of different races. That is, a police officer is said to be biased against race A (or to have a taste for discrimination) if $t_A < t_W$. If a police officer has no taste for discrimination and yet chooses search probabilities that differ by race, then the equilibrium is said to exhibit statistical discrimination. Statistical discrimination is motivated out of efficiency considerations and not out of racial bias. (Statistical discrimination is used here in the same sense as in Arrow 1973.)

KPT (2001) study the equilibrium implications of this model for the case where officers are homogeneous in their costs of search and motorists are heterogeneous in the benefits they derive from carrying contraband. The model implies the following equilibrium conditions, for all c

$$
\begin{aligned}
&P^*(G|\,c,A) = t_A, P^*(G|\,c,W) = t_W \\
&\gamma^*(c,A) = \frac{v(c,A)}{[v(c,A) + j(c,A)]}, \\
&f\gamma^*(c,W) = \frac{v(c,W)}{[v(c,W) + j(c,W)]},
\end{aligned} \tag{2}
$$

where * denotes equilibrium values.

Suppose that $t_A = t_W = t$ (that is, police officers are not biased). Then, for all c, guilt probabilities at equilibrium must be equal across races:

$$P * (G|c,A) = t = P^*(G|c,W). \tag{3}$$

If guilt probabilities were not equalized, a police officer could do better by reallocating searches towards the group with the higher hit rate.

An important observation is that equalization of hit rates does not imply equalization of search rates. The equilibrium search intensity may be higher for African-Americans even in the absence of racial bias. This happens if $v(c, W)/[v(c, W) + j(c, W)] < v(c, A)/[v(c, A) + j(c, A)]$. That is, if the expected value of carrying drugs is higher or the cost of being convicted lower for black motorists, then the search rate on that group would have to be higher in order to equalize the guilt probability to that of whites.

Equation 3 is the basis for the outcomes-based test proposed in KPT as a test for racial bias (a test for $t_W \neq t_A$). An advantage of the test is that it is implementable even in the absence of complete data on c and on γ^*. It suffices to have data on the frequency of guilt by race conditional on being searched,

$$D(r) = \int P^*(G|c,r)\frac{\gamma^*(c,r)f(c|r)}{\int \gamma^*(s,r)f(s|r)ds}dc.$$

Using Eq. 3 to substitute for $P(G|c,r)$ we get the implication

$$D(W) = t = D(A), \tag{4}$$

which KPT empirically test. In the model, there is nothing special about the characteristic 'race'. The analogue of Eq. 4 should hold for any observed characteristic on which police can condition their search decision. Thus, the model has the strong implication that the guilty rates should be equal for any set of observed conditioning variables, such as age, gender, or type of car.

The assumption that motorists respond to the probability of being searched is key to obtaining a test for bias that is applicable even without data on all the characteristics that police use in the search decision. If motorists did not react to the probability of being searched, testing for prejudice would require data on c. To see why, consider a model where the probability that a motorist with characteristic c and race r carries drugs does not depend on the actions of police, and the only optimizing agents are the police. Let $\pi(c, r)$ denote the probability that a type c,r carries drugs and suppose that $\pi(c, r)$ is increasing in c. Then, it is optimal for police officers to choose two cut-offs k_W and k_A and to search any motorist of race r with a c greater than k_r. In the absence of prejudice, police will choose k_W and k_A so that the probability that types k_W,W and k_A,A are guilty equals the marginal cost t of searching motorists. Without data on c, one cannot empirically identify the marginal motorists and so cannot test the equilibrium implications of this model, in the absence of strong assumptions on the shape of $\pi(c, r)$ and on the distribution of the unobservables. When $\pi(c, r)$ is determined endogenously, the only equilibrium is for $\pi(c, r)$ to equal t_r for all c. Thus, allowing for endogenous response of motorists to the probability of being searched eliminates the problem of having to identify the marginal motorists.

A number of papers have explored extensions or variations of the KPT model. For example, Antonovics and Knight (2004) raise the concern that police heterogeneity is a potential threat to the validity of the outcomes-based tests, and they present evidence that police are more likely to search the vehicles of drivers of a different race. Persico and Todd (2006) generalize the KPT model to allow for police heterogeneity in costs of search and for the possibility that drivers can adapt some of their characteristics to reduce the probability of being monitored. (For example, if drivers with sports cars are subject to high monitoring rates, an individual might choose to drive a different type of car.) They show that the hit rate test is still valid under these extensions. Persico and Todd (2005) further extend the model to allow for imperfections in the monitoring technology, namely, that searches do not always uncover contraband. Even with varying detectability rates across groups, the hit rate test can still be justified as a test for racial bias. (The extensions are developed in an application of the model to monitoring of passengers at airports.)

Hernandez-Murillo and Knowles (2004) consider how to test for racial bias with aggregated data that is contaminated by observations on non-discretionary stops. The KPT model assumes that police searches are discretionary, whereas Hernandez-Murillo and Knowles analyse a dataset from Missouri that mixes discretionary and non-discretionary stops. They derive tests for racial bias (inspired by the nonparametric bounding approach of Horowitz and Manski 1995) that are robust to the contamination.

Dharmapala and Ross (2004) extend the KPT model by relaxing the assumption that police can search any motorist. They impose a technological limitation on the search capacities of police, whereby police observe motorists with probability less than 1. In this case, there can be some motorists for whom the constraint leads them to carry contraband all the time. Police would like to search this group harder to equalize guilty rates, but cannot. In equilibrium, such motorists will be searched whenever police observe them. Dharmapala and Ross (2004) demonstrate that, if this type of motorist is distributed differently among racial/ethnic groups, then the hit rate test breaks down. They also consider the set of equilibria in a modified version of the KPT model in which there are offences of varying levels severity and motorists sort over the level of severity.

The Model of Anwar and Fang (2006)

A limitation of the KPT model is that it assumes that police officers first see some motorist characteristics and then decide whether to search them. A more realistic assumption is that police see some information prior to the stop decision and then acquire more information from interacting with the motorist. Anwar and Fang (2006) develop a framework in which the officers' search decisions can depend on the additional information they acquire after the initial stop. They also allow for the possibility that police behaviour varies with the race of the police officer. For example, white police may be biased against minority drivers and minority police biased against white drivers. (Persico and Todd (2004) also allow for police heterogeneity in the bias, but do not allow for the sign of the bias to differ for individual officers. That is, they do not allow for the possibility that some police may be biased while others may exhibit favouritism, which is the case considered in Anwar and Fang 2006.)

The model of Anwar and Fang (2006) is in the spirit of statistical discrimination models (see, for example, Coate and Loury 1993). It assumes that during the stop and prior to the search decision the police officer observes a noisy but informative signal about whether the driver carries contraband. The signal is informative in the sense that guilty drivers (those carrying illegal contraband) are more likely than innocent drivers to generate suspicious signals, such as nervousness in answering the police officer's questions. (It is assumed that the drivers themselves do not know at the time of deciding whether to carry contraband whether they will generate a signal, only the probability that they will generate one.) As in the KPT model, police officers are considered racially biased if their cost of search depends on the race of the motorist and the objective of officers is to maximize the number of successful searches.

Anwar and Fang (2006) develop two tests. The first is a test for whether police officers of different races use different search criteria when dealing with motorists of the same race, which would indicate police heterogeneity. The test is based on the observation that, if officers do not differ in search costs, then the search rates and success rates of different groups of officers should on average be the same. The second test they develop is a test for racial prejudice that can uncover whether at least one of the groups of officers (for example, white or minority officers) is searching in a racially biased manner, although it cannot distinguish which group is biased.

Models of Crime Minimization

The previous class of models assumed that individual officers adopt search strategies that maximize successful search rates, or so-called hit rates. As noted above, a hit rate objective function would be a reasonable approximation to police behaviour if officers are rewarded on the basis of criminals apprehended, something that is easily observed. As demonstrated in Persico (2002), however, the hit rate objective function does not minimize the aggregate crime rate.

An alternative modelling framework assumes that there is a centralized authority, a police chief, say, who can direct officers to focus their searches on particular subgroups. In such a model, the hit rate test fails as a test of the unbiasedness of the police chief, because, in the equilibrium of such a model, an unbiased police chief will allocate searches to equate the deterrence effect and not the hit rates across groups. Crime deterred is unobserved, making it difficult to test for whether the deterrence effect is being equalized across groups. While one could conceivably introduce racial bias in a crime minimization model in the same way as in the hit rate maximization models – as a difference in the costs of searching different types of motorists – there is currently no empirically implementable test for racial bias in such a model. (Eeckhout et al. 2003, study optimal monitoring strategies for police assuming that the objective is to minimize crime. They show that in some cases it can be optimal to randomly subject even identical motorists to different levels of monitoring. This could be considered random profiling, in that motorists are randomly divided into different groups and are subjected to different levels of monitoring.)

Imposing a Race-Blind Constraint on Police Behaviour

Persico (2002) studies the effects of constraints on police behaviours, within a model where police maximize hit rates, but the assumed socially efficient objective is to minimize the aggregate crime rate. He shows that imposing a 'race-blind' constraint on police search behaviour does not necessarily entail any loss in efficiency and can sometimes increase efficiency. That is, not allowing police to condition their search probabilities on race can sometimes lead to a lower-cost way of achieving a given crime rate. This somewhat surprising result follows because search strategies that aim to maximize arrests do not take into account the deterrence value that arrests have on different groups. The incentive scheme that minimizes the crime rate would place a higher value on arresting motorists of the race that is more likely to be deterred by the prospect of being arrested.

To see how a fairness constraint can increase efficiency, suppose, for example, that whites were more numerous in the population and were also less likely than blacks to carry drugs at a given search rate. In the absence of any constraints on search behaviour, police would search blacks at a higher rate so as to equalize the hit rates across groups. Under a fairness constraint, however, the two groups are pooled and experience the same probability of being searched. In equilibrium, the overall carrying rate will remain the same as in the unconstrained equilibrium and will equal the cost of search. However, equalizing search rates by race leads to an increase in the black carrying rate, and an offsetting decrease in the white rate. If whites are deterred by a relatively small increase in the probability of search and they are more numerous in the population, then it is possible to achieve the same overall carrying rate at a lower search cost in the constrained equilibrium than in the unconstrained equilibrium. Persico (2002) finds that whether imposing a fairness constraint leads to an increase in efficiency (defined as a decrease in the crime rate for the same cost) crucially depends on the proportion of blacks in the population relative to the cost of search.

For further consideration of how to incorporate efficiency and equity considerations into an assessment of racial profiling as a public policy, see Durlauf (2005) and Risse and Zeckhauser (2004). Also, see Dominitz (2003) for discussion of the statistical relationship between various outcomes that could be considered when formulating public policy, such as search rates, find rates, thoroughness of search, rates of detention of the innocent, and rates of apprehension of the guilty.

Racial Profiling, Table 1 Summary of hit rate findings for racial profiling studies

Location	Whites	Blacks	Hispanics	Source
Maryland	22.7	22.0	18.9	Knowles, Persico and Todd (2001)
Florida	32.0	34.0	11.0	Anwar and Fang (2006)
Tennessee	25.1	20.9	11.5	Cohen-Vogel and Doss (2002)
New Jersey	20.1	19.2	10.3	Verniero and Zoubek (1999)
Rhode Island	10.5	13.5	n/r	Farrell et al. (2003)
New York (pedestrian)	23.5	17.8[a]	17.8[a]	Spitzer (1999)
Charlotte, NC	13.0	11.0	n/r	Smith et al. (2004)
Lansing, MI	30.9	24.2	n/r	Carter et al. (2002)
Missouri	6.8	8.7	n/r	Nixon (2003)
San Antonio, TX	23.2	17.5	14.7	Lamberth (2003)
Denver, CO	17.2	14.6	14.9	Thomas and Hansen (2004)
Denver, CO (pedestrian)	16.5	19.7	11.3	Thomas and Hansen (2004)
Los Angeles, CA	18.7	20.6	14.6	Tabulations provided by the LAPD
Sacramento, CA	23.8	18.2	17.2	Greenwald (2003)
San Diego, CA	26.5	22.4	28.0	Cordner et al. (2002)
Washington State	11.0	12.0	5.0	Lovrich et al. (2003)
Wichita, KS	32.0	21.0	n/r	Persico and Todd (2006)

[a]A single hit rate is reported for all minorities

Empirical Evidence

As noted above, models that assume that police maximize hit rates and that motorists take into account the probability of being caught when deciding whether to break the law lead to simple procedures to test for racial bias. The KPT test compares hit rates across different groups of motorists, which can be performed using the type of data that is conventionally available.

Table 1 (reproduced from Persico and Todd 2006) summarizes findings from 16 different city-level and state-level racial profiling studies/ reports, in which the hit rates by race/ethnicity are reported. The table displays what appears to be an empirical regularity: there is not a large disparity in hit rates for black and white drivers, especially when compared with the disparity in search/stop rates. This regularity is puzzling in the context of a crime-minimizing police chief, but not in light of a simple hit rate maximization model, which offers a rationale for the equalization of hit rates across races, namely, (*a*) that police are allocating searches in a way that maximizes efficiency in catching criminals and (*b*) that police departments, on average, are not afflicted by widespread bias against African Americans. In Table 1, the hit rates for Hispanics are in many cases notably lower than that of whites or blacks, which is suggestive of bias against Hispanics. Whether in fact this is really the case can be ascertained only with more work on the police data-sets that are becoming available.

Anwar and Fang (2006) apply their alternative test (not described here, for the sake of brevity) to a data-set on highway stops and searches collected by the Florida Highway Patrol. The data reveal search patterns that differ significantly by race of the officer. Despite the differences in search behaviour, however, the test does not reject the null hypothesis of no relative racial prejudice between black and white officers. (The authors advise caution in interpreting the results, as the test is informative only about relative racial bias and cannot rule out the possibility that all police – of both races – might be biased.)

Summary and Conclusions

Recent advances in the economic literature have led to a better appreciation of the assumptions that underlie different approaches to testing for racial/ ethnic bias in policing. Simple benchmark tests for discrimination only uncover whether a disparity

R

exists; they do not reveal the motivation for the disparity, which plays a key role in racial profiling investigations. Assuming a hit rate objective function and strategic behaviour on the part of both police officers and motorists leads to a simple, empirically implementable outcomes-based test for racial bias. Such a model can potentially explain an observed empirical regularity in many police data-sets to the effect that there is little disparity in hit rates for black and white drivers, despite large disparities in search/stop rates. An alternative modelling framework is one in which a police chief allocates resources so as to minimize the aggregate crime rate. An implication of unbiased policing in this type of model is that the deterrence effect be equated across different groups of motorists, which is difficult to test empirically.

Even if an outcomes-based test concludes that racial disparities in search rates do not reflect racial bias, there is still the question of whether statistical discrimination is justified. Statistical discrimination may be considered unfair, because drivers experience different probabilities of being searched, depending on their race. An intriguing aspect of Persico's (2002) findings is that, when police are maximizing a second-best objective (hit rates), imposing a race-blind constraint can bring them closer to a first-best objective (minimizing overall crime). In practice, though, it may be difficult to ask police officers to simply ignore race in their decision-making. More race-neutral policing might instead be achieved by giving police differential rewards for hit rates on white and minority drivers. Designing optimal incentive schemes for achieving a particular objective is a current area of research.

Finally, the economic literature on racial profiling is relatively nascent, and there are many ways of extending existing models. For example, none of the existing models specifies how a police chief allocates police officers to patrol particular areas. For this reason, existing tests are usually applied to data on highway searches where selective allocation of officers to monitor certain populations is less of an issue. Also, existing theory has mainly been developed for discretionary stops, but a major proportion of police stops and searches are triggered by events that make a search of the vehicle mandatory.

See Also

► Arrow, Kenneth Joseph (Born 1921)
► Becker, Gary S. (Born 1930)

Acknowledgments I am grateful to Steven Durlauf, Hanming Fang and Nicola Persico for helpful discussions.

Bibliography

Antonovics, K., and B. Knight 2004. A new look at racial profiling: Evidence from the Boston Police Department. Working Paper No. W10634. NBER, Cambridge, MA.

Anwar, S., and H. Fang. 2006. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review* 96: 127–151.

Arrow, K. 1973. The theory of discrimination. In *Discrimination in labor markets*, ed. O. Ashenfelter and A. Rees. Princeton: Princeton University Press.

Ayres, I. 2002. Outcomes tests of racial disparities in police practices. *Justice Research and Policy* 4: 131–142.

Becker, G. 1957. *The economics of discrimination*. Chicago: Chicago University Press.

Becker, G. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 76: 169–217.

Border, K., and J. Sobel. 1987. Samurai accountant: A theory of auditing and plunder. *Review of Economic Studies* 54: 525–540.

Carter, D., A. Katz-Bannister, and J. Schafter. 2002. Analysis of the lansing police department MATS data: An eighteen month status report. Report submitted to Chief Mark Alley, Lansing Police Department.

Coate, S., and G.C. Loury. 1993. Will affirmative action policies eliminate negative stereotypes? *American Economic Review* 83: 1220–1240.

Cohen-Vogel, D., and B. Doss. 2002. Vehicle stops and race: A study and report in response to public Chapter 910 of 2000. Report to Comptroller of the Treasury, Comptroller of the Treasury, State of Tennessee, Nashville.

Cordner, G., B. Williams, and A. Velasco. 2002. *Vehicle stops in San Diego: 2001*. San Diego: San Diego Police Department.

Dharmapala, D., and S. Ross. 2004. Racial bias in motor vehicle searches: Additional theory and evidence. *Contributions to Economic Analysis & Policy* 3(1), Article 12. Online. Available at http://www.bepress.com/bejeap/contributions/vol3/iss1/art12. Accessed 2 Sept 2006.

Dominitz, J. 2003. How do the laws of probability constrain legislative and judicial efforts to stop racial profiling? *American Law and Economics Review* 5: 412–432.

Donohue, J. 1999. Expert witness testimony in the case. Chavez v Illinois State Police.

Donohue, J. 2005. The law and economics of anti-discrimination law. Public Law Working Paper No. 101, Yale Law School.
Donohue, J., and S. Levitt. 2001. The impact of race on policing, arrest patterns, and crime. *Journal of Law and Economics* 44: 367–394.
Durlauf, S. 2005. Racial profiling as a public policy question: Efficiency, equity, and ambiguity. *American Economic Review* 95: 132–136.
Eeckhout, J., N. Persico, and P. Todd. 2003. *A theory of random crackdowns*. Mimeo: University of Pennsylvania.
Epstein, R. 1992. *Forbidden grounds: The case against employment discrimination laws*. Cambridge, MA: Harvard University Press.
Farrell, A., D. McDevitt, S. Cronin, and E. Pierce. 2003. *Rhode island traffic stop statistics act final report*. Boston: Institute on Race and Justice, Northeastern University.
Greenwald, H. 2003. Police vehicle stops in Sacramento, CA: Second of three reports. School of Policy, Planning and Development, University of Southern California.
Gross, S., and K. Barnes. 2002. Road work: Racial profiling and drug interdiction on the highway. *Michigan Law Review* 101: 653–754.
Harcourt, B. 2004. Rethinking racial profiling: A critique of the economics, civil liberties, and constitutional literature and of criminal profiling more generally. *University of Chicago Law Review* 71: 1276–1380.
Hernandez-Murillo, R., and J. Knowles. 2004. Racial profiling or racist policing? Bounds tests in aggregate data. *International Economic Review* 45: 959–989.
Horowitz, J.L., and C.F. Manski. 1995. Identification and robustness with contaminated and corrupted data. *Econometrica* 63: 281–302.
Kennedy, R. 1997. *Race, crime and the law*. New York: Pantheon Books.
Knowles, J., N. Persico, and P. Todd. 2001. Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy* 109: 203–229.
Lamberth, J. 2003. Racial profiling data analysis study: Final report for the San Antonio Police Department. Online. Available at http://www.sanantonio.gov/sapd/pdf/LamberthSanAntonioRpt_2003.pdf. Accessed 3 Oct 2006.
Lovrich, N., M. Gaffney, C. Mosher, M. Pickerill, and M. Smith. 2003. WSP traffic stop data analysis project: Data analysis project report. Division of Governmental Studies and Services, Washington State University.
Nixon, J. 2003. Executive summary on 2003 Missouri traffic stops. Missouri Attorney General's Office, Jefferson City. Online. Available at http://www.ago.mo.gov/racialprofiling/2003/racialprofiling2003.htm. Accessed 2 Sept 2006.
Persico, N. 2002. Racial profiling, fairness and effectiveness of policing. *American Economic Review* 92: 1472–1497.
Persico, N., and D. Castleman. 2005. Detecting bias: Using statistical evidence to establish intentional discrimination in racial profiling cases. *2005 University of Chicago Legal Forum,* 217–235.
Persico, N., and P. Todd. 2005. Passenger profiling, imperfect screening and airport security. *American Economic Review* 95: 127–131.
Persico, N., and P. Todd. 2006. Generalizing the hit rates test for racial bias in law enforcement, with an application to vehicle searches in wichita. *Economic Journal* 116(November): F351–F367.
Reinganum, J., and L. Wilde. 1986. Equilibrium verification and reporting policies in a model of tax compliance. *International Economic Review* 27: 739–760.
Risse, M., and R. Zeckhauser. 2004. Racial profiling. *Philosophy and Public Affairs* 32(2): 31–70.
Ross, S. 1996. Mortgage lending discrimination and racial differences in loan default. *Journal of Housing Research* 7: 117–126.
Scotchmer, S. 1987. Audit classes and tax enforcement policy. *American Economic Review* 77: 229–233.
Sherman, L. 1990. Police crackdowns initial and residual deterrence. In *Crime and justice: A review of research*, ed. M. Tonry and N. Morris, Vol. 12. Chicago: University of Chicago Press.
Smith, W., E. Davison, M. Zingraff, K. Rice, and D. Bissler. 2004. An empirical investigation of the possible presence and extent of arbitrary profiling in the Charlotte-Mecklenburg Police Department. Final Report to Charlotte-Mecklenburg Police Department. Department of Sociology and Anthropology, North Carolina State University.
Spitzer, E. 1999. *A report to the people of the state of New York from the Office of the Attorney General*. New York: Civil Rights Bureau.
Stigler, G. 1970. The optimum enforcement of laws. *Journal of Political Economy* 78: 526–536.
Thomas, D. and Hansen, R. 2004. Second annual report, Denver police department, contact card data analysis. Department of Geography, University of Colorado at Denver.
US General Accounting Office. 2000. Customs service: Better targeting of airline passengers for personal searches could produce better results. GAO/GGD-00-38. US Government Printing Office, Washington, DC.
Verniero, P., and P. Zoubek. 1999. *Interim report of the state police review team regarding allegations of racial profiling*. Trenton: Attorney General's Office.

Radical Economics

Diane Flaherty

Abstract

Contemporary radical economics comprises a broad set of methodological approaches, including Marxian political economy,

R

institutionalism, Post Keynesianism, analytical political economy, radical feminism and postmodernism. Unlike radical economics in the mid-1980s, radical thought today emphasizes conflict other than class conflict, policy-relevant analysis and incorporation of more mainstream methods into radical research. Nonetheless, despite substantial evolution, radical economics remains faithful to its original vision. Uniting the various approaches is a set of unchanged core principles, the three most salient of which are the importance of history, embeddedness of individual choice in an institutional environment, and the centrality of conflict to understanding capitalism.

Keywords
Agency; Aggregate demand; Alienation; Altruism; Analytical political economy; Capital controls; Capitalism; Chaos theory; Class; Collective action; Complexity; Conflict; Convergence; Credit controls; Distribution; Endogeneity of preferences; Equilibrium; Experimental economics; Exploitation; Game theory; Globalization; Heterodox economics; *homo economicus*; Inequality; Intrahousehold welfare; Keynesianism; Labour theory of value; Marxian political economy; New institutional economics; Optimization; Path dependence; Post Keynesianism; Postmodernism; Profit; Race; Radical economics; Radical feminism; Simultaneous equation models; Social construction of norms; Surplus; Trade unions; Washington consensus

JEL Classifications
B5

Contemporary radical economics comprises a broad set of methodological approaches, including Marxian political economy, institutionalism, Post Keynesianism, analytical political economy, radical feminism and postmodernism (see for example Pietrykowski 2000; Colander et al. 2004; Dutt 2005). This inclusive definition of radical economics differs in many respects from the radical economics of the mid-1980s. Today radical thought emphasizes conflict outside of class conflict, policy-relevant analysis and incorporation of more mainstream methods into radical research.

Nonetheless, despite substantial evolution, radical economics (or, as it is often called now, heterodox economics) remains at the core faithful to its original vision. All approaches are grounded in a stable set of uniting principles, the three most salient of which are the importance of history, embeddedness of individual choice in an institutional environment and the centrality of conflict to understanding capitalism.

Roots of Radical Economics

Radical economics has always identified its fundamental project as the construction of realistic representations of the capitalist system, the better to identify and redress exploitation, alienation and inequality. Specifically, the shared goal from the beginning has been to incorporate a degree of reality not available in neoclassical models based on assumptions of atomistic individuals optimizing under conditions of complete information and perfect foresight.

The first unifying proposition of radical thought is the importance of history. The past shapes the present through inherited initial conditions: all choices in the current period are made within constraints imposed by history. Moreover, projections into the future must eschew assumptions of perfect foreknowledge or even, in some strands of radical thought, a known risk embodied in a fixed distribution of outcome probabilities. Path dependence renders the future unpredictable and unknowable.

Social construction of norms and endogeneity of preferences to institutional constraints constitute a second set of shared radical concepts. Individuals make choices on the basis of ‘background’ criteria derived from social norms, which in turn arise from stable institutions and rules of the game (Searle 1995).

The radical critique of exogenous preferences does not simply replace individual agency with structure. Rather, most strands of radical thought assume interaction between individual and social structure. The dominant view posits that intentional human agency exerts its most powerful effect on institutions in contentious historical conjunctures (Setterfield 2005; Searle 1995). Conflictual historical periods give rise to re-examination, rejection and perhaps supersession of existing institutions because individuals challenge established norms.

Conflict in turn is seen by radicals as endemic to capitalism and a cause of chronic inefficiency. Conflict at the workplace makes capitalism operate at less than maximum achievable output (Bowles and Jayadev 2005). At the macroeconomic level, conflict over distribution is implicated in lack of stability or consistent growth (Bhaduri 2006; Pollin 1997). Chronic conflict periodically and unavoidably rises to an acute level, crises emerge and institutions once supportive of growth fail to resolve the crisis.

Impetus to the Evolution of Radical Thought

While most current practitioners of radical economics are, as noted above, still working broadly within the original radical vision, since the mid-1980s internal criticisms of radical research have precipitated significant change within the field. Citing lack of coherence or realism, critics urge theoretical overhaul to achieve a truly integrated analysis of history, institutions and conflict.

Immanent criticism argues that radical theory must explain collective action, both to establish the centrality of class and to sustain a critique of the neoclassical behavioural model predicated on atomistic, self-interested individuals. Radical theory in its original incarnation provided little analysis of how individuals decide to engage in collective action. Thus, radical analysis fell short in explaining how or when the working class decides to act in its own interest. Most radical economics rejected early on any explanation derived from an overly simplified materialist assertion that individuals are simply bearers of their class roles. Nonetheless, lacking a clear alternative, radical explanations still tended to fall back into an unexamined and unacknowledged use of functionalist arguments, in which behaviours derive from the structure and requirements of capitalism.

Exploitation, also a central proposition of radical theory, similarly required further elaboration. The labour theory of value as the theoretical foundation of the role of labour as the sole source of surplus and profit sustained damaging attack from analytical Marxists, losing pride of place as the radical theory of profit and prices. Without a labour theory of value, much of crisis theory inherited from Marx in turn became unsustainable.

Finally, despite theoretical commitment to a path-dependent theory of historical change, radical economics tended to fall back into an overly deterministic theory of history, in which capitalism moves according to knowable and immutable laws. The inconsistency of a deterministic theory of history with the notion of path dependence weakened claims to realism and superior understanding of capitalist development.

In addition to identifying such theoretical vulnerabilities of radical theory, critics argued for recognition of new empirical realities in capitalism. Most important has been the growing belief that class is simply too crude a tool with which to analyse conflict. Inequality along many dimensions, particularly race and gender, has emerged as a topic of radical investigation, further undercutting a classical Marxian analysis of capitalism. Radical feminists, for example, question the relevance of exploitation of labour to understanding inequality within the household. Resilience of capitalism in the face of crisis, in contrast to fragility of socialist countries, confronted radicals with another dilemma, since radical theory based on historical materialism had predicted opposite results.

Absence of radical analysis from policy debates fuelled the criticism that radicals neglected empirical work. Complaints have been widespread, coming even from the more Marxist of the radicals. Howard Sherman, for example, has argued that radicals are obliged ‘to always start from actual problems, not from ideal models, universal laws or any rigid rules of research’

(Sherman 1995, p. 262). Empirical Marxists bemoaned radicals' insistence on the purity of Marxian categories, which led radicals to avoid empirical analysis, resulting in exclusion from policy discussions (Dunne 1991). Prominent non-Marxist radicals too have argued that radicals must confront reality through policy-relevant research if they are to claim superior reality of their analysis (Reich 1995).

Contours of Change

Three shifts in the body of radical research most clearly embody these critiques of early radical theory. First, characterization of the core injustice of capitalism has moved from a narrowly defined concept of exploitation at the point of production to broader inequality beyond both production and class. Second, the role of individual choice relative to structure has become more prominent in explanations of the nature and development of capitalism. The third shift marks movement away from a still 'virtual' radical representation of capitalism to ever more realism.

The prominence given now to inequality signals a further loosening of bonds to Marxian theory. Inequality encompasses race and gender relations as sites of conflict independent of class. As a result, class is now one of several fault lines along which capitalism is both unequal and fragile, and cannot be said to be determinate of relations in other spheres of conflict. Attention to agency reflects that radicals now take seriously the need to demonstrate the inapplicability of *homo economicus* either at the point of production or at other sites such as the household. Thus, the new radical economics unpacks the 'black boxes' of the Marxian theory from which it originated. Gone are the stylized facts of homogenous workers confronting homogeneous capitalists or of a unitary household providing a sanctuary from, as well as valueless inputs to, capitalist production. On the agenda now are, for example, the effects of altruism on intra-household allocation decisions and the role of race in worker decisions to participate in strikes.

Similarly, a key point of the shift to the 'real' is to examine actually existing capitalist countries rather than construct an ideal capitalist type in relation to which actual economies are to be interpreted. The implication for policy is clear. Capitalist countries differ; some provide considerable room for redress of inequality and reduction of conflict, hence improved efficiency. Rather than condemning capitalism as an abstract, necessarily exploitative system, radicals increasingly focus on specific sites within capitalism where potential for improvement may be found. Again, even radicals more firmly within a Marxian tradition call for and applaud a new determination to 'come to grips with . . . the realities of contemporary capitalism as opposed to the creation of a "virtual world"' (Fine 2002, p. 2062).

Contested Evolution

These three evolutionary shifts, while evident throughout radical research, are nonetheless both contested and incomplete. Movement towards fulfilling the promise of more realistic representations of capitalism has raised thorny issues of subject and method. Radicals now contend over the appropriate definition of reality and the ability of competing theories and methods to represent reality.

Radical analysis of globalization provides one example of competing visions of what is real and of methods appropriate to represent reality. On this topic there is still general agreement on the core principles of a radical theory and on the inadequacies of mainstream theory. Radicals reject the so-called Washington consensus, a set of liberalizing policies derived from the claim that opening of markets will lead to improved efficiency and economic growth. Radicals counter both efficiency and growth propositions of the liberalizing story, while focusing instead on distributional consequences of globalization. (Arestis and Sawyer 2002, and Baker et al. 1998, provide overviews of critiques of liberalization.)

Beyond unity in opposition to the free market position, however, radicals disagree profoundly on critical features of globalization. The appropriate definition and indicators of globalization are in dispute, meaning that radicals cannot agree on such basic issues as the degree of globalization of the world economy and role of nation states versus transnational corporations (Glynn and Sutcliffe 1999). Differences also are apparent in competing theoretical understandings of how globalization affects distribution and growth.

The Asian Tigers' perspective on globalization focuses on aggregate demand as the engine of growth and maintains that aggregate demand must be sustained domestically, requiring both controls on capital flows and avoidance of excessive competition (Crotty 2001; Baker et al. 1998). In contrast to the Washington consensus perspective, unemployment in this view is caused by insufficient aggregate demand rather than inflexible labour markets. The role of the state in a globalized economy, often modelled on recent Korean experience, is not only to manage domestic demand but also to develop non-traditional export industries through central control of credit. Class conflict plays at most a peripheral role. Instead, the main locus of conflict is between developed capitalist and developing countries.

This nation-state version of the Keynesian approach calls for controls on capital flows in part to sanitize domestic economics from the effects of 'hot' money. Within the same general aggregate demand framework, however, others deny that nation states have the capacity to control capital flows. A 'one-world' position, for example, calls instead for an international financial authority to regulate financial flows. This authority further must intervene even into the shaping of firm-level decisions to achieve consistency between firm incentives and international financial stability (Eatwell and Taylor 2002).

An alternative radical vision disputes the relevance of a Keynesian model in the current historical conjuncture because history is not reversible. If history is indeed path-dependent, a theory grounded in history and institutions must conclude that it is not possible to revive the golden age of capitalism through a return to Keynesian policy. The current leaden age cannot be regilded to reproduce the growth of the post-war period exactly because the institutions that once supported growth through demand management policy are not sustainable in the globalized economy. In the new economic environment, it is not possible to reconstruct the institutions that once fuelled growth, namely, concentrated industrial sectors, strong trade unions (at least in manufacturing sectors), growing export markets and a world trade system denominated in dollars.

The golden age is also questioned for lack of attention to intra-country conflict of class, gender or race. The nation state is not a homogeneous unit in which aggregate growth benefits all. Moreover, this position contends that a new world order has emerged in which transnational corporations have the character and goals of a supra-national capitalist class. If capitalism has moved to a fundamentally new structure in which nation states have lost the power to determine policy for good, the restorative power attributed to Keynesian macro policy is a delusion.

A third radical vision takes a more positive view of globalization by examining intra-country class relations. This approach sees globalization as offering new space to national policies that can simultaneously increase growth and improve distribution. Focusing on capitalist inefficiency due to conflict, Bardhan et al. (2006) conclude that pressures of globalization can force national governments to implement policies that both improve efficiency and reduce inequality. While the Bardhan et al. story rests on microeconomic behaviour, at the macro level, too, national policy and internal class dynamics are linked. The main contention is that poor countries too often merely use protection from the world economy to redistribute income and assets from the poor to the rich (see for example Griffin 1998). Globalization can undercut the entrenched power of exploitative elites and hence enhance equality as well as growth.

The last divide among radicals to note here is disagreement about the role of foreign investment. The dominant position for some time has been that foreign investment is detrimental to developing countries because of concentration in low- wage industries and competition with existing domestic production. The stylized facts of this view are that foreign investment is footloose and relentlessly pursues ever-lower wages. Therefore, on balance, foreign investment creates negative rather than positive spillover effects. Recently, however, several radical research projects have challenged the accuracy of this picture, pointing to a different set of realities. Nations are seen to have leverage over the terms and flow of foreign investment. Leverage in turn comes from political stability and skilled workers, which, rather than low wages, are the major determinants of investment patterns. Moreover, open countries have succeeded and closed countries have failed, depending upon the role of the domestic government in exploiting opportunities of openness (Chang 1998).

Nihilism or High Theory?

The globalization debate highlights the point that as radical theory has attempted to move towards less 'virtual' representations of capitalism, hitherto concealed splits have surfaced. Much of the purported greater realism comes from analytical methods based in mathematical or formal models, forcing radicals to confront underlying divisions about the proper role of mathematics and analytical models. Competing radical positions now contend over fundamental issues of method, particularly over whether radical core principles are sustained or abandoned by use of formal analytical techniques.

Debate over use of game theory and experimental economics is one illuminating example of current controversy. If the radical project includes understanding endogeneity of preferences and providing a more realistic concept of individual rationality than neoclassical economics, game theory would seem to provide a powerful tool. Feminist economics in particular seeks to move beyond a *homo economicus* model of behaviour to demonstrate that 'many alternatives exist to the traditional self-interested model, with motivations responding, for instance, to notions of altruism, fairness, and reciprocity' (Beneria 1999, p. 71; see also, Folbre and Goodin 2004). Further, policy relevance is the explicit goal of much radical experimental economics. Cross-country experiments are seen to yield insights to improve 'the design of institutions and contracts, the allocation of property rights, the conditions for successful collective action . . .', all considerations dear to radical economists (Henrich et al. 2001, p. 76).

Game theory from this perspective achieves greater realism by identifying parameter values or relationships from which a range of outcomes, here varying preferences or levels of collective action, may arise. To its proponents game theory offers a mechanism for demonstrating that there are no immutable laws or behaviours applicable to all times and all places. Norms and preferences are endogenous to and, therefore, vary with institutional arrangements, whether across countries or households.

Critics vehemently oppose game-theoretic models of preferences or norms exactly on grounds of insufficient realism. In this view, any reduction of complex reality to a model privileging at most a few variables and relationships is antithetical to the radical claim to realism. Models which use any form of optimization are considered to be ahistorical and lacking in institutional specificity, including work by new institutionalists like Bowles and Gintis (1998). Ben Fine, among many others, asserts that such work 'relies on utility, production, inputs and informational asymmetries, timeless and rootless optimizing of individuals . . .' (Fine 2002, p. 2060).

Equilibrium, too, is a matter of much dispute. The strongest repudiation of equilibrium comes from the recent temporal single system interpretation of Marx, which maintains that incorporating history into theory requires neither equilibrium nor disequilibrium models, but non-equilibrium (Kliman and McGlone 1999). To many radicals the complexity of history and the necessity of non-equilibrium render inadmissible use of techniques such as optimization or simultaneous equation models (Lawson 2006).

Defenders of the use of analytical models and mathematics respond that all analysis, mathematical or not, requires the same process of model construction in simplifying reality to a small set of main variables and relationships. The narrow neoclassical concept of equilibrium, with implausible assumptions about foresight and information, not equilibrium itself, must be abandoned. Without some notion of equilibrium, theory is simply nihilistic: 'the social world is complex and determinate but it is impossible to say anything systematic about it' (Foley 2003, p. 3). An alternative equilibrium concept, compatible with heterodox theory and goals, can be defined 'in the general sense of the balancing of forces within a particular model . . .' without any market clearing or settlement to 'tranquil states' (Dutt 1994, p. 3).

Adding fuel to the debate over the requirements of radical theory is the proposition that theories are converging and radicals are all post-Walrasians now. Convergence contends that mainstream economics is no longer bound to what Colander et al. (2004) and Colander (2005) call the Walrasian unholy trinity of rationality, equilibrium and greed. Method and message are no longer linked because new methods explicitly are not ahistorical and asocial and hence are appropriate for radical inquiry (Gibson 2005). Michael Reich has argued further that the liberal wing of neoclassical economics now can accommodate analyses of 'disequilibrium economics, non-market-clearing equilibria, multiple equilibria and the new institutional economics, which have brought radical economics 'out of the ghetto' and into the liberal mainstream' (Reich 1995, p. 50). Duncan Foley expands this point, asserting that complexity and chaos theory finally can liberate radicals from dependence on the concept of 'determination in the last instance' or functionalist arguments to close the system (Foley 2003).

The convergence contention poses a sharp choice for radicals. If new methods like complexity and chaos theory are not just consistent with but necessary for preservation of core radical principles, radicals who repudiate new methods are abandoning history, institutions and conflict as central concerns. The other pole of the dilemma emerges starkly from Colander's otherwise sympathetic assessment of post-Walrasian theory. While supporting Foley's call for more sophistication in radical analysis, Colander's contrasting conclusion is that, in the face of increasing complexity, recent theoretical innovations cannot provide a guide to policy but only an 'aid to one's intuition' (Colander 2005, p. 23). With more realistic representations of capitalism the very complexity of the analytical tools means that no precise policy can be devised and we can only, as Colander says, muddle through.

Conclusion

Radical economics remains, as it was in the mid 1980s, a body of thought defined by common core principles while divided on method. What has changed is the depth of division across the several strands of radical economics. Attempts to develop more realistic and nuanced analyses of capitalism, together with the emergence of new methods of analysis, have generated sharp conflict over both method and object of radical inquiry. To the extent that the choice facing radicals is indeed nihilism or high theory, the centre of the paradigm is eroding and common ground is being lost. Nonetheless, the shifts in radical thought since the mid-1980s have yielded significant positive results. Radicals are indeed more involved in policy discussions and more engaged with data, achieving successful policy interventions as exemplified by living wage legislation (Pollin 2002). Self-criticism also has opened space for re-examination of basic tenets of radical theory and energized debate, which bodes well for continued dynamism and evolution of the broad radical paradigm.

See Also

▶ Economy as a Complex System
▶ Family Economics
▶ Feminist Economics
▶ Institutional Economics
▶ Keynesianism
▶ Marx's Analysis of Capitalist Production
▶ Methodology of Economics
▶ Post Keynesian Economics
▶ Postmodernism
▶ Social Norms

Bibliography

Arestis, P., and M. Sawyer. 2002. *Neo-liberal economic policy*. Cheltenham: Edward Elgar.

Baker, D., G. Epstein, and R. Pollin. 1998. *Globalization and progressive economic policy*. Cambridge: Cambridge University Press.

Bardhan, P., S. Bowles, and M. Wallerstein. 2006. *Globalization and egalitarian redistribution*. New York: Russell Sage Foundation.

Beneria, L. 1999. Globalization, gender and the Davos man. *Feminist Economics* 5: 61–83.

Bhaduri, A. 2006. Endogenous economic growth: A new approach. *Cambridge Journal of Economics* 30: 69–84.

Bowles, S., and H. Gintis. 1998. *The evolution of strong reciprocity*. Amherst: Department of Economics, University of Massachusetts.

Bowles, S., and A. Jayadev. 2005. *Guard labor*, Working paper. Department of Economics, University of Massachusetts.

Chang, H.J. 1998. Globalization, transnational corporations and economic development: Can the developing countries pursue strategic industrial policy in a globalizing world economy? In *Globalization and progressive economic policy*, ed. D. Baker, G. Epstein, and R. Pollin. Cambridge: Cambridge University Press.

Colander, D. 2005. Post-Walrasian macro policy and the economics of muddling through. In *Interactions in analytical political economy: Theory, policy and applications*, ed. M. Setterfield. Armonk: M.E. Sharpe.

Colander, D., R.P.F. Holt, and J.B. Rossiter Jr. 2004. The changing face of mainstream economics. *Review of Political Economy* 16: 485–499.

Crotty, J. 2001. *Core industries, coercive competition and the structural contradictions of global neo-liberalism*, Working paper. Department of Economics, University of Massachusetts.

Dunne, P. 1991. *Quantitative marxism*. Cambridge: Polity Press.

Dutt, A. 1994. Analytical political economy: An introduction. In *New directions in analytical political economy*, ed. A. Dutt. Aldershot: Edward Elgar.

Dutt, A. 2005. On Post-Walrasian economics, macroeconomic policy and heterodox economics. In *Interactions in analytical political economy: Theory, policy and applications*, ed. M. Setterfield. Armonk: M.E. Sharpe.

Eatwell, J., and L. Taylor. 2002. *International capital markets*. Oxford: Oxford University Press.

Fine, B. 2002. Economics imperialism and the new development as Kuhnian paradigm shift? *World Development* 30: 2057–2070.

Folbre, N., and R. Goodin. 2004. Revealing altruism. *Review of Social Economy* 62: 1–25.

Foley, D. 2003. *Unholy trinity: Labor, capital and land in the new economy*. London: Routledge.

Gibson, B. 2005. Thinking outside the Walrasian box. In *Interactions in analytical political economy: Theory, policy and applications*, ed. M. Setterfield. Armonk: M.E. Sharpe.

Glynn, A., and R. Sutcliffe. 1999. Still underwhelmed: Indicators of globalization and their misinterpretation. *Review of Radical Political Economics* 31: 111–131.

Griffin, K. 1998. Comment on Harris and Mitchie 'The effects of globalization on policy formation in South Africa'. In *Globalization and progressive economic policy*, ed. D. Baker, G. Epstein, and R. Pollin. Cambridge: Cambridge University Press.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economics Association Papers and Proceedings* 91: 73–78.

Kliman, A., and T. McGlone. 1999. A temporal single-system interpretation of Marx's value theory. *Review of Political Economy* 11: 33–52.

Lawson, T. 2006. The nature of heterodox economics. *Cambridge Journal of Economics* 30: 483–506.

Pietrykowski, B. 2000. A primer in political economy. In *Political economy and contemporary capitalism*, ed. R. Baiman, H. Boushey, and D. Saunders. Armonk: M.E. Sharpe.

Pollin, R. 1997. *The macroeconomics of saving, finance and investment*. Washington, DC: University of Michigan Press.

Pollin, R. 2002. What is a living wage? Considerations from Santa Monica, CA. *Review of Radical Political Economics* 34: 267–273.

Reich, M. 1995. Radical economics: Successes and failures. In *Heterodox economic theories*, ed. F. Moseley. Aldershot: Edward Elgar.

Searle, J.R. 1995. *The construction of social reality*. London: Penguin.

Setterfield, M. 2005. Interactions in political economy: An introduction. In *Interactions in analytical political economy: Theory, policy and applications*, ed. M. Setterfield. Armonk: M.E. Sharpe.

Sherman, H. 1995. *Reinventing marxism*. Baltimore: Johns Hopkins University Press.

Rae, John (1796–1872)

K.H. Hennings

Keywords

Bentham, J.; Böhm-Bawerk, E. von; Capital accumulation; Capital goods; Capital theory; Conspicuous consumption; Division of labour; Free trade; Infant-industry protection; Internal rate of return; Inventions; Lauderdale, Eighth

Earl of; Liquidity preference; Rae, J. (1798–1872); Schumpeter, P. A.; Smith, A.; Social preferences; Social rate of return; Technical progress; Time preference; Veblen, T.; Wicksell effects

JEL Classifications

B31

John Rae was born in Aberdeen on 1 June 1796 into a merchant and shipping family. He graduated from the University of Aberdeen in 1815 and read medicine in the University of Edinburgh, but had to abandon his studies when his father's business failed in 1817. He emigrated to Canada in 1822 and turned to medical practice (whence 'Dr' Rae) and school teaching. He also participated in public affairs, but his career was shattered in 1848 when he was dismissed on spurious grounds after becoming embroiled in controversies about church control of education. Rae set out to start a new life first in California, and then the Hawaiian island of Maui. After another 20 years of farming, teaching, providing medical services to the natives, and serving as district judge and notary public, Rae went to live with a former pupil in New York, where he died on 12 July 1872.

None of Rae's many misfortunes and distractions could quell his scientific curiosity. He reported scientific experiments and inventions, lectured on scientific subjects, and wrote on public affairs, geology, and Polynesian language and customs (James 1965). The only book he ever managed to get published, his *Statement of Some New Principles on the Subject of Political Economy* (1834), originally intended as an appendix to a larger work on the natural history and statistics of Canada, is one of the highlights of classical economic theory.

Rae's economics is rooted in a natural history of man which he had conceived in the tradition of Montesquieu, Turgot and the Scottish Enlightenment, but never came to execute. Political power and economic progress are seen to result not from the pursuit of self-interest, but to require 'social instincts' which create 'an intelligent and moral community' that furthers both the 'effective desire of accumulation' and the 'rational spirit of invention'. Charging Adam Smith with building his system exclusively on the pursuit of self-interest, and neglecting the role of inventions, Rae contended that economic activity is based primarily on an unselfish regard for the future. In consequence Rae emphasized the temporal aspect of economic activity, and developed a theory of capital accumulation and technical progress which goes far beyond what can be found in Adam Smith or other classical writers.

In language which Fisher was to take up, Rae argued that 'provident forethought' leads man to create 'instruments', that is, capital goods, in order to change the course of events. The sum total of such instruments constitutes the wealth of a society. All instruments are formed, directly or indirectly, by labour; all have the capacity to provide, directly or indirectly, for future wants; and they need time before they are finally exhausted (land being a special case). Rae assumes that the cost of production, and capacity, of any instrument can be measured, in a given society, in exogenously given wage units. All instruments whose capacity exceeds their cost of production can 'be arranged in . . . a series, of which the orders are determined, by the proportions existing between the labour expended in the formation of instruments, the capacity given to them, and the time elapsing from the period of formation to that of exhaustion' (1834, p. 100). Rae expresses this 'order' by the time which elapses before the instrument has yielded twice its cost of production, that is, by n in the expression $(1 + \mathrm{r})^n = 2$ where r is the internal rate of return of the quasi-rents associated with the instrument. Rae rejected working with the latter because it leads, in his view, to the assumption that the stock of all instruments is 'an homogeneous quantity' which he 'found to be the foundation of much of the contradictions, in which the reasonings on these subjects are involved' (1834, p. 197). His calculation rests on the assumption that every instrument can be associated with a unique rate of return. This need not be the case, but the possible multiplicity of internal rates of return does not affect his argument.

R

Rae argued that with knowledge stationary, both capital widening and capital deepening (that is, increasing the durability of instruments) necessarily lower the internal rate of return. Nevertheless, capital goods will be created as long as their internal rate of return is higher than the 'effective desire of accumulation', or rate of time preference, which Rae also expresses in time periods, that is, by m in the expression $(1 + s)^m = 2$ where s is the rate of time preference. Such time preference exists because life is finite and its end uncertain, and because ' passion' is often stronger than 'reason'. But it is counteracted by the concern for future generations, or what Rae called 'social and benevolent affections' (1834, p. 122), which depend on a healthy climate that increases life expectancy, or on social circumstances such as internal and external security, good government, and so on. Hence the strength of the 'effective desire of accumulation', which Rae considers as much a social habit as an individual inclination, varies from one society to another. Variations from one person to another, Rae shows in an almost neoclassical manner, will be equalized by the exchange of instruments among them, so that a social rate of time preference can be juxtaposed to an internal rate of return which is equalized across different 'employments' by profit-seeking 'merchants'.

Rae defines the equality of the social rate of return with the social rate of time preference as a stationary state in which accumulation ceases. 'Gravitation' towards it is slow. In a comparative static analysis Rae shows that the division of labour – which he views as a consequence of the accumulation of capital rather than its cause, as Adam Smith did – reduces the time for which instruments lie idle, and consequently increases their quasi-rents; hence more instruments can be created before the stationary state is reached, and wealth is increased. Similarly, foreign trade is said to increase the productivity of instruments, while conspicuous consumption (his term) will lower the effective desire of accumulation. Rae also argues that as accumulation proceeds, more and more wealth will be tied up in instruments of increasing durability; hence the value of cash balances, and thus liquidity preferences, will increase. But far and away the most important factor making for changes in the progress of accumulation was in Rae's view the progress of inventions. Apart from raising quasi-rents, and hence the internal rate of return, and thus providing scope for more accumulation, inventions also raise the value of existing capital goods. Obviously assuming that these Wicksell effects were positive, Rae placed such capital 'augmentation' alongside capital accumulation as a factor in creating wealth. Indeed, Rae ascribes to inventions a more important role for economic progress (and thus the creation of political power) than capital accumulation, and criticizes Adam Smith for emphasizing savings too much, and neglecting technical progress.

The policy conclusions Rae draws from his analysis are also used to controvert Adam Smith. Instead of pursuing a policy of non-intervention, the 'legislator' should stimulate foreign trade and technical progress, encourage the transfer of knowledge, tax luxuries, and use tariffs to protect infant industries.

It was in this sense that Rae tried to expose 'the fallacies of the system of free trade, and of some other doctrines maintained in the *Wealth of Nations*', as he announced on his title page. But, issued in the midst of a protectionist campaign, Rae's book was mistaken as a heavy-going anti-free-trade tract, and ignored. It did find a champion in Nassau Senior (Bowley 1937, ch. 4) and through him in J.S. Mill, who quoted from it copiously in his *Principles* (1848), comparing Rae on accumulation to Malthus on population. But there the matter rested, except that it seems to have had a strong influence on Hearn's *Plutology* (1863). Rae was re-discovered by Mixter (1897) as a forerunner of Böhm-Bawerk, who acknowledged him as such (1900, ch. XI) despite some criticism. Together with a (botched, because re-arranged) reprint of Rae's book by Mixter (Rae 1905), this brought Rae's work to the attention of capital theorists such as Irving Fisher (1907, 1930) who dedicated one of his main works to Rae, as well as Wicksell and Åkerman. It also influenced Schumpeter's (1911) concept of economic development, and Veblen's (1899) notion of conspicuous consumption.

In his criticism of Adam Smith, Rae did not go beyond Bentham (1787) and Lauderdale (1804). But he added poignancy because he derived it from a theory of economic development which was altogether novel. Based on a materialist conception of capital and a vintage-type approach complete with the distinction between capital goods and their value, Rae clearly separated the supply of from the demand for capital goods, and investigated their determinants. He saw but dimly the equality between discounted marginal returns and marginal costs, but he was clear about the equality of opportunities to invest to the 'inclination . . . to yield up a present good'. He was quite clear, too, about the equality between the rate of return on capital and on money, and about what brought about such equalities: and also about the effects technical progress and the growth of knowledge have upon both demand and supply of capital. All this adds up to a remarkably original and creative performance which was, like that of Gossen, Cournot or Thünen, ahead of its time.

Selected Works

1834. *Statement of some new principles on the subject of political economy, exposing the fallacies of the system of free trade, and of some other doctrines maintained in the 'wealth of nations'*. Boston: Hilliard Gray & Co.
1905. *The sociological theory of capital, being a complete reprint of the new principles of political economy 1834*, ed. C.W. Mixter. New York: Macmillan.

Bibliography

Åkerman, G. 1923–4. *Realkapital und Kapitalzins*. Stockholm: Centraltryckeriet.
Bentham, J. 1787. Defence of usury. In *Jeremy Bentham's economic writings*, ed. W. Stark, vol. 1. London: George Allen and Unwin.
Birchler, U.W. 1980. *John Rae (1796–1872). Seine Theorie der wirtschaftlichen Entwicklung*. Bern: Lang.
Böhm-Bawerk, E. von. 1900. *Kapital und Kapitalzins*, vol. 1, *Geschichte und Kritik der Kapitalzines-Theorien.* 2nd ed. Innsbruck: Wagner. Trans. as: *Capital and interest, vol. 1, history and critique of interest theories.* South Holland: Libertarian Press, 1959.
Bowley, M. 1937. *Nassau senior and classical economics*. London: George Allen & Unwin.
Edmonson, N. 1970. John Rae and liquidity preference. *History of Political Economy* 2: 432–440.
Fisher, I. 1907. *The rate of interest*. New York: Macmillan.
Fisher, I. 1930. *The theory of interest*. New York: Macmillan.
Hearn, W.E. 1863. *Plutology, or the theory of the efforts to satisfy human wants*. Melbourne: Robertson.
James, R.W. 1965. *John Rae, political economist. An account of his life and a compilation of his main writings*. 2 vols. Toronto: University of Toronto Press.
Lauderdale, [J. Maitland] Earl of. 1804. *An inquiry into the nature and origin of public wealth.* Edinburgh: Constable.
Lehmann, H. 1937. *John Raes Werk, seine philosophischen und methodologischen Grundlagen*. Dresden: Dittert.
Mill, J.S. 1848. *Principles of political economy*, ed. W.-J. Ashley. London: Longmans, 1909.
Mixter, C.W. 1897. A forerunner of Böhm-Bawerk. *Quarterly Journal of Economics* 11: 161–190.
Mixter, C.W. 1902. Böhm-Bawerk on Rae. *Quarterly Journal of Economics* 16: 385–412.
Robbins, L. 1968. *The theory of economic development in the history of economic thought*. London: Macmillan.
Schumpeter, J.A. 1911. *Theorie der wirtschaftlichen Entwicklung*. Leipzig: Duncker & Humblot. Trans. as *The theory of economic development*. Cambridge, MA: Harvard University Press, 1934.
Spengler, J.J. 1959. John Rae on economic development: A note. *Quarterly Journal of Economics* 73: 393–406.
Veblen, T.B. 1899. *The theory of the leisure class*. New York: Macmillan.

Rae, John (1845–1915)

Andrew Skinner

R

Keywords
Rae, J. (1845–1915); Eight-hour movement

JEL Classifications
B31

John Rae was born in Wick, Caithness, Scotland on 26 May 1845, the eldest son of William Rae who was for some years Provost of the town. Rae received his early education at Edinburgh Academy, before proceeding to university in the city.

He graduated in 1866 with first class honours in philosophy. Rae was awarded an honorary doctorate in 1897 by his alma mater. He died on 19 April 1915, having spent the last 15 years of his life in London, and was buried in Wick.

John Rae has been variously described as 'author and journalist' (1953, p. 582) and as 'economist, writer on socialism' (1966, p. 1057). He was certainly all of these, publishing numerous articles, notably in the *Fortnightly Review* (1885), the *Temple Bar* (1882, 1883, 1897), *MacMillan's Magazine* (1893) and the *National Review* (1889). The great bulk of his considerable output is to be found in the *Contemporary Review* (from 1880) of which he was assistant editor.

Rae's contributions to the *Contemporary Review* disclose an interest in at least five major areas. These include a review of contemporary literature on social philosophy (in seven parts), and a number of articles on the Socialism of Karl Marx and the Hegelians, Christian Socialism in Germany, and State Socialism and Social Reform. Rae also contributed articles on the crofting problem in the Highlands, supplementing these with pieces on the Highland Shealing *(Temple Bar,* 1883) and on the Scotch Village Community *(Fortnightly Review,* 1885). Rae wrote a number of articles on taxation and a review of recent economic literature. Finally, he addressed questions of industrial relations, in considering the implications of the eight-hour day in the context of unemployment and of foreign competition.

Rae's journalistic interests resulted in three major books. The first of these was entitled *Contemporary Socialism* (1884). This was followed by *Eight Hours for Work* (1894), a book which consisted largely of his articles on labour questions, supplemented by chapters on the connection between hours and wages, the eight- hour movement of 1833, and current legislative proposals.

John Rae is now best known for his admirable *Life of Adam Smith* (1895) which was favourably reviewed in *The Times* for 8 March 1895 as presenting a 'vivid picture' of his subject. The review also drew attention to the point that the book's real merit lay 'not in the originality of the matter, but in the patient industry, with which Mr Rae has collected his materials, old and new, and in the skill and judgement with which he has presented them to the reader'.

While more critical of Rae's scholarship (1965, p. 12), Jacob Viner has noted that Rae was a trained writer who made his *Life* 'an interesting and highly readable book' (p. 13). Viner also drew attention to the remarkable fact that 'As a comprehensive biography, it had no substantial predecessor. Seventy years after its publication, it still has no substantial successor' (1965, p. 5). These judgements are still valid.

See Also

► Smith, Adam (1723–1790)

Selected Works

1884. *Contemporary socialism.* London/Wick: W. Isbister. 2nd ed., enlarged. London: Sonnenschein, 1891; 3rd ed. (with additional chapter). London: Swan Sonnenschein, 1901; 4th ed., London: Swan Sonnenschein, 1908.
1894. *Eight hours for work.* London: Macmillan.
1895. *Life of Adam Smith.* London: Macmillan.

Bibliography

The Wellesley index to Victorian periodicals, 1824–1900. 1966. vol. 1, ed. W.E. Houghton. Toronto: Toronto University Press.
Viner, J. 1965. Guide to John Rae's *Life of Adam Smith*. In *Rae's Life of Adam Smith*. New York: Kelley.
Who was who 1897–1915. 1953. vol. 1, 4th ed. London: Adam & Charles Black.

Ramsey Model

David M. Newbery

Keywords
Arbitrage; Bellman's dynamic programming; Bliss; Calculus of variations; Capital gains; Capital theory; Commodity taxation; Dual

economies; Exhaustible resources; Golden rule; Hamiltonians; Heterogeneous capital goods; Intertemporal competitive equilibrium; Intertemporal welfare maximization; Investment; Keynes-Ramsey rule; Life-cycle savings; Marginal utility of consumption; Neoclassical growth theory; Optimal accumulation; Optimal growth; Optimal monetary policy; Optimal saving rate; Optimal taxation; Planning; Pontryagin maximum principle; Population growth; Production functions; Public investment criteria; Ramsey model; Ramsey, F. P.; Rate of interest; Social welfare function; Stationary state; Taxation of capital

JEL Classifications

O4

Frank Plumpton Ramsey died at the age of 26 after making brilliant contributions to philosophy mathematical logic, and, of course, economics. His two contributions to economics both appeared in the *Economic Journal,* then edited by J.M. Keynes. The first, ‘A Contribution to the Theory of Taxation’, published in March, 1927, laid the foundation for the modern theory of commodity taxation. The second, the subject of this entry, was ‘A Mathematical Theory of Saving’, published in December, 1928. Keynes, in his obituary notice published two months after Ramsey’s death, in the *Economic Journal* of March, 1930, described the latter as ‘one of the most remarkable contributions to mathematical economics ever made, both in respect of the intrinsic importance and difficulty of its subject, the power and elegance of the technical methods employed, and the clear purity of illumination with which the writer’s mind is felt by the reader to play about its subject’.

Ramsey asked how much of its income should a nation save and derived a remarkably simple rule, usually known as the Keynes-Ramsey rule, as Keynes provided a non-technical argument for the result. The rule states that the rate of saving, multiplied by the marginal utility of consumption, should always be equal to the amount by which the total net rate of enjoyment of utility falls short of the maximum possible rate.

Ramsey’s formulation of the problem served as a model for almost all subsequent studies of optimal economic growth, and, with the critical addition of a growing population, might have created neoclassical growth theory about 30 years before Solow’s (1956) contribution. He assumed a one-good world, in which labour with a stock of capital would produce a flow of output, part of which was consumed, and the balance was saved and thereby added to the stock of capital. The objective, or criterion, was to achieve the maximum level of enjoyment, summing over all time, where enjoyment was the utility of consumption, $U(C)$, less the disutility of working, $V(L)$. Ramsey made three crucial assumptions which together allowed him to solve explicitly an otherwise intractable problem. He assumed that there was no population growth, no technical progress, and no discounting of utility, ‘a practice which is ethically indefensible and arises merely from the weakness of the imagination’ (Ramsey 1928, p. 543). He further supposed that there was a ‘maximum *obtainable* rate of enjoyment’ called *Bliss, B,* either because of capital or consumption saturation. As neither population grows nor future utilities are discounted, Ramsey then argues, rather informally, that it must be desirable to save enough to eventually reach bliss, or approximate to it indefinitely. To stop short means forgoing a finite amount of utility, which, summed over an infinite time horizon, is infinitely costly. Formally, Ramsey deals with this problem of a potentially unbounded integral of utility (summed without discounting over infinite time) by *minimizing* the amount by which enjoyment falls short of bliss integrated throughout time:

$$\min \int_0^\infty [B - U(C) + V(L)]\mathrm{d}t \qquad (1)$$

subject to

$$\frac{\mathrm{d}K}{\mathrm{d}t} + C = F(K, L). \qquad (2)$$

Ramsey attacks the problem from two directions: economic and mathematical. His economic argument first solves for the relationship between consumption and the effort by equating the marginal disutility of labour to the product of the marginal product of labour and the marginal disutility of consumption. He then solves the basic arbitrage relationship equating the marginal utility of consuming a unit now with the marginal utility of consuming the product of investing the unit until the next instant of time. This key relationship implies that the marginal utility of consumption, $U'(C)$, must fall at the rate of interest, equal to the marginal product of capital, $\partial F/\partial K$. These two conditions, together with (2), the initial stock of capital, and a terminal condition as $t \to \infty$, produce a differential equation which can be integrated to give the result.

The mathematical approach observes that the calculus of variations gives the first two conditions directly, but also observes that the variable of integration in (1) can be changed from t to K by using (2) to give

$$\min \int_{K_0}^{\infty} \frac{B - U(C) + V(L)}{F(K,L) - C} \mathrm{d}K \tag{3}$$

and since C and L are arbitrary functions of K all that is needed to minimize the integrand is to set its partial derivations to zero. Differentiating with respect to C gives

$$F(K,L) - C = \frac{B - [U(C) - V(L)]}{U'(C)}. \tag{4}$$

The left-hand side of (4) is the rate of investment or saving, while the right-hand side is equal to bliss minus the additional rate of enjoyment, divided by the marginal utility of consumption, and the whole is the Keynes-Ramsey rule.

Ramsey concluded from this rule that the optimal rate of saving should be ‘greatly in excess of that which anyone would normally suggest’ and gave an illustration in which the savings rate should be 60 per cent of income. One of the main themes explored by later writers was whether this was a robust conclusion, or whether the optimal rate of saving was very sensitive to the simplifying assumptions – a theme which is discussed below. Ramsey recognized that discounting utility would destroy the simple reasoning which led to (4), and was thus anxious to have an ethical reason for rejecting it. He believed that population growth would argue for higher rates of saving whilst technical progress would have ambiguous effects – as proved to be the case in later formal models.

Ramsey drew attention to two remarkable features of the rule. The first is that the level of saving does not depend on the production function. The second is that it does not depend on the rate of interest, unless this is actually zero. In fact, the first feature is only apparently the case, for in (4), C will depend on the level of output, F, and since savings, given by the right-hand side, also depends on C, it will depend on F. In his Section III, Ramsey clearly pointed out that the level of saving was motivated by the demand for future consumption, while the rate of interest was determined by the current stock of capital (in this one-sector model). In a concluding remark to this section he notes that ‘in the accounting of a Socialist State the function of the rate of interest would be to ensure the wisest use of existing capital, not to serve in any direct way as a guide to the proportion of income which should be saved’. The second result does not survive in more general models which allow for utility discounting. Nevertheless, the arbitrage relationship does suggest a way in which the rate of interest can guide the rate of saving. If the rate of decline of the marginal utility of consumption is less than the rate of interest, taken to be the rate of return on investment, then the rate of saving is too low, and vice versa.

The main contribution of the paper was to pose a fruitful question – what should the rate of savings be – and propose a method of analysis – that of intertemporal welfare maximization using the techniques of dynamic optimization, in this case the calculus of variations. The main result was striking – the rate of saving should apparently be rather high. In addition to this contribution, the paper also contained various remarkable extensions. It considers the choice of savings rate for an individual facing constant factor prices, who wishes to optimize his lifetime consumption pattern, and as such provides a positive theory of life-cycle savings. It shows that if utility is to be discounted, then it must be discounted at a constant rate if one is to escape the contradiction 'that successive generations are motivated by the same system of preferences'. Later, Strotz (1956) would return to this issue and the related problem of dynamic consistency. Finally, Ramsey shows that if a society consists of individuals who differ in their rate of discount, and if it is in steady state, then the equilibrium would be attained by a division of society into two classes, the thrifty enjoying bliss and the improvident at the subsistence level. In short, he characterizes the long-run general equilibrium of a society of heterogeneous individuals.

Ramsey thus laid the foundations for the study of optimal accumulation and optimal growth, as well as the positive theory of savings and the rate of interest. Space precludes a full assessment of the subsequent work his paper stimulated, though Burmeister and Dobell's (1970) textbook lists 107 references in their chapter on optimal economic growth, and much has happened since that date. Instead we shall briefly mention some of the themes of this subsequent work.

Ramsey's model represented a significant advance on the classical analysis of stationary states, since it made possible the analysis of non-stationary time paths of capital accumulation, but ultimately his model would tend towards a stationary state. With the development of growth theory the profession acquired a more appealing concept of long-run equilibrium – that of steady growth. In due course this suggested the obvious extension to Ramsey's model of incorporating these dynamic features – population growth at the steady rate n and Harrod-neutral technical progress at a steady rate g. The instantaneous level of national welfare was variously taken as $U(C_t)$, $U(C_t/L_t)$, or, most satisfactorily, $L_t u(C_t/L_t)$, where L_t was the total population or workforce, and C_t was total consumption. Since welfare now depended on time, it made no drastic difference to include a utility discount rate, δ, and to propose a more general objective such as

$$\begin{aligned} W &= \int_0^\infty U(C_t, t)\mathrm{d}t \\ &= \int_0^\infty L_t u(C_t/L_t)\, e^{-\delta t}\mathrm{d}t. \end{aligned} \tag{5}$$

Steady growth now raised the question of the *existence* of an optimal savings policy in an acute form, for the integral in (5) might diverge unless δ was sufficiently large. Ramsey had faced a similar problem and avoided it by minimizing the shortfall from a reference path (or bliss). Similar devices were invoked to deal with divergent integrals, and much effort was expended on devising criteria of optimality and categorizing conditions under which an optimal savings plan existed, though many apparently reasonable problems nevertheless failed to possess an optimal savings plan, as Hammond and Mirrlees (1973) demonstrated. (They also give references to earlier discussions of the problem of non-existence.) They observe that no restrictions on the class of utility function will ensure existence, nor will any realistic restrictions on the production assumptions by themselves be enough to avoid the problem. They then argue that if we could specify date after which events are of no significance, then the problem reduces to a finite horizon model, for which the utility integral would converge. Different people might disagree on the horizon date, but if the initial T_0 years of the plan were relatively insensitive to any horizon date later than some date T_1, then everyone would agree with the T year plan, and, in their language, the plan would be *agreeable.* Hammond and Mirrlees show that in the one good model with a general instantaneous utility function $U(C_t, t)$ the agreeable path is unique and locally optimal, and that if an optimal path exists it is agreeable. Establishing the existence of agreeable paths is, however, considerably easier than establishing the existence of optimal paths.

While existence problems are important and raise intriguing philosophical problems (what if optimal growth paths do not exist?), they are not central to the economics of the problem. One of the key issues that has engaged the attention of subsequent researchers is whether the optimal savings rate is indeed as high as Ramsey argued (though, as Samuelson 1969, pointed out, Ramsey's conclusion depended on a particular choice of utility function). Certainly, Tinbergen (1956) was inclined to agree, but Mirrlees (1967) argued that Ramsey's model was seriously misleading, and that once population growth, technical progress and utility discounting were admitted, the initial value of the optimal rate of saving was typically very different from that implied by the Keynes-Ramsey rule. Once time enters the production function, it is no longer possible to obtain explicit solutions and an alternative solution strategy is required. Mirrlees argued that it was preferable to find the asymptotic form of the optimally developing economy in which output, consumption and consumption per head all grow at steady rates along a 'modified Golden Rule', and in which the savings rate is constant. The initial value of the savings rate could then be estimated by expanding around this asymptotic solution.

Mirrlees, in common with a large number of other optimal growth theorists, used a particular utility function – the iso-elastic form

$$u(c) = -c^{1-v}, v > 1, \; = \log c, v = 1 \qquad (6)$$

for which Ramsey's rule gives a savings rate of $1/v$ (providing an optimum exists). Mirrlees was impressed that for plausible values of the parameters of his model, the optimum savings rate was very different from the Ramsey value, and might be quite low. He also pointed out that the asymptotic solution, or the 'modified Golden Rule', would differ from the Golden Rule (according to which the rate of savings should equal the share of profit), if utilities were discounted – for the obvious reason that one would expect optimum policies to reflect the values regarding the distribution between generations.

Ramsey's model made skilful use of the classical calculus of variations, and in that vein Samuelson and Solow (1956) extended the model to deal with heterogeneous capital goods. In so doing they made possible two notable contributions to capital theory. The first was to argue that on the optimum path it was not too misleading to think in terms of an abstract quantity of capital – heterogeneity did not significantly alter the Ramsey theory. Second, the Hahn-Samuelson problem of the indeterminacy of equilibrium with capital heterogeneity disappeared on the optimal path, though the significance of this did not emerge until the paper by Hahn (1966).

As Samuelson and Solow pointed out, the classical calculus of variations could be replaced by Hamiltonian methods which would be able to deal with inequality constraints. The powerful techniques of the Pontryagin Maximum Principle and Bellman's Dynamic Programming were in due course applied to various extensions of the Ramsey problem to good effect, and their advantages and interrelationships are well discussed in the textbook of Intriligator (1971). In both approaches shadow prices or co-state variables play an important role both in characterizing the solution and in demonstrating the relationships between optimality, intertemporal efficiency, and a set of intertemporal (shadow) prices (prices on futures markets) which might be used to decentralize the optimum. These shadow prices have a natural interpretation, for they value the capital stock in terms of the objective function, that is social welfare or the utility of consumption. The price guides the instantaneous allocation of output between consumption and investment, for consumption should be increased, if possible, until its value (the marginal utility of consumption) falls to the value of investment, that is, of the capital stock. The evolution of the price over time then satisfies the fundamental arbitrage relationship, so that asset holders obtain a return (including capital gains) on the asset equal to the return on other assets and to the return from delaying consumption.

The strengths of these alternative approaches are best appreciated in multisector models when there are constraints on reallocating resources. If investment goods are physically different from consumption goods, and capital is immobile between sectors, then savings will be constrained by the feasible output of the investment goods sector, and the planners' problem is primarily one of choosing the allocation of investment between the two sectors. In such a model the rate of return on capital will depend on the level of investment, and Ramsey's observation that in his model the two are independent is shown to be a feature of the one-good assumption. With two sectors corner solutions are quite likely (in the early stages) and the inequality constraints require the extra power of the new approaches.

The shadow prices are arguably most useful for cost–benefit analysis, rather than the more ambitious planning problems which so engaged the attentions of optimal growth theorists in the 1960s. Little and Mirrlees (1969, 1974), Newbery (1972), and Stern (1972) were concerned to develop methods for calculating shadow or accounting prices in dual economy models of developing countries in which the level of aggregate savings was constrained. The two key accounting prices on which optimal growth models can shed light are the wage rate and the rate of discount to use in investment projects. The former emerges from the constraints on the allocation of labour and on the level of wages which must be paid, while the latter is again given by an arbitrage relation, or the rate of change of the shadow price of capital itself. The arbitrage equation gives a differential equation for the shadow price which, together with the equation for saving and the accumulation of capital, can be numerically integrated backwards from the asymptotic solution. Modern computers allow this to be done quickly, as illustrated in Newbery (1972).

The arbitrage equation comes into its own in exhaustible resource models where the return to the exhaustible resource must, while it remains in the ground, take the form of a capital gain equal to the return on other assets. This rule, due originally to Hotelling (1931), and nicely exposited by Solow (1974), has achieved prominence since the dramatic rise in the oil price of 1973–4.

Although the revival of interest in the Ramsey model in the 1960s was initially motivated by the post-war popularity of national economic planning, a popularity which waned rapidly in the 1970s, the model and its successors remain useful for the more modest aims of characterizing intertemporal competitive equilibrium in asset markets, especially for exhaustible resources like oil and gas, and for providing a more satisfactory neoclassical theory of equilibrium growth with individually rational savers. The common feature of Ramsey's two contributions to economics was that they were normative, and postulated an additive (utilitarian) social welfare function as the objective to be maximized. Several writers have taken the natural step of combining both of Ramsey's two interests and enquiring what optimal tax (and monetary) policy should be in an intertemporal model in which savings and investment are affected by these policies. Arrow and Kurz (1969) were the first to explore these issues and the closely related issues of the problem of public investment criteria systematically in a growth model in which full optimality is not achieved.

Diamond (1973) extended their work to a model with many goods, and demonstrated the desirability (under constant returns) of equal efficiency, on average, between public and private production, even though aggregate efficiency was not desired. In particular the public and private sectors should use the same discount rates. Later work (surveyed, for example, by Kotlikoff 1984) has explored the efficiency losses involved in an economy of intertemporal optimizing individuals in the presence of distortionary taxes on capital, and have used these estimates to rank alternative capital tax reform programmes – a compromise between the optimal tax approach of Diamond and the need to incorporate more of the complex features of particular economies.

In short, if the central question which Ramsey addressed of the right level of saving and investment has fallen from favour recently, nevertheless the spirit of the Ramsey model with its emphasis on intertemporal optimization lives on strongly, whether it be in the study of the oil market, the derivation of public investment rules, or the reform of the corporate tax system.

See Also

▶ Calculus of Variations
▶ Neoclassical Growth Theory

Bibliography

Arrow, K., and M. Kurz. 1969. Optimal public investment policy and controllability with fixed private savings ratio. *Journal of Economic Theory* 1: 141–177.

Burmeister, E., and A.R. Dobell. 1970. *Mathematical theories of economic growth*. New York: Macmillan.

Chakravarty, S. 1969. *Capital and development planning*. Cambridge, MA: MIT Press.

Diamond, P.A. 1973. Taxation and public production in a growth setting. Ch. 10 in Mirrlees and Stern (1973).

Hahn, F.H. 1966. Equilibrium dynamics with heterogeneous capital goods. *Quarterly Journal of Economics* 80: 633–646.

Hammond, P.J., and Mirrlees, J.A. 1973. Agreeable plans. Ch. 13 in Mirrlees and Stern (1973).

Hotelling, H. 1931. The economics of exhaustible resources. *Journal of Political Economy* 39: 137–175.

Intriligator, M.D. 1971. *Mathematical optimization and economic theory*. Englewood Cliffs: Prentice-Hall.

Kotlikoff, L.J. 1984. Taxation and savings: a neoclassical perspective. *Journal of Economic Literature* 22: 1576–1629.

Little, I.M.D., and J.A. Mirrlees. 1969. *Manual of industrial project analysis for developing countries*, Social Cost Benefit Analysis. Vol. 2. Paris: OECD Development Centre.

Little, I.M.D., and J.A. Mirrlees. 1974. *Project appraisal and planning for developing countries*. London: Heinemann.

Mirrlees, J.A. 1967. Optimum growth when technology is changing. *Review of Economic Studies* 34: 95–124.

Mirrlees, J.A., and N.H. Stem, eds. 1973. *Models of economic growth*. London: Macmillan.

Newbery, D.M.G. 1972. Public policy in the dual economy. *Economic Journal* 82: 567–590.

Ramsey, F.P. 1927. A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Ramsey, F.P. 1928. A mathematical theory of saving. *Economic Journal* 38: 543–559.

Samuelson, P.A. 1969. Foreword to Chakravarty (1969).

Samuelson, P.A., and R. Solow. 1956. A complete capital model involving heterogeneous capital goods. *Quarterly Journal of Economics* 70: 537–562.

Solow, R.M. 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70: 65–94.

Solow, R.M. 1974. The economics of resources or the resources of economics. *American Economic Review* 64 (2): 1–14.

Stern, N.H. 1972. Optimum development in a dual economy. *Review of Economic Studies* 39 (2): 171–184.

Strotz, R.H. 1956. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23 (3): 165–180.

Tinbergen, J. 1956. The optimal rate of savings. *Economic Journal* 66: 603–609.

Ramsey Pricing

William J. Baumol

Keywords

Contestable markets; Hyperplanes; Inverse elasticity; Lump-sum taxes; Marginal cost pricing; Optimal taxation; Pareto efficiency; Ramsey pricing; Ramsey, F. R.; Second best

JEL Classifications

D0

Ramsey prices are prices that are Pareto optimal subject to a constraint on the total profits of a single supplier or group of suppliers. In particular, because a firm whose activities are characterized by scale economies will lose money if it sets the prices of its products equal to their marginal costs, Ramsey prices become for that firm the prices that are optimal (economically efficient) given the financial feasibility requirement that the firm's profits be non-negative. The same Ramsey prices can also be shown to be those necessary for maximization of the sum of consumers' and producers' surpluses.

The concept is named after Frank Ramsey, its discover, whose 1927 paper on the subject was one of several revolutionary contributions to economics, mathematics and philosophy this extraordinary man made before he died at the age of 26. Since then and until the 1970s, the principle was largely forgotten even though it was rediscovered and expanded upon by Pigou, Boiteux and Samuelson. In 1970 it was publicized and its history explored in an article by Baumol and Bradford, and the principle has since been widely recognized and accepted by economists and practitioners. As an illustration, in 1983 the Interstate Commerce Commission adopted Ramsey pricing as the underlying principle it would follow in the regulation of railroad rates.

Ramsey prices are an outstanding example of the use of pure economic theory to derive an operational solution to a difficult set of practical problems. It may also be as definitive as any available second-best theorem. The extraordinary achievement of the theorem lies in the very explicit formulae it is able to derive from so weak a premise – the Pareto optimality requirement that the prices be those which elicit such a set of outputs and purchase quantities that it is impossible to increase the welfare of any one individual without harming anyone else. Aside from the apparent weakness of this assumption, the definitive character of the Ramsey theorem is surprising in light of the conclusion suggested by much of the second-best literature, that where additional constraints are superimposed on the usual requirements of optimality, one can expect no simple and straightforward results to emerge.

The Ramsey Theorem and its Interpretation

The Ramsey theorem is expressed in a variety of formulae all of which are essentially equivalent. Perhaps its simplest form asserts that when a producer supplies n commodities then Pareto optimality subject to a profit constraint requires the prices, p_j of these goods to satisfy

$$\frac{p_j - mc_j}{p_n - mc_n} = \frac{mr_j - mc_j}{mr_n - mc_n}, (j = 1, \cdots, n-1),$$
$$\sum_{j=1}^{n} p_j y_j = c(y_1, \cdots, y_n) + k \tag{1}$$

where mc_j and mr_j are, respectively, the marginal cost and marginal revenue of output j, $c(\cdot)$ is the supplier's total cost function and k is any constant.

In the special case in which none of the seller's goods is either a complement or a substitute in demand, the preceding relationship is easily shown to take the special form which is widely known as 'the inverse elasticity formula':

$$\frac{(p_j - mc_j)/p_j}{(p_n - mc_n)/p_n} = \frac{E_n}{E_j}, (j = 1, \cdots, n-1),$$
$$\sum p_j y_j = c(\cdot) + k, \tag{2}$$

where E_j is the price elasticity of demand for product j.

In the particular case where an optimum satisfies locally the requirements of constant returns to scale, so that marginal cost pricing yields zero economic profits exactly, then (for $k = 0$) conditions (1) and (2) are automatically transformed into the marginal cost pricing conditions

$$p_j = mc_j, (j = 1, \ldots, n). \tag{3}$$

It is easy to show that no prices can be Pareto optimal subject to the profit constraint indicated unless they satisfy (1). Moreover, as long as the proper concavity-convexity conditions hold, any prices which satisfy (1) will be consistent with Pareto optimality so constrained.

One can suggest in rough intuitive terms why constrained Pareto optimality requires prices which satisfy (1), or (2) – in the case of demand independence. The latter is perhaps the most illuminating case, and so it is useful to summarize the argument briefly.

As a starting point, one should recall that the reason marginal cost pricing is necessary for a 'first best' (unconstrained) optimum is that such prices equate the pecuniary cost to the consumer of purchasing an additional unit of the item and the economic cost of producing it, that is, its marginal cost . Thus, when the consumer selects his purchases so as to maximize the utility he derives from a given outlay of money, he thereby automatically maximizes the utility derivable from a bundle of economic resources.

However, where returns to scale are not constant at the vector of purchases elicited by the prices $p_i = mc_i$ then the requirement $\Sigma p_i y_i = c(\cdot)$ will be violated by those marginal cost prices. Consequently, prices will have to deviate from marginal costs in some pattern that satisfies the profit constraint. Of course, every such deviation will affect consumer purchases, and so the quantities produced, making them depart in different degrees from the optimal quantities that would have been selected under marginal cost pricing. The objective is to cause the p_i to deviate from the mc_i in a manner that satisfies the profit constraint and yet distorts consumer purchases from their optimal levels as little as possible.

For this purpose, consider two of the pertinent commodities, i and j, with i's demand highly elastic and j's very inelastic. Start with $p_i = mc_i$ and $p_j = mc_j$ and assume that at those prices profits are negative. Because of the high demand elasticity of i a small rise in p_i above mc_i will cause a relatively large 'distortion' in consumer demand from its Pareto optimal quantity. Moreover, also because of the high elasticity, the rise in p_i will yield a relatively small increase in revenue to help eliminate losses. In contrast, a similar percentage increase in p_j will cause a smaller percentage change in quantity of j demanded and a larger gain in revenue. Clearly less damage will be done to welfare if a larger share of the task of meeting the shortfall of total revenue relative to total cost is carried out via a rise in p_j, the price of the commodity with the more inelastic demand. This is, in essence, the logic of the inverse elasticity formula.

Informal Derivation of the Theorem

A simplified and rather informal derivation of the formulae is straightforward. For brevity only a single consumer and a single input, labour, is used in the following, but the proofs in the k consumer – m input cases are virtually identical. Let

y_i = the supplier's output of i ($i = 1,\ldots, n$)

x = the vector of outputs of all other goods

R = the available quantity of resource

r = unused resource (leisure)

p_i = the price of i

w = wage (price of leisure)

$U(y_1,\ldots, y_n, x, r)$ = the consumer's utility function

$C(y_1,\ldots, y_n)$ = the firm's input requirement function

$K(x)$ = the input requirement for production of x. Then, optimality requires maximization of

$$U(y_1, \ldots, y_n, x, r)$$

subject to the resource constraint

$$C(y_1,\ldots,y_n) + K(x) + r = R$$

and the budget constraint

$$\sum p_j y_i = wC(y_1, \ldots, y_n).$$

This yields the Lagrangian

$$\begin{aligned} L = U(\cdot) + \alpha[R - C(\cdot) - K(\cdot) - r] \\ + \beta\left[\sum p_i y_i - wC(\cdot)\right] \end{aligned}$$

Using the notation U_i for $\partial U/\partial y_i$, $U_r = \partial U/\partial r$, and so on, we have the first order conditions

$$U_i - \alpha C_i + \beta(mr_i - wC_i) = 0, \qquad (4)$$

where $mr_i = \partial \sum p_i y_i / \partial y_i$ is the marginal revenue of i, and

$$U_r - \alpha = 0. \qquad (5)$$

Since consumer equilibrium requires equality between price ratios and marginal rates of substitution (the ratios of marginal utilities) we have

$$\frac{U_i}{p_i} = \frac{U_r}{w} = k, (i = 1, \ldots, n)$$

so that (5) yields $U_i = kw = \alpha$, and therefore (4) becomes

$$p_i - wC_i + (\beta/k)(mr_i - wC_i) = 0, \qquad (6)$$

which, writing $mc_i = wC_i$, yields the general Ramsey formula (1). To obtain the inverse elasticity formula (2) we simply use a standard relationship for the case of independent demands,

$$mr_i = p_i(1 - 1)/E_i),$$

substituting this into (6) we have

$$p_i - mc_i = (\beta/k)(p_i - mc_i - p_i/E_i),$$

or

$$(1 - \beta/k)(p_i - mc_i) = -(\beta/k)p_i/E_i$$

which immediately yields (2).

Applications

Aside from its obvious connection with pricing by the firm, the theorem also has applications to the principles of taxation and to the general equilibrium analysis of the economy. Indeed, Frank Ramsey presented his result as a theorem on taxation rather than pricing. The point is that, the theoretical concept of lump-sum taxes aside, any tax must be a levy on some sort of economic activity. Even if the price of that activity's product is equal to its marginal cost, the tax will in general drive a wedge between the two, particularly if the total tax revenue is required to meet some particular target. Thus the problem of determining the vector of tax rates on the economy's activities that will meet the overall revenue target with minimum social welfare loss is equivalent to determining the optimal vector of deviations between prices and marginal costs that will satisfy that revenue (budget) constraint. In sum, the search for the optimal (budget constrained) prices and the optimal (revenue constrained) tax rates are formally equivalent.

The analysis also has direct implications for general equilibrium theory, for it tells us that if lump-sum taxes are impossible, then a vector of (first-best) marginal cost prices may also be ruled out for the economy as a whole. Indeed, such a first-best parametric price solution is possible only if, at the corresponding vector of activity levels (outputs), the production frontier happens to be locally linear and homogeneous (meaning, in the differentiable case, that it must be tangent to a hyperplane through the origin in input–output space).

For suppose this is not so – say, that there are increasing returns to scale at any such point. Then marginal cost pricing will yield negative profits for the economy, and suppliers as a class will be able to survive financially with such prices only if they receive subsidies. But subsidies must be paid for by taxes, and any such taxes on activities whose pre-tax prices equal their marginal costs must yield after-tax prices which do not. In sum, one cannot escape the problem of finding the deviations of prices from their 'first best' magnitudes which meet the budget requirement that every subsidy must be covered by tax revenues. This, then, is the inescapable Ramsey problem for the entire economy if prices are parametric and no optimal output vector is a point of (at least local) linear homogeneity.

The case of diminishing returns poses corresponding problems, even though it is often thought to be compatible with competitive equilibrium and marginal cost pricing. As long as input quantities (including the input of entrepreneurship) can be expanded, marginal cost pricing will be incompatible with equilibrium at an optimal point because marginal cost pricing will then yield positive economic profits and the number of firms will therefore increase. There can be no finite number of firms at which this manifestation of disequilibrium ceases unless marginal cost pricing is abandoned. But then the best equilibrium prices in terms of Pareto optimality must again be the Ramsey prices.

In sum, Ramsey pricing is no mere artifact of regulation of industry or tax policy. It is deeply embedded in the logic of the general equilibrium mechanism.

History of Ramsey Analysis

The basic theorem apparently first appeared in Frank Ramsey's classic article (1927). While the article has sometimes attracted the attention it deserved, it did not effectively convey to the profession the wider implications of its second-best pricing analysis. In 1928 A.C. Pigou, who had apparently posed the original issue to Ramsey, published a restatement of the theorem. Here, too, it was presented as a result on the principles of taxation and not related to pricing. Ursula Hicks (1947) independently provided a similar discussion.

Perhaps the first work on Ramsey theory that was expressed in terms of pricing issues occurred in the aftermath of Hotelling's (1938) classic paper on marginal cost pricing. There the author had advocated a system of subsidies to firms subject to scale economies, but he himself came to recognize the tax implications and the consequences for the overall optimality of the solution. He and J.R. Hicks discussed the problem, and Hicks emerged with an independently discovered Ramsey theorem, which was never published.

Early after the Second World War, two major contributions were made to the literature. Paul

Samuelson (1951) prepared a memorandum for the US Treasury pointing out the logic of the Ramsey approach to taxation. As is to be expected, Samuelson's contribution was highly sophisticated and offered substantial original insights, but, although widely circulated in public finance circles, it was never published. After having published a less sophisticated version of the theorem in 1951, Marcel Boiteux, Directeur-Général of Électricité de France 1967–87, published a major article on the subject in 1956. It explicitly dealt with the topic as an issue in pricing policy for nationalized or regulated firms and derived its results directly from a Pareto optimality model. Moreover, it provided a result more general than the inverse elasticity form of the theorem on which Ramsey and Pigou had focused.

An even deeper exploration of the subject was provided by Diamond and Mirrlees (1971) as part of their continuing work on the theory of optimal taxation. Their papers are important not only because of their careful analysis but also because they played a major role in bringing the subject to the profession's attention. Within a year or two of the appearance of their articles and that of Baumol and Bradford (1970), 'everyone' in the profession was fully aware of the notion of Ramsey pricing and its logic. Since then there has been an explosion of writings on the subject and it occurs centrally or peripherally in a wide variety of fields.

An illustrative and perhaps surprising application which suggests the unexpected places in which the construct can turn up, is the 'weak invisible hand theorem', that occurs in the contestable markets literature (see Baumol et al. 1977). That theorem states that if a monopolist who is constrained by a regulatory (or other) profit ceiling chooses to adopt the Ramsey price vector rather than some other set of prices that enable him to earn his allowed return, then under a fairly attractive set of assumptions the monopolist will be rewarded for his virtuous decision by being protected from entry by those prices. In other words, self-interest may impel a monopolist to adopt Ramsey prices because those prices are *sustainable* against entry, meaning that at those prices the monopolist will earn the profits that the constraint allows to him, but any rival firm that undertakes to enter the field will be predestined to lose money even if the incumbent undertakes no strategic (retaliatory) response.

Today Ramsey pricing is accepted as a basic proposition of microanalysis and appears with great frequency in new writings on the theory of the firm, industrial organization and public finance; it recurs regularly in the pricing discussions of American regulatory agencies.

See Also

► Optimal Taxation

Bibliography

Baumol, W.J., and D.F. Bradford. 1970. Optimal departures from marginal cost pricing. *American Economic Review* 60: 265–283.

Baumol, W.J., E.E. Bailey, and R.D. Willig. 1977. Weak invisible hand theorems on the sustainability of prices in a multiproduct monopoly. *American Economic Review* 67: 350–365.

Boiteux, M. 1956. Sur la géstion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica* 24: 22–40.

Diamond, P.A., and J.A. Mirrlees. 1971. Optimal taxation and public production: II. *American Economic Review* 61: 261–278.

Hicks, U. 1947. *Public finance*. London: Nisbet.

Hotelling, H. 1938. The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica* 6: 242–269.

Pigou, A.C. 1928. *A study in public finance*. London: Macmillan.

Ramsey, F. 1927. A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Samuelson, P.A. 1951. *Theory of optimal taxation*. Unpublished memorandum for the US Treasury.

Ramsey, Frank Plumpton (1903–1930)

Peter Newman

Abstract

There are interesting parallels in the careers of Frank Ramsey and John von Neumann. Each was born in 1903, one the product of the 'High

Intelligentsia of England' (Keynes 1933, p. vii) and the other the son of a wealthy banker in Budapest (Ulam 1976, p. 79). Each was a creative mathematician of high order but each also made major contributions to at least two other disciplines. Each wrote just three papers in economic theory, all six of which were of fundamental importance. Moreover, with one exception every one of these seminal papers had to wait many years for its proper recognition; even the exception – the utility theory set out in the Appendix to von Neumann and Morgenstern (1947) – at first encountered serious misunderstanding within the profession. Indeed, considering them purely as economists, one wonders how these two geniuses would fare today, when promotion and tenure so often depend on a good immediate showing in citation indexes and the like.

Keywords

Bayes, T.; Bernoulli, D.; Bootstrap; Choice under uncertainty; De Finetti, B.; Dutch books; Growth theory; Optimal taxation; Probability; Ramsey, F.P.; Sraffa, P.; Subjective probability; von Neumann, J.; von Neumann–Morgenstern; Wittgenstein, L.

JEL Classifications

B31

There are interesting parallels in the careers of Frank Ramsey and John von Neumann. Each was born in 1903, one the product of the 'High Intelligentsia of England' (Keynes 1933, p. vii) and the other the son of a wealthy banker in Budapest (Ulam 1976, p. 79). Each was a creative mathematician of high order but each also made major contributions to at least two other disciplines. Each wrote just three papers in economic theory, all six of which were of fundamental importance. Moreover, with one exception every one of these seminal papers had to wait many years for its proper recognition; even the exception – the utility theory set out in the Appendix to von Neumann and Morgenstern (1947) – at first encountered serious misunderstanding within the profession. Indeed, considering them purely as economists, one wonders how these two geniuses would fare today, when promotion and tenure so often depend on a good immediate showing in citation indexes and the like.

The three papers of Ramsey are in subjective probability and utility (1926), optimal taxation (1927), and optimal one-sector growth (1928), while those of von Neumann are in game theory (1928), optimal multi-sector growth (1937, 1945–6), and objective probability and utility (1947). It is quite striking that their work both on growth theory and on choice under uncertainty should be so complementary, especially since there is no evidence that von Neumann knew of Ramsey's work in either field.

Another and grievous similarity was that both men died early, Ramsey on 19 January 1930 of complications associated with jaundice, and von Neumann (twice Ramsey's age) on 8 February 1957 of cancer. Both losses were tragic, especially that of the 26-year-old Frank Ramsey, whose 'death at the height of his powers deprives Cambridge of one of its intellectual glories and contemporary philosophy of one of its profoundest thinkers' (Braithwaite's Introduction to Ramsey 1931, p. ix).

Life

Frank Plumpton Ramsey was born in Cambridge on 22 February 1903. His father was a mathematician, Fellow and later President of Magdalene College (Harrod 1951, pp. 141, 320), and his brother Michael became Archbishop of Canterbury. He was educated at Winchester and at Trinity College Cambridge, and was a Scholar of both those ancient foundations. In the autumn of 1924 he became Fellow of King's College and University Lecturer in Mathematics and soon afterwards married Lettice Baker, who had been a student in the Moral Sciences Tripos. After his death she became a founder of Ramsey and Muspratt, a firm of portrait photographers that has long been an Oxbridge institution. She survived into the 1980s, in vigorous old age.

In physical appearance Ramsey was tall and portly, the latter a feature he shared with von

Neumann; 'I take no credit for weighing nearly 17 stone [238 pounds]' (1931, p. 291). All accounts agree as to his simplicity and modesty, qualities which are happily reflected in his engaging literary style. 'Ramsey reminds one of Hume more than of anyone else, particularly in his common sense and a sort of hard-headed practicality towards the whole business' (Keynes 1933, p. 301). But his unfailing cheerfulness did not disguise 'the amazing, easy efficiency of the intellectual machine which ground away behind his wide temples and broad, smiling face' (Keynes 1933, p. 296). 'He comes down to earth, however, with a satisfying bump, and earth is certainly the natural element of my old friend Lettice' (Partridge 1981, p. 129).

Ramsey and Wittgenstein

For many years it was thought that while still an undergraduate Ramsey assisted in the translation of the German text of Wittgenstein's *Tractatus Logico-Philosophicus* (1922). It now appears that 'the first draft of the translation was produced by F.P. Ramsey alone' (von Wright 1982, p. 102). Just 19, he dictated it directly to a stenographer in the University Typing Office in Cambridge in the winter of 1921–2 (reminiscent, on a smaller scale, of the 19-year-old 'John S. Mill' beginning in 1825 to edit Bentham's massive Rationale of Judicial Evidence). Wittgenstein seems to have been pleased with Ramsey's translation (1973, p. 77), and a fast friendship was thereby established between the two philosophers that lasted for the rest of Ramsey's short life.

In September 1923 the Tractatus had been published for almost a year. Not only had Ramsey been its main translator but he had also written a long and penetrating review of it for Mind (reprinted in 1931, pp. 270–86). But still there were many passages which remained unclear to him. To remedy this he made a special journey to Austria, where Wittgenstein was teaching in the local school of a small village and living in spartan conditions. The eccentric philosopher and the brilliant undergraduate hit it off immediately. Ramsey stayed two weeks, spending every afternoon from 2 to 7 elucidating the great man's work: 'we get on about a page an hour' (Wittgenstein 1973, p. 79).

In the several letters that Ramsey afterwards wrote to Wittgenstein we can glimpse what Keynes meant in referring to 'the simplicity of his feelings and reactions, half-alarming sometimes and occasionally almost cruel in their directness and literalness' (Keynes 1933, p. 296). Consider for example these passages from his letters of 12 November and 20 December 1923 (Wittgenstein 1973, pp. 81–3):

> I have not been doing much towards reconstructing mathematics; partly because I have been reading miscellaneous things, a little Relativity and a little Kant, and Frege . . . But I am awfully idle; and most of my energy has been absorbed since January by an unhappy passion for a married woman, which produced such psychological disorder, that I nearly resorted to psychoanalysis, and should probably have gone at Christmas to live in Vienna for nine months and be analysed, had not I suddenly got better a fortnight ago, since when I have been happy and done a fair amount of work.
>
> I think I have solved all problems about finite integers, except such as are connected with the axiom of infinity, but I may well be wrong.
>
> [December 20th] I was silly to think I had solved those problems. I'm always doing that and finding it a mare's nest . . . *I have been trying to prove a proposition in the Mengenlehre* either $2^{\aleph_0} = \aleph_1$ *or* $2^{\aleph_0} \neq \aleph_1$, which it is no one knows but I have had no success. (His italics)

In 1924 Ramsey actually did spend six months in Vienna in psychoanalysis (rarer then than now), after which 'I feel that people know far less about themselves than they imagine, and am not nearly so anxious to talk about myself as I used to be, having had enough of it to get bored' (1931, p. 290). The mathematical problem referred to in his second letter was of course the famous Continuum Hypothesis. His lack of success in this is scarcely surprising, since in the 1960s Paul Cohen showed the Hypothesis to be an undecidable proposition within Zermelo–Fraenkel set theory (see, for example, Cohen 1966). It was, incidentally, a continual disappointment to von Neumann that it was not he but his hero Kurt Gödel who made the startling discovery, in 1930–31, of the necessary existence of such undecidable propositions (Ulam 1976, pp. 76, 80).

Wittgenstein returned to Cambridge early in 1929 and began those ‘innumerable conversations’ with Ramsey that are acknowledged in the Preface/Foreword (dated January 1945) to his *Philosophical Investigations* (1953, p. x). Unfortunately, these were cut short by Ramsey’s tragic death, a moving account of which may be found in Frances Partridge’s *Memories* (1981, pp. 169–82); the grieving Wittgenstein was at Ramsey’s bedside in the hospital until a few hours before he died.

The only other person acknowledged by name in the Preface to the *Investigations*, and for even greater help than Ramsey gave, was Piero Sraffa. The trio of Ramsey, Sraffa and Wittgenstein must have been a formidable discussion group indeed; a treasured piece of Cambridge folklore is a lunch at which the three of them discussed Keynes’s theory of probability with its author. The odd pattern of belated recognition of intellectual indebtedness was continued in Sraffa’s acknowledgement (1960, pp. vi–vii) of Ramsey’s help, a mere 30 years after the fact.

Works

Ramsey’s early work in philosophy was a continuation of the methods of Russell and Whitehead’s *Principia*, but it is clear that the influence of Wittgenstein and the evolution of his own thinking were moving him towards the end of his life in a quite different, more pragmatic direction. These later contributions were left fragmentary and incomplete at his death, but a very brief account of them and their relations to modern philosophy may be gleaned from the first two Introductions to the revised edition (1978) of (1931).

In mathematics proper, as distinct from the foundations of mathematics, his main contribution is a fundamental theorem which appeared actually as a byproduct of a paper of 1928 on formal logic (reprinted in 1931, pp. 82–111). It reads (1931, p. 82):

Theorem A Let Γ be an infinite class, and and r positive integers; and let all those sub-classes of Γ which have exactly r members, or, as we may say, let all r- combinations of the members of Γ be divided into μ mutually exclusive classes $C_i (i = 1, 2, \ldots, \mu)$, so that every r-combination is a member of one and only one C_i; then, assuming the Axiom of Selections [i.e. the Axiom of Choice], Γ must contain an infinite sub-class Δ such that all the r-combinations of the members of Δ belong to the same C_i.

This beautiful result was ignored until 1935, when it was essentially rediscovered by Paul Erdös and Esther Szekeres. Gradually, it led to the formation of a subdiscipline of combinatorial analysis known as *Ramsey Theory*, which already contains many hundreds of papers and is growing at a remarkable rate (see the survey by Graham et al. 1980).

Ramsey’s pioneering paper on optimal taxation seems to have been written in response to a request by Pigou to look into the problem (see Pigou 1928, pp. 126–8) but his work on the theory of growth was apparently his alone, although greatly admired by and discussed with Keynes.

Mathematical Expectation, Probability and Utility

The present discussion of Ramsey’s great Chapter VII (1931, pp. 156–98) will consider it quite narrowly, as a contribution only to the theory of choice under uncertainty, and thus neglect the important question of its relation to traditional theories of probability. Ramsey himself adopted throughout a modest and peaceable tone towards probability theory, stressing that ‘the meaning of probability in logic’ may be quite different from ‘its meaning in physics’ (p. 157).

The chapter is entitled ‘Truth and Probability’ and dated 1926; presumably most of it was written then, in spite of a reference which bears the date 1927. It contains almost all of what he has to say on the subject, although further on in Chapter VIII and pages 256–7 there are a few unsystematic comments and glosses on the earlier work. The first ten pages form a critique of Keynes’s theory of probability (1921), which may well have stimulated his own interest in the whole subject, so it is not until Section 3 that Ramsey begins his 'inquiry . . . [into] . . . the logic of partial belief'.

Ignoring here all his careful qualifications, the theory outlined in that Section begins as follows (pp. 172–4):

> The old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds which he will accept. This method I regard as fundamentally sound; . . . I propose to take as a basis a general psychological theory . . . that we act in the way we think most likely to realize the objects of our desires . . . The question then arises how . . . to take account of varying degrees of certainty in his beliefs. I suggest that we introduce as a law of psychology that his behaviour is governed by what is called the mathematical expectation; . . . We thus define degree of belief in a way which *presupposes* the use of the mathematical expectation. (Italics added).

Ramsey was fully aware of the crucial dependence of his approach on mathematical expectation. Later in the *Foundations* he asks: 'The question . . . why just this law of mathematical expectation. The answer to this is that if we use probability to measure utility, as explained in my paper, then consistency [for which see below] requires just this law' (p. 251).

Putting the matter in its crudest (and so necessarily inaccurate) form, mathematical expectation as a principle of choice involves the use for any risky line of action α of a 'probability' π_i and a 'valuation' v_i attached to each of the possible outcomes α_i, that constitute α, in such a way that: (i) the expected valuation $E(\alpha)$ of α is $\Sigma\pi_i v_i$ (or an appropriate integral if α has infinitely many members, an alternative which Ramsey expressly rejects: pp. 183−4); and (ii) α is chosen rather than another risky line of action β if and only if $E(\alpha) > E(\beta)$.

Implicit in this crude form is a conflation between events and outcomes. Outcomes depend upon decisions and events, and it is in events and not outcomes that the randomness present is usually held to reside, so that given the occurrence of an event the relevant outcome on which it depends follows deterministically. Nevertheless, the randomness that inheres in the events may be transferred to the outcomes that are conditional upon those events. In the words of Arrow (1951, 1971, p. 26): 'no matter how complicated the structure of a game of chance is, we can always describe it by a single probability distribution of the final outcomes.'

Notice that because mathematical expectation depends linearly both on the probabilities and on the valuations, choice that follows this principle is made according to a *bilinear* form; there is however no necessity for the valuations of the possible outcome themselves to depend linearly upon those outcomes.

Essentially, given any two of the three concepts, mathematical expectation, probabilities and valuations, the remaining one follows more or less naturally. For example, in Daniel Bernoulli's account of the theory of risk (1738), the π_i are apparently given 'objectively', for example by the tosses of a coin. Wishing to preserve the principle of mathematical expectation, and citing the St Petersburg Paradox as evidence for the inappropriateness of using money itself as valuation, Bernoulli was thus led to a specific *utility function* to compute the correct valuations, this being a nonlinear (actually, concave) function of wealth. This did not in fact resolve the basic difficulty of the Paradox (which resides in unboundedness of the mathematical expectation) but it was a novel and important idea that was very influential.

A quite different approach was used by Bayes (1763, 1958), who actually *defined* probability in terms of mathematical expectation: 'The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening' (1958, p. 298; Jeffreys 1961, pp. 30–4, stresses the similarity here between Bayes and Ramsey). Possibly in ignorance of the earlier contribution, Bayes retained monetary valuations rather than replace them by Bernoullian utilities.

Both authors regarded the maximization of the mathematical expectation of gain as the appropriate principle of choice in an uncertain situation. But whereas Bernoulli accepted probabilities from the outside and altered the meaning of valuations so as to achieve consonance between the maximization of mathematical expectation and rational choice, Bayes started with the outside monetary valuations and thence determined probabilities so as to square rational choice with mathematical expectation.

Ramsey was more subtle. He effectively 'bootstrapped' both the valuations *and* the probabilities from mathematical expectation, at the small cost of: (*a*) a very general assumption about preferences; (*b*) an assumed existence of a certain kind of event; and (*c*) a further principle, original with him, that no agent's subjective probabilities should be *inconsistent*. To be inconsistent means that 'He could have a book made against him by a cunning better and would then stand to lose in any event' (1931, p. 182); this no-win situation is now usually called a *Dutch book*.

Sketch of a Proof

Ramsey provided sufficient detail for a formal proof of the existence of valuations and probabilities to be constructed from his system of axioms, but he did not construct one himself. Such proofs, for varying circumstances, have been given by Davidson and Suppes (1956) and Vickers (1962), while more informal discussions may be found in Jeffrey (1965, 1983, ch. 3) and Luce and Suppes (1965, pp. 291–4). Only the merest sketch is attempted here, and its mild technical detail follows Davidson and Suppes (1956) rather than Ramsey's original treatment, which was couched mainly in the concepts of Wittgenstein's *Tractatus* and the language of Russell and Whitehead's *Principia*, both long since unfamiliar.

Ramsey begins by considering the case where the agent has 'certain [that is, sure] beliefs about everything'. He then adopts assumption (*a*) above, which expressed in modern language says that the agent has a complete preference preordering over 'all possible courses of the world ...[though] ... we ... have no definite way of representing them by numbers' (1931, p. 176). Vickers points out that if different preferences can themselves be parts of different 'courses of the world' then the argument is ambiguous, and if not then the question is begged (1962, pp. 6−11); however, he shows how to resolve these problems by suitable amendment of Ramsey's definitions.

When 'the subject is capable of doubt' (p. 177), the theory proceeds by offering options. Suppose that the agent has two options: the first is α, in which he receives x if an event e occurs and a preferentially different outcome y if it does not; and the other is β, in which he receives r if e occurs and another outcome s if it does not. Assuming that probabilities $\pi(e)$ and $\pi(e')$ can be attached to the events e and to e' (the complement of e), respectively, and that valuations $v(x)$, $v(y)$, and so on, can be placed on the outcomes x, y, r and s, then the principle of mathematical expectation says that α is better than, indifferent to, or worse than β, according as

$$\pi(e)v(x)+\pi(e')v(y) >;=;< \pi(e)v(r)+\pi(e')v(s). \qquad (1)$$

Ramsey's next assumption is (*b*) above, to the effect that there exists some event, say e^*, such that for *every* pair (m, n) of preferentially distinct outcomes the subject is indifferent between the option γ consisting of m if e^* and n if not e^*, and another option δ consisting of n if e^* and m if not e^*. According to the principle of mathematical expectation, this implies

$$\pi(e^*)v(m)+\pi(e^{*\prime})v(n) = \pi(e^*)v(n)+\pi(e^{*\prime})v(m). \qquad (2)$$

Since m and n are preferentially distinct, their valuations must be such that $v(m) \neq v(n)$. Then from this and (2) it follows that necessarily

$$\pi(e^*)=\pi(e^{*\prime}). \qquad (3)$$

Although quantitative probabilities have not yet been defined, (3) shows that there is a clear qualitative sense in which event e^* has a (subjective) probability of 1/2, *provided* that the subjective probabilities of an event and its complement sum to unity. Ramsey terms *ethically neutral* any event (in his language, prpt) that has the properties of e^*; the force of the word 'ethically' is not explained (p. 177). The assumption that such events (prpts) exist is perhaps the weakest part of his theory of choice under uncertainty, although before it is rejected out of hand the careful philosophical discussion of it by Vickers (1962) and the equally careful empirical applications of it discussed by Davidson and Suppes (1956) should be consulted.

Now take the case of (1) where e is an ethically neutral event e^* and the option α is indifferent to the option β. Then from (1) and (3),

$$v(x) - v(r) = v(s) - v(y). \tag{4}$$

This says that differences in valuations can be equated, so that the latter are measurable by an *interval scale*; or what comes to the same thing, that they are measurable up to choice of unit and origin, so that for any other such scale μ , $\mu(\cdot) = \alpha + bv(\cdot)$, where $b > 0$.

A valuation having been obtained in this fashion for each outcome, and assuming again that for any event e, $\pi(e) + \pi(e') = 1$, it follows from the case of equality in (1) that

$$[\pi(e)]^{-1} = 1 + [v(x) - v(r)]/[v(s) - v(y)]. \tag{5}$$

This gives a way of calculating the subjective probability $\pi(e)$ of any event, ethically neutral or not, in a way compatible simultaneously with the principle of mathematical expectation and with the valuations $v(\cdot)$ of the possible outcomes. Thus both valuations ('utilities') and subjective probabilities have been bootstrapped, in that order, from the simple assumptions (a) and (b), plus the assumption that any event and its negation have subjective probabilities that add up to 1. Ramsey dispenses with this last, auxiliary assumption by means of his principle (c) of consistency, which in effect insists upon the impossibility of Dutch books.

Dutch Books

Although his paper is crystal clear that consistency means that the subjective probabilities of any set of disjoint and exhaustive events must sum to 1, and it is twice stated explicitly (pp. 182–3) that anyone who is not consistent in this sense can have a Dutch book made against him, Ramsey provided no formal proof of equivalence between these two ideas. Hence this result is usually attributed to de Finetti (1937, 1964), who gave a very neat proof. Not having read Ramsey's paper, de Finetti like Bayes worked with monetary valuations in his account of personal probability, though he admitted later (1964, p. 102 fn(a)) that 'Such a formulation could better, like Ramsey's, deal with expected *utilities*; I did not know of Ramsey's work before 1937, but I was aware of the difficulty of money bets.' What follows is a free adaptation of de Finetti's proof to Ramsey's problem.

Let there be n mutually incompatible and together exhaustive events e_i, for example, the faces of a die. Suppose then that I, knowing your subjective probabilities π_i, offer you the following wager: If e_i occurs I pay you σ_i. In return, you pay me an initial stake of $\Sigma\pi_i\sigma_i$ valuation units, where the sum is taken over the n events. If you behave according to Ramsey's theory of choice under uncertainty, then you should be on the margin of accepting this wager, since for you to attach probability π_i to e_i is to say that you would be indifferent between the following offers: receive σ_i valuation units contingent on the occurrence of e_i, and the amount $\pi_i\sigma_i$ for sure. Since by hypothesis the e_i are exclusive events, the separate amounts $\pi_i\sigma_i$ may be added together.

If event e_h, occurs, your gain is

$$\gamma_h = \sigma_h - \sum \pi_i\sigma_i \; h = 1, 2, \ldots, n. \tag{6}$$

These are n linear equations, which can be put into matrix–vector notation. Writing g and s for the vectors of the γ_i and σ_i, respectively, I for the $n \times n$ identity matrix, and P for the matrix whose (i, j)th element is π_j, the equations (6) become

$$g = (I - P)s. \tag{7}$$

Computation shows that det $(I - P) = 1 - \Sigma\pi_i$. So if $\Sigma\pi_i \neq 1$, then for any desired vector of gains g stakes $s = (I - P)^{-1}\, g$ can be computed that will guarantee me the vector $-g$. In particular, I can specify g to be strictly negative, thus ensuring that you will lose whatever event occurs.

Conversely, suppose that your subjective probabilities are what de Finetti called coherent (and Ramsey, consistent), so that by definition $\Sigma\pi_i = 1$. Then, multiplying each equation in (6) by π_h and adding over all events,

$$\sum \pi_h \gamma_h = \sum \pi_h \sigma_h - \sum \pi_h \sum \pi_i \sigma_i = 0 \quad (8)$$

Since each $\pi_h \geq 0$ and their sum is non-zero, it follows from (8) that not all the γ_h can be negative. Hence the condition that your subjective probabilities π_i , sum to 1 for all complete sets of incompatible events e_i, that is, that you obey the rules of probability calculus, is necessary and sufficient in order that no Dutch book can be made against you.

The Reception of 'Truth and Probability'

Ramsey's theory of choice under uncertainty was deeply original. Emile Borel, in his review (1924, 1964) of Keynes's theory of probability, had earlier sketched an interesting theory of subjective probability in terms of bets (note in particular his remark that 'the method of betting permits us in the majority of cases a numerical evaluation of probabilities that has exactly the same characteristics as the evaluation of prices by the methods of exchange'; 1964, p. 57), but nobody had come close to the depth and comprehensiveness of Ramsey's theory. He was characteristically modest about its range of application: 'I only claim for what follows approximate truth … like Newtonian mechanics … [it] can, I think, still be profitably used even though it is known to be false' (p. 173).

Perhaps because the theory was too original, such modesty did not help its author, any more than his high reputation as a philosopher. I can find no evidence that anyone, let alone any economist, took any serious notice of Ramsey's work until after von Neumann and Morgenstern's quite separate utility theory had appeared in 1947. The latter theory was very much in the Bernoullian tradition, in which the probabilities are given from outside, 'objectively'. Coupling these with a complete preference preordering for such alternatives, suitable continuity, and the principle of mathematical expectation in the form of the independence axiom, the authors were able to deduce the existence of a utility function, unique up to positive affine transformations, which gave valuations compatible both with the outside probabilities and that principle.

The first published reference to Ramsey's theory known to me appears in Little (1950, p. 29, fn.1), who considered it 'essentially the same' as that of von Neumann and Morgenstern. Little's reference was soon followed by one in Arrow (1951), who acknowledged that Ramsey was brought to his attention by Norman Dalkey. Though complaining that 'Ramsey's work was none too clear' (1971, p. 26), Arrow did see that it originated 'a new stage' in decision theory, 'in which a priori probabilities are derived from behavior postulates' (1971, p. 22). Thereafter there was a gradual increase in the appreciation of Ramsey's contribution, although even as late as 1954 an excellent collection of papers on decision theory (Thrall et al. 1954) contained not one reference to his work.

It is a common mistake to suppose that the line of descent in the theory of personal probability is direct from Ramsey to de Finetti (1937) to Savage (1954). We have seen that de Finetti did not know of Ramsey's work, his own remarkable contribution being very much in the Bayesian tradition which takes the valuations from outside and thence derives the probabilities. Moreover, a careful reading of Savage's fine book shows that Ramsey's influence was at best peripheral, the axiomatization of probabilities and valuations proceeding far more along the lines developed by de Finetti.

There have in fact been relatively few explicit exponents of Ramsey's approach. The most notable are probably Davidson and Suppes (for example, Davidson and Suppes 1956) and Anscombe and Aumann (1963), who used an interesting bootstrapping argument to go from assumed probabilities for what they called 'roulette' lotteries to valuations, and thence to subjective probabilities for the much wider class of 'horse' lotteries, all very much in the Ramsey manner.

The direct heirs to Ramsey's work have been few but there is no doubt that its influence has been pervasive, to such extent that chairs in decision theory at US business schools have been named after him (though with what warrant is hard to say). Arrow (1965, p. 57) claimed that all arguments involving the expected-utility hypothesis 'are only variations of Ramsey's', while Savage et al. (1962, p. 10) wrote that the 'more thorough-going . . . formulation of Ramsey (1931) . . . is in no way obsolete'. Even now, not to experience that 'clear purity of illumination with which the writer's mind is felt by the reader to play about its subject' (Keynes on Ramsey 1928) is a sad loss for the modern student.

See Also

▶ Expected Utility Hypothesis
▶ von Neumann, John (1903–1957)

Selected Works

1926. Truth and probability. Chapter VII, 156–198 in (1931). Reprinted in (1978) and in Kyburg and Smokler (1964).
1927. A contribution to the theory of taxation. *Economic Journal* 37: 47–61. Reprinted in (1978).
1928. A mathematical theory of saving. *Economic Journal* 38: 543–549. Reprinted in (1978).
1931. *The foundations of mathematics*, ed. R.B. Braithwaite, with a Preface by G.E. Moore. London: Routledge & Kegan Paul.
1978. *Foundations*, ed. D.H. Mellor, with Introductions by the ed. L. Mirsky, T.J. Smiley, and J.R.N. Stone. London: Routledge & Kegan Paul.

Bibliography

Anscombe, F.J., and R.J. Aumann. 1963. A definition of subjective probability. *Annals of Mathematical Statistics* 34: 199–205.
Arrow, K.J. 1951. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica* 19: 404–437.
Arrow, K.J. 1965. *Aspects of the theory of risk-bearing*. Helsinki: Yrjö Jahnsson Foundation.
Arrow, K.J. 1971. *Essays in the theory of risk-bearing*. Chicago: Markham.
Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society* 53: 370–418. Reprinted in *Biometrika* 45: 293–315, (1958).
Bernoulli, D. 1738. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* V. Trans. as Exposition of a new theory on the measurement of risk. *Econometrica* 22: 23–36, (1954).
Borel, E. 1924. A propos d'un traité de probabilité. *Revue Philosophique* 98: 321–336. Trans. in Kyburg and Smokler (1964).
Cohen, P.J. 1966. *Set theory and the continuum hypothesis*. New York: W.A. Benjamin.
Davidson, D., and P. Suppes. 1956. A finitistic axiomatization of subjective probability and utility. *Econometrica* 24: 264–275.
De Finetti, B. 1937. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7: 1–68. Trans. in Kyburg and Smokler (1964).
Graham, R.L., B.L. Rothschild, and J.H. Spencer. 1980. *Ramsey theory*. New York: Wiley.
Harrod, R.F. 1951. *The life of John Maynard Keynes*. London: Macmillan.
Jeffrey, R.L. 1965. *The logic of decision*. New York: McGraw-Hill. 2nd ed., Chicago: University of Chicago Press, 1983.
Jeffreys, H. 1961. *Theory of probability*. 3rd ed. Oxford: Clarendon Press.
Keynes, J.M. 1921. *A treatise on probability*. London: Macmillan.
Keynes, J.M. 1933. *Essays in biography*. London: Macmillan.
Kyburg, H.E., and H.E. Smokler, eds. 1964. *Studies in subjective probability*. New York: Wiley.
Little, I.M.D. 1950. *A critique of welfare economics*. Oxford: Clarendon Press.
Luce, R.D., and P. Suppes. 1965. Preference, utility, and subjective probability. In *Handbook of mathematical psychology*, ed. R.D. Luce, R.R. Bush, and E. Galanter, vol. 3. New York: Wiley.
von Neumann, J. 1928. Zur Theorie der Gesellschaftspiele. *Mathematische Annalen* 100: 295–320.
von Neumann, J. 1937. Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunksatzes. *Ergebnisse eines mathematischen Kolloquiums* 8: 78–83. Trans. as A model of general economic equilibrium. *Review of Economic Studies* 13: 1–9, 1945–6.
von Neumann, J., and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press.
von Neumann, J., and O. Morgenstern. 1947. 2nd ed., of (1944).
Partridge, F. 1981. *Memories*. London: Victor Gollancz.

Pigou, A.C. 1928. *A study in public finance*. London: Macmillan.
Savage, L.J. 1954. *The foundations of statistics*. New York: Wiley.
Savage, L.J., et al. 1962. *The foundations of statistical inference: A discussion*. London: Methuen.
Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.
Thrall, R.M., C.H. Coombs, and R.L. Davis, eds. 1954. *Decision processes*. New York: Wiley.
Ulam, S.M. 1976. *Adventures of a mathematician*. New York: Charles Scribner's Sons.
Vickers, J.M. 1962. A critical investigation of Frank Ramsey's theory of value and belief. Ph.D., thesis, Department of Philosophy, Stanford University.
Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*, with an Introduction by Bertrand Russell. London: Kegan Paul, Trench, Trubner & Co.
Wittgenstein, L. 1953. *Philosophical investigations*. Trans. G.E. Anscombe. Oxford: Basil Blackwell.
Wittgenstein, L. 1973. In *Letters to C.K. Ogden*, ed. G.H. von Wright. Oxford: Basil Blackwell.
von Wright, G.H. 1982. *Wittgenstein*. Minneapolis: University of Minnesota Press.

Random Coefficients

P. A. V. B. Swamy and J. R. Barth

Random coefficients models generalize conventional fixed coefficients models to avoid inconsistent and inaccurate assessments of relationships among variables.

Fixed Coefficients Models

As Goldberger (1964, pp. 380–88) indicates, there are three alternative ways to formulate fixed coefficients models: (1) structural-form, (2) reduced-form, and (3) recursive-form. The view of the economic mechanism underlying these formulations is that there is a joint probability distribution of the current endogenous (random) variables $y_t^* = \left(y_t^*, \ldots, y_{Lt}^*\right)'$ conditional on the values of the predetermined variables $\mathrm{x_t} = (x1_{t,\ldots,}\ x_{kt})'$. This conditional distribution may be written as

$$p(y_t|x_t, \theta), \tag{1}$$

where θ is a fixed parameter vector taking values in a parameter space Ω and Y is a sample space in which y_t takes on its values.

One way to specify this distribution is to postulate a reduced-form model. The vector θ then includes the fixed coefficients as well as the disturbance variances and covariances of the reduced-form. Alternatively, one could postulate a structural-form model and then deduce the conditional distribution by transforming the structural-form into the reduced-form. Finally, one could postulate a model in recursive-form, thereby imposing upon the structural-form the triangularity restriction on the coefficients matrix of the endogenous variables and the diagonality restriction of the covariance matrix of the disturbances.

Inaccuracies and Inconsistencies

Regardless of the formulation postulated, the usefulness of fixed coefficients models is limited because of inherent inconsistencies and inaccuracies. For example, even when the fixed coefficients model is formulated in structural-form, the equations frequently are subjected to episodic breakdowns that are usually handled by judgemental 'add factors' and dummy variables. Even in those situations when microeconomic relationships remain invariant under changed circumstances, the corresponding macroeconomic relationships obtained by aggregating across individual units may not remain invariant, as shown by Swamy et al. (1982).

Besides these inaccuracies, the reduced-form is invariant under nonsingular transformations of the structural-form, see Goldberger (1964, p. 312). Different values of the structural parameters therefore imply the same conditional distribution so that the structural equations may not be identified. Yet, identification is a necessary condition for statistical consistency, see Gabrielsen (1978). Of course, one can achieve identification by imposing restrictions on the parameters of structural equations. But if these restrictions are

R

overidentifying, they can be inconsistent, as shown by Conway et al. (1984, p. 7). Furthermore, there are cases in which the reduced-form parameters are also not identified without appropriate restrictions. Imposing identifying restrictions on the reduced-form parameters, however, may contradict the structural identifying restrictions, see Swamy (1980) and Swamy and Mehta (1983), leading to a logically inconsistent model. Surely no one would wish to construct this type of model, since as Boland convincingly argues, even if one cannot prove a model is true, to be true it must be at least logically consistent, see Swamy et al. (1985). Furthermore, if a model is logically inconsistent, the notion of the true value of a parameter and the related concept of statistical consistency do not apply.

Finally, according to Lane (1984), there are three possible interpretations of the elements of Ω:

(a) $\boldsymbol{\theta}$ is compatible with the conditional distribution;
(b) Ω is an abstract set and $\boldsymbol{\theta}$ merely indexes the conditional distributions;
(c) $\boldsymbol{\theta}$ is a possible value for some 'real' physical parameter and function (1) is to be regarded as the distribution of the random quantity y^* (conditional on x) should $\boldsymbol{\theta}$ be the true value of that parameter.

Interpretation (c) raises the difficult philosophical question: When and in what sense do 'real' physical parameters exist? Furthermore, must one believe that each structural parameter has a propensity to take a single value? It is difficult to believe that there are model-free physical quantities underlying each model parameter, without guidance as to what constitutes reality and how reality is linked to the mathematics embodied in specific models. Such guidance is impossible, however, because the truth status of a logically consistent model cannot be established, see Swamy et al. (1985). Thus, one never knows when interpretation (c) is appropriate. An appeal to statistical consistency based on the notion of the true values of parameters therefore cannot be made with any conviction.

Since interpretations (a) and (b) are defined solely in terms of the assumed (mathematical) model and do not necessarily refer to the physical reality that model is intended to represent, they are mathematically precise. However, interpretation (a) provides no scope for the mixture principle (that is, permitting an assumed underlying distribution for the parameters to affect the ultimate values of the endogenous variables), since only models whose sampling distributions are identical share 'the same Ω '. Fixed coefficients models thus apply only to situations when the model structure can be represented solely in terms of probability distributions on the sample space indexed by the fixed and unknown $\boldsymbol{\theta}$.

Interpretation (b), on the other hand, does provide wide scope for mixing. Indeed, any two fixed coefficients models with the same index set can be mixed. Consequently, if there exists a pair of observations, one from each model, yielding the same likelihood function on the index set Ω, the likelihood principle holds that the 'evidence' or 'inference' derived from the two models with these two observations must be identical. Yet, this conclusion may not only be incorrect, but is inconsistent with the Bayesian approach, as Lane (1984) points out.

In sum, the foundational status of a fixed coefficients model cannot be determined until Ω is interpreted. Depending upon whether one adopts interpretation (a), (b), or (c), fixed coefficients models are either devoid of interesting consequences (since θ is fixed and unknown), wrong (since inferences may be incorrect and unacceptable to Bayesians), or severely and ambiguously restricted in its domain of applicability (since the truth status of models cannot be known).

Random Coefficients Models

A way to avoid the difficulties mentioned in the preceding section is to use the following random coefficients model (for earlier models, see Swamy (1971) and the references therein) developed by Swamy and Tinsley (1980):

$$\begin{aligned} &\text{(i)} \quad y_t^* = x_t'\beta_t^*; \\ &\text{(ii)} \quad \beta_t^* = Bz_t + J\xi_t^*; \\ &\text{(iii)} \quad \xi_t^* = \Phi\xi_{t-1}^* + \upsilon_t^*; \end{aligned} \tag{2}$$

with $E\left(\upsilon_t^*|z_t, x_t, \xi_{t-1}\right) = E\left(\upsilon_t^*\right) = 0$ for all z_t, x_t, and $\xi_{t-1}, E\left(\upsilon_t^*\upsilon_s^{*\prime}|z_t, x_t, \xi_{t-1}\right) = \Delta_\upsilon$ if $t = s$ and 0 if $t \neq s$. One element of each of x_t and z_t may be identically equal to 1 for all t, with the coefficients corresponding to these unit elements representing a random intercept and a constant vector, respectively. Although Swamy and Tinsley set $J = [I, 0,\ldots, 0]$, alternative choices for J are possible.

Since the disturbance term is indistinguishable from a time-dependent random element of β_t^* corresponsing to the unit element of x_t, both specifications are combined into a single element of β_t^*. When an equation is part of a larger model, regressors may be jointly determined with the regressand and hence correlated with the contemporaneous disturbance term, see Goldberger (1964, p. 292). If so, elements of x_t in equation (2) (i) are correlated with β_t^*, which means they also appear in z_t. Equation (2) (ii) admits such correlations. The term $J\xi_t^*$ is that part of β_t^* that is mean-independent of x_t and z_t.

Clearly, fixed coefficients models are special cases of random coefficients models. This is the case, for instance, when all the elements of β_t^* corresponding to the non-constant elements of x_t have zero variances and when the columns of B corresponding to the non-constant elements of z_t are null. Thus, in fixed coefficients models, the intercept but not the slopes may be interpreted as random, see Swamy (1970, 1971, p. 8).

When time series of cross-sections data are available, equation (2) has been generalized by Swamy and Mehta (1975) to

$$\begin{aligned} &\text{(i)} \quad y_{it}^* = x_{it}'\beta_{it}^* = \sum_{j=1}^{K} x_{jit}\beta_{jit}^* \\ &\text{(ii)} \quad \beta_{it}^* = \beta + \alpha_i^* + \xi_{it}^*, \end{aligned} \tag{3}$$

where i indexes cross-section observations, t indexes time series observations, the α are independently distributed with mean vector zero and constant covariance matrix Δ_α, the ξ_{it}^* are distributed with mean vector zero and a general covariance matrix Δ the α are independent of the ξ_{it}^* and the β_{it}^* are mean-independent of the x_{it}.

Since these random coefficients models are designed to provide only a convenient approach to modelling relationships, they do not carry a metaphysical burden of 'reality' for the parameters they contain. Furthermore, equations (2) (ii) and (3) (ii) provide rich classes of coherent mixing functions. Swamy and Mehta (1975) and Swamy and Tinsley (1980) use data based methods to select the mixing functions, though purely subjective beliefs can form the basis for these functions. Based on these observations, the correct interpretation for the coefficients β_t^* and β_{it}^* is Lane's interpretation (b). Unlike fixed coefficients, random coefficients are not subject to inconsistent restrictions. Since the ultimate aim of inference is typically to generate an accurate prediction about the value of some future observables, Swamy and Lad (1985) employed a random coefficients model to generate predictions about the future values of stock prices based upon the current and past values of dividends. The resulting forcasts are substantially better than those obtained from the corresponding fixed coefficients model, demonstrating the potential gain in accuracy provided by this consistent approach to modelling relationships among variables.

In conclusion, rather than indicting fixed coefficients models, the comments presented here emphasize the shortcomings of that approach as compared with random coefficients models, thereby providing the researcher with more complete information when deciding upon an empirical model.

See Also

▸ Estimation

Bibliography

Conway, R.K., P.A.V.B. Swamy, J.F. Yanagida, and P. von zur Muehlen. 1984. The impossibility of causality testing. *Agricultural Economic Research* 36(3): 1–19.

Gabrielsen, A. 1978. Consistency and identifiability. *Journal of Econometrics* 8(2): 261–263.

Goldberger, A.S. 1964. *Econometric theory*. New York: Wiley.
Lane, D.A. 1984. *Discussion of the likelihood principle by J.O. Berger and R.L. Wolpert*, Lecture notes – monograph series, vol. 6. Hayward: Institute of Mathematical Statistics.
Swamy, P.A.V.B. 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38(2): 311–323.
Swamy, P.A.V.B. 1971. *Statistical inference in random coefficient regression models*. New York: Springer.
Swamy, P.A.V.B. 1980. A comparison of estimators for undersized samples. *Journal of Econometrics* 14(2): 161–181.
Swamy, P.A.V.B., and F. Lad. 1985. *Forecasting stock prices with the stochastic coefficients models*. Washington, DC: Federal Reserve Board, Special Studies.
Swamy, P.A.V.B., and J.S. Mehta. 1975. Baysian and non-Bayesian analysis of switching regressions and of random coefficient regression models. *Journal of the American Statistical Association* 70(Pt I): 593–602.
Swamy, P.A.V.B., and J.S. Mehta. 1983. Further results on Zellner's minimum expected loss and full information maximum likelihood estimators for undersized samples. *Journal of Business and Economic Statistics* 1(2): 154–162.
Swamy, P.A.V.B., and P.A. Tinsley. 1980. Linear prediction and estimation methods for regression models with stationary stochastic coefficients. *Journal of Econometrics* 12(2): 103–142.
Swamy, P.A.V.B., J.R. Barth, and P.A. Tinsley. 1982. The rational expectations approach to economic modelling. *Journal of Economic Dynamics and Control* 4(2): 125–147.
Swamy, P.A.V.B., R.K. Conway, and P. von zur Muehlen. 1985. The foundations of econometrics – Are there any? *Econometric Reviews* 4(1): 1–61 (With discussion.).

Random Fields

Nazgul Jenish

Abstract

Random fields are stochastic processes indexed by a multidimensional parameter. They possess some interesting properties, e.g. isotropy and the Markov property, and satisfy laws of large numbers and weak convergence theorems under fairly general conditions. As such, random fields provide a powerful tool for modelling spatial phenomena in physics, biology, economics, and other social sciences.

Keywords

Markov random fields; Random fields; Spatial processes

JEL Classifications

C10; C31; D71

Definition and Examples

Many physical and economic phenomena are described by random variables that depend on a multidimensional parameter. For example, the height of the sea surface, unemployment and crime rates in cities depend on spatial coordinates $\mathbf{r} = (r_1, r_2)$. Let $D \subseteq \mathbb{R}^d$ be a subset of $\mathbb{R}^d$, $d \geq 2$.

Definition 1 *A random field*, $X_\mathbf{r}$, $\mathbf{r} = (r_1, \ldots, r_d) \in D$ *is a family of random variables* $\{X_\mathbf{r}(\omega), \mathbf{r} \in \mathrm{D}, \omega \in \Omega\}$ defined on a probability space $(\Omega, \mathfrak{F}, P)$.

This definition generalises the notion of stochastic processes indexed by a scalar parameter t interpreted as time. The random field $X_\mathbf{r}$ may take its values in $\mathfrak{X} \subseteq \mathbb{R}^k$, $k \geq 1$, i.e. $X_\mathbf{r}$ may be vector-valued, e.g. average income, crime and unemployment rates in cities can be viewed as a vector-valued random field indexed by $\mathbf{r} = (r_1, r_2)$. For given ω, $x_\mathbf{r} = X_\mathbf{r}(\omega)$ is a realization or sample function. Thus a random field can be equivalently defined as a measurable mapping from Ω to the sample space: $\Omega \rightarrow \mathfrak{X}_D$.

An important example of random fields is a Poisson random field used to model random point patterns in space, e.g. the location of trees or firms. For any Borel subset $B \in \mathbb{R}^d$, let $N(B)$ be the random number of points in B. Suppose for any disjoint sets, $B_1, \ldots, B_m$ in $\mathbb{R}^d$, the random variables $N(B_1), \ldots, N(B_m)$ are independent, and there are no multiple points within infinitesimal spatial domains. Then the random point field $N(B)$ is called a Poisson field with the intensity measure $\mu(B)$ if

$$P\{N(B)=n\}=\frac{[\mathrm{M}(B)]^n e^{-\mathrm{M}(B)}}{n!}, n=0,1,2,\ldots$$

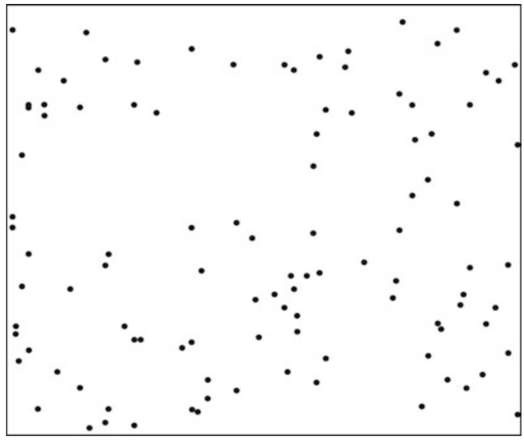
If $\mu(B) = a\lambda(B)$, $a > 0$, where λ (B) is the Lebesgue measure of B, the Poisson field is said to be homogeneous. A homogenous Poisson field on the unit square is shown in Fig. 1.

If D is continuous, the random field is called continuous. For example, D may coincide with $\mathbb{R}^d$ or $[0,\ 1]^d$. Continuous random fields are used in turbulence theory, geology and image processing. Since economic agents (individuals, firms, cities, states etc.) are intrinsically discrete, discrete-index random fields, e.g. lattice random fields on $D = \mathbb{Z}^d$, are more popular in economic applications.

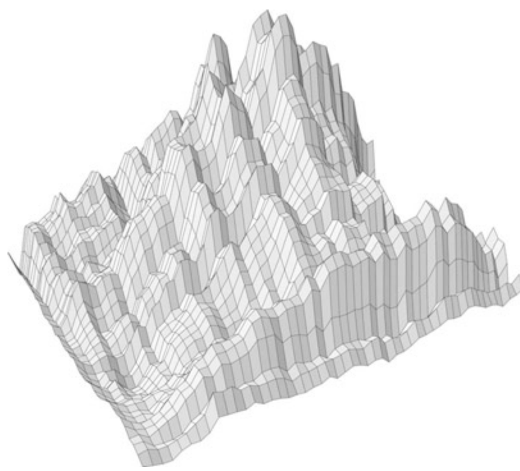
As with stochastic processes, a random field can be specified by a collection of finite-dimensional distributions:

$$F_{\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m}(x_1,x_2,\ldots,x_m) \\ = P\left(X_{\mathbf{r}_1} \ll x_1,\ldots,X_{\mathbf{r}_m} \ll x_m\right)$$

for $m \geq 1$. The Kolmogorov consistency theorem ensures the existence and uniqueness of a random field with the given finite-dimensional distributions. The field is called Gaussian if all its finite-dimensional distributions are normal. The Brownian sheet $B_\mathbf{r}$, $\mathbf{r} \in [0,\ 1]^d$ is a Gaussian field with zero mean and covariance Cov; $(B_\mathbf{r}, B_\mathbf{s}) = \prod_{k=1}^d \min\{r_k, s_k\}$. The Brownian sheet on the unit square $[0,\ 1]^2$ is depicted in Fig. 2. This figure resembles a rough landmass jutting out of the ocean or sea surface during a storm.



Random Fields, Fig. 1 Poisson random field on the unit square



Random Fields, Fig. 2 Brownian sheet on the unit square

Classes of Random Fields

Some important classes of random fields are discussed below. We begin with various characterizations of homogeneity.

Definition 2 (i) *A random field $X_\mathbf{r}$ is strictly homogeneous if for every collection $\mathbf{r}_1,\mathbf{r}_2,\ \ldots,\ \mathbf{r}_m$ on D and every $\mathbf{h}$ such that $\mathbf{r}_i + \mathbf{h} \in D$, $i = 1, \ldots, m$*

$$F_{\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m}(x_1,x_2,\ldots,x_m) \\ = F_{\mathbf{r}_1+\mathbf{h},\mathbf{r}_2+\mathbf{h},\ldots,\mathbf{r_h}+\mathbf{r}}(x_1,x_2,\ldots,x_m).$$

That is, the finite-dimensional distributions are translation-invariant. (ii) *A random field $X_\mathbf{r}$ with* $\mathrm{E}|X_\mathbf{r}|^2 < <\infty$ *is weakly homogeneous if for all* $\mathbf{r}$, $\mathbf{r} + \mathbf{h} \in D$

$$\mathrm{E}(X_\mathbf{r}) = \mathrm{const},\ Cov(X_\mathbf{r}, X_{\mathbf{r}+\mathbf{h}}) = K(\mathbf{h}) \\ = K(h_1,\ldots,h_d).$$

These definitions generalise, respectively, strictly stationary and weakly stationary stochastic processes. The spectral representation of stationary stochastic processes extends easily to homogenous random fields. In higher-dimensional spaces, there is yet another characterization of homogeneity:

Definition 3 *A weakly homogenous random field $X_{\mathbf{r}}$ is isotropic if its co-variance function $Cov(X_{\mathbf{r}}, X_{\mathbf{r+h}})$ depends only on the length of vector $\mathbf{h}$: $Cov(X_{\mathbf{r}}, X_{\mathbf{r+h}}) = K(|\mathbf{h}|)$.*

Finite-dimensional distributions of an isotropic field are invariant to rotation of $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ around axes through the origin. Moreover, $K(|\mathbf{h}|)$ can be expressed in terms of Bessel functions (cf. Cressie 1993), which simplifies statistical analysis considerably.

We now turn to various dependence conditions for random fields. In contrast to the time-line, there is no natural order in a higher-dimensional space. For this reason, the dependence concepts that rely on linear order, including martingales and Markov processes, do not extend easily to random fields.

Suppose the random field $X_{\mathbf{r}}$, $\mathbf{r} \in V$ is defined on the graph $\Gamma = \{V, E\}$, where $V = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ is the set of vertices or sites, and E is the set of edges. For any subset $\Lambda \subseteq V$, let $X_\Lambda = \{X_{\mathbf{r}}, \mathbf{r} \in \Lambda\}$ denote a configuration on Λ. The neighbourhood of $\mathbf{r}$ is the set $N(\mathbf{r}) = \{\mathrm{s} \in V: (\mathbf{r}, \mathbf{s}) \in E\}$.

Definition 4 *A random field $X_{\mathbf{r}}$ on $\Gamma = \{V, E\}$ is called Markov if its conditional distributions satisfy the following condition*

$$P\left(X_{\mathbf{r}} | X_{V/\{r\}}\right) = P\left(X_{\mathbf{r}} | X_{N(r)}\right).$$

Thus a Markov random field is defined by means of conditional distributions. Given this, the natural question arises: what is the joint distribution whose conditional distributions define a Markov field? The answer is given by the Hammersley-Clifford theorem (cf. Cressie 1993), which states that for a Markov field on Γ with a discrete state space and positive probabilities, the joint distribution can be represented as

$$P(X_V) = \prod_{C \in cl(\Gamma)} \Phi_C(X_C) \tag{1}$$

where C is a subset of V, called a clique, that consists either of a single site or sites that are all pairwise neighbours, $cl(\Gamma)$ is the set of cliques on Γ, and $\phi_C(\cdot)$ are positive functions, referred to as the clique potentials.

This representation of Markov fields is closely related to Gibbs random fields in statistical mechanics. In fact, the use of Markov random fields originated in statistical physics of interacting particles. Imagine a system of n interacting particles located at sites $\mathbf{r}_i \in \mathbb{Z}^d$, $i = 1, \ldots, n$. The random field $X_{\mathbf{r}}$, $\mathbf{r} \in V$ on $\Gamma = \{V, E\}$ is called a Gibbs field (or a Gibbs measure) if

$$P(X_V) = Z^{-1}\exp\{-U(X_V)\} \tag{2}$$

where Z is the normalising constant such that $P\,(X_V)$ is a well-defined density, and

$$U(X_V) = \sum_{\mathbf{r}_i \in V} \left(u_0(X_{\mathbf{r}_i}) + \sum_{\mathbf{r}_i \neq \mathbf{r}_j \in V: \left\|\mathbf{r}_i - \mathbf{r}_j\right\| \leq m} u_1\left(X_{\mathbf{r}_i}, X_{\mathbf{r}_j}\right) \right)$$

is the potential or energy of finite-range interaction between m-neighbours. If $m = 1$, interactions only between nearest neighbours are allowed, e.g. within cliques. For instance, the Ising model of ferromagnetism (cf. Kindermann and Snell 1980) is a nearest neighbour model of interacting magnets located at sites $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ on $\mathbb{Z}^2$ with spins $X_{\mathbf{r}} \in \{1, -1\}$ and the energy:

$$U\,(X_V) = \sum_{i=1}^{n} \left(-hX_{\mathbf{r}_i} - J \sum_{j \neq i:\ \left\|\mathbf{r}_i - \mathbf{r}_j\right\| = 1} X_{\mathbf{r}_i} X_{\mathbf{r}_j} \right).$$

where h and J are parameters. Comparing (1) and (2), it is clear that a Markov field can be represented as a Gibbs field with suitable choice of the potential function, e.g. $\phi_C\,(X_C) \equiv Z^{-1} \exp\, u_k(X_{\mathbf{r}}, X_{\mathbf{s}})$ for $\|\mathrm{r} - \mathrm{s}\| = k$, $k = 0, 1$. Conversely, the conditional probabilities of a Gibbs field satisfy the Markov property. For more detailed discussion of Markov fields, see Kindermann and Snell (1980).

More generally, Dobrushin (1968) established conditions under which a random field defined by conditional probabilities exists and is unique. Existence obtains under fairly mild conditions, e.g. homogeneity, while the uniqueness of a random field does not come from its existence; for example the Ising model exhibits phase transition for different values of J and h. Intuitively, uniqueness requires the autocorrelation to decay sufficiently fast, i.e. interactions declining with distance. This implies that unique Gibbs fields satisfy some mixing conditions, e.g. φ-mixing, and hence satisfy laws of large numbers and central limit theorems under appropriate regularity conditions (cf. Dobrushin 1968).

Economic Applications

Random fields and, in particular, Gibbs fields, have been used widely in economics to model phenomena as diverse as voting, education, crime, residential segregation, location and concentration of economic activity, technology adoption, asset prices, growth and business cycles. For a recent survey, see Brock and Durlauf 2001b. The common feature of these models is the explicit description of interactions among heterogeneous agents that give rise to interesting aggregate dynamics or patterns, including multiple equilibria and phase transition. Blume (1993) provides a rigorous game-theoretical formulation of interaction models. Brock and Durlauf (2001a) develop a general discrete-choice model of I interacting agents that has a structure similar to that of the Ising model. Specifically, they define the utility function of agent i as

$$U(\Omega_i) = u(\Omega_i) + J\Omega_i(I-1)^{-1}\sum_{j\neq 1} m^e_{i,j} + \varepsilon(\Omega_i)$$

where $u(\omega_i)$ is the agent i's private utility from her choice $\omega_i \in \{-1,1\}$, the second term represents a social utility associated with i's choice, which depends on the expected values, $m^e_{i,j}$, of other agents' choices from i's perspective, and $\varepsilon(\omega_i)$ is a random utility term. If the $\varepsilon(\omega_i)$ are independent and extreme-value distributed, then the joint probability measure over $\omega = (\omega_1, \ldots, \omega_I)$ is a Gibbs measure with

$$\Pr(\boldsymbol{\Omega}) = \exp\left[\beta\left(\frac{\sum_{i=1}^{I}\left(u(\Omega_i) + J\Omega_i\overline{m}^e_i\right)}{\sum_{v_1\in\{-1,1\}}\cdots\sum_{v_I\in\{-1,1\}}\exp\left[\beta\left(\sum_{i=1}^{I} u(v_i) + Jv_i\overline{m}^e_i\right)\right]}\right)\right]$$

where $\overline{m}^e_i = (I-1)^{-1}\sum_{j\neq i} m^e_{i,j}$, β and J are some parameters. The model generates multiple equilibria and phase transitions for different values of β and J. Furthermore, laws of large numbers and weak convergence results for random fields can be exploited to study dynamics and the emergence of aggregate states in economic systems. For more examples and the methodology of interaction-based models, see Brock and Durlauf (2001b) and references therein. Thus, random fields provide an elegant and powerful tool for economic models.

See Also

- ► Markov Processes
- ► Spatial Economics

Bibliography

Blume, L.E. 1993. The statistical mechanics of strategic interaction. *Games and Economic Behavior* 5: 387–424.

Brock, W.A., and S.N. Durlauf. 2001a. Discrete choice with social interactions. *Review of Economic Studies* 68: 235–260.

Brock, W.A., and S.N. Durlauf. 2001b. Interaction-based models. In *Handbook of econometrics*, vol. 5, ed. J.J. Heckman and E.E. Leamer, 3297–3380. Amsterdam: North-Holland.

Cressie, N. 1993. *Statistics for spatial data*. New York: Wiley.

Dobrushin, R.L. 1968. The description of a random field by means of conditional probabilities and conditions on its regularity. *Theory of Probability and its Applications* 13: 197–224.

Kindermann, R., and J.L. Snell. 1980. *Markov random fields and their applications*. Providence: American Mathematical Society.

Random Variables

I. Richard Savage

Scientific statements often have a probabilistic element, for example, 'In population Ω the distribution of individual income, *I*, can be approximated by a log-normal distribution'. The formal interpretation of this statement requires a moderate amount of structure, such as,

> The population Ω has n members, $\omega_1, \ldots, \omega_n$ Associated with each ω is an income, $I(\omega)$. Each ω has the same probability $P(\omega_i)$ of being observed so that $P(\omega_i) = 1/n$ for $i = 1, \ldots, n$ Finally, $P(I \leq t) \doteq F(t, \alpha, \beta, \gamma)$ for $-\infty < t < \infty$ where F is the 3-parameter log-normal distribution function.

For this formal description, the following terms are often used. The set of all *elementary events*, ω that is Ω is the *sample space*. A function, such as $I(\omega)$ defined on Ω is called a *random variable*. The *distribution function* of $I(\omega)$ is given by the probabilities of the events that if ω_i is selected, then $I(\omega_i) \leq t$ as a function of t. In this example, $F(t, \alpha, \beta, \gamma)$ is a *model* for the distribution of I. The model contains unspecified *parameters*, α, β, γ, which could depend on units of measurement, the population, time, etc. The sign $\doteq$ indicates the approximation.

Basic Properties of Random Variables

In studying a random variable, $X(\omega)$ attention is focused on finding probabilities of events described in terms of $X(\omega)$, such as, $a \leq X(\omega) \leq b$ or in terms of concepts derived from those probabilities, such as, the average or *expected value* of $X(\omega)$, see (16). Other approaches could have been taken: the development could use expected value instead of probability as the basic concept, or the sample space concept could be omitted, proceeding directly to distribution functions. The approach taken here is in the mainstream. The theory requires a σ-field of sets, $\mathcal{F}$, whose members are subsets of Ω; that is, if $A \in \mathcal{F}$ then the complement of A is in $\mathcal{F}$, $\Omega \in \mathcal{F}$, and if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then $\cup A_i \in \mathcal{F}$ The basic theory permits us to compute probabilities of events, B, only when $B \in \mathcal{F}$; these sets are called *measurable*. For the real line we select the smallest σ-field which includes all intervals of the form $(-\infty, t)$. Random variables must be *measurable*; that is, events defined in terms of a random variable must belong to $\mathcal{F}$. Thus, if X is a random variable, then for each t it is required that $\{\omega: X(\omega) \leq t\} \in \mathcal{F}$.

All probabilities of events determined in terms of $X(\omega)$ can be obtained from the *distribution function* of X, denoted by $F(t)$, where $F(t) = P(X \leq t)$ for $-\infty < t < \infty$.

The necessary and sufficient conditions for $F(t)$ to be a distribution function are:

(a) $$F(s) \leq F(t) \text{ for } -\infty < s < t < \infty.$$

(b) $$\operatorname*{limit}_{t \to -\infty} F(t) = 0.$$

(c) $$\operatorname*{limit}_{t \to \infty} F(t) = 1.$$

(d) $$\operatorname*{limit}_{x \to t^+} F(x) = F(t). \tag{1}$$

Notice, F can be used to compute probabilities for events not of the form $X \leq t$ for examples,

$$P(X = t) = F(t) - \operatorname*{limit}_{x \to t^-} F(x),$$

and

$$P(a \leq X \leq b) = F(b) - \operatorname*{limit}_{x \to a^-} F(x).$$

Every distribution function has a unique decomposition of the form,

$$F = w_{ac}F_{ac} + w_s F_s + w_d F_d, \tag{2}$$

where $w_{ac} + w_s + w_d = 1$, $0 \leq w_{ac}, w_s, w_d$. Hence F_{ac} is an *absolutely continuous* distribution function, that is, there exists a function f_{ac} such that

Random Variables, Table 1 Discrete distributions

Name	Probability function	Support	Characteristic function	Mean	Variance
(1) Degenerate (x_0)	1	$x \in \{x_0\}$	e^{itx_0}	x_0	0
(2) Bernoulli (p) $B(p)$	$f(0)=1-p$ $f(1)=p$	$x \in \{0, 1\} 0 \leq p \leq 1$	$1-p+pe^{it}$	P	$p(1 - p)$
(3) Binomial (n,p) $B(n,p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, \ldots, n\}$ n is a positive interger $0 \leq p \leq 1$	$(1 - p + pe^{it})^n$	Np	$np(1 - np)$
(4) Poisson (λ) $P(\lambda)$	$e^{-\lambda}\lambda^x/x!$	$x \in \{0, 1, \ldots\}$ $0 \leq \lambda$	$e^{-\lambda(1-e^{it})}$	λ	λ
(5) Geometric (p)	$(1-p)^{x-1}p$	$x \in \{1, 2, \ldots\}$ $0 < p \leq 1$	$\frac{pe^{it}}{1-(1-p)e^{it}}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
(6) Uniform (a) $U(a)$	$1/a$	$x \in \{1, 2, \ldots, n\}$ a is a positive integer	$\frac{e^{it}(1-e^{ita})}{a(1-e^{it})}$	$\frac{1+a}{2}$	$\frac{a^2-1}{12}$
7) Hypergeometric	$\frac{\binom{n}{x}\binom{m}{r-x}}{\binom{m+n}{r}}$	$x \in \{0, 1, \ldots \min(m, n)\}$ n, m, r non – negative integers with $r \leq m + n$	$\left(\sum p_j e^{itj}\right)^n$	$\frac{rn}{m+n}$	$\left(\frac{m+n-r}{m+n-1}\right) r \frac{mn}{(m+n)^2}$
8) Multinomial	$n! \prod_{j=1}^{J} \left(p_j^{xj}/x_j!\right)$	$x_j = 0.1, \ldots$ $\sum x_j = n$ $p_j \geq 0, \quad j = 1, \ldots, J$ $\sum p_j = 1$ $\left(\sum p_j e^{it_j}\right)^n$	$\left(\sum p_j e^{it_j}\right)^n$	$EX_j = np_j,$ $j = 1, \ldots, J$	$\sigma_j^2 = np_j(1-p_j)$ $\sigma_{jk} = -np_j p_k$ $j \neq k$

$$F_{ac}(t) = \int_{-\infty}^{t} f_{ac}(x)dx \quad -\infty < t < \infty. \tag{3}$$

The function f_{ac} is a *density*. Notice

$$f_{ac} \geq 0 \quad \text{and} \quad \int f_{ac} = 1$$

The distribution function F_s is *singular* in that although it is not identically zero, its derivative exists and is zero almost everywhere. The distribution function F_d is *discrete*, that is, it is a right continuous step function with at most countably many jumps at $\{t_i\}$. The *probability function* f_d for Fd is zero everywhere except on $\{t_i\}$ where

$$f(t_i) = F(t_i) - \operatorname*{limit}_{x \to t_i^-} F(x).$$

The most common situations are the *discrete distributions* ($w_d = 1$) and the (*absolutely*) *continuous distributions* ($w_{ac} = 1$) In the discrete case the most common situation is the non-negative integer lattice, that is $\{t_i\} = \{i\}$ where the range of i is a set of non-negative integers, see Table 1.

In Table 1, x is used to designate a value of the random variable. All values that have a positive probability of occurrence are called the *support*. Also found in the 'support' column are specific restrictions on the constants or parameters of the probability function. In Table 1

$$A! = \Gamma(A+1) = \int_0^{\infty} x^A e^{-x} dx, \quad A \geq -1, \tag{4}$$

and when A is a non-negative integer,

$$\binom{B}{A} = \frac{B(B-1)\cdots(B-A+1)}{A!}. \tag{5}$$

The name of a random variable or of a distribution, for example $B(p)$, is used also to represent the random variable. The symbol $\sim$ between two random variables means that they have the same distribution function, and in the case of discrete

R

Random Variables, Table 2 Continuous distribution

Name	Density	Support	Characteristic function	Mean	Variance
(1) Exponential	e^{-x}	$x \in (0, \infty)$	$1/(1-it)$	1	1
(2) Logistic	$\frac{e^{-x}}{(1+e^{-x})^2}$	$x \in (-\infty, \infty)$	$\pi t \operatorname{cosech} \pi t$	0	$\pi^2/3$
(3) Normal (0,1) $N(0,1)$	$\frac{1}{(2\pi)^{1/2}} e^{-x^2/2}$	$-\infty < x < \infty$	$e^{-t^2/2}$	0	1
(4) Uniform (0,1) $U(0,1)$	1	$x \in (0, 1)$	$(e^{it}-1)/it$	½	1/12
(5) Chi-Square (n) $\chi^2(n)$	$\frac{1}{2^{n/2}\Gamma(n/2)} e^{-x/2} x^{(n/2)-1}$	$x \in (0, \infty)$ $n \in (0, \infty)$	$(1-2it)^{-n/2}$	n	$2n$
(6) Cauchy	$[\pi(1+x^2)]^{-1}$	$x \in (-\infty, \infty)$	$e^{-\lvert t\rvert}$	Sec(15)	Sec(15)
(7) Student's $t(n)$ $t(n)$	$\frac{1}{n^{1/2} B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-[(n+1)/2]}$	$x \in (-\infty, \infty)$ $n \in (0, \infty)$		0 if $n > 1$	$n/(n-2)$ if $n > 2$
(8) Fisher's F(m,n) F(m,n)	$C \frac{x^{(m-2)/2}}{(n+mx)^{(n+m)/2}}$ with $C = \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)}$	$x \in (0, \infty)$ $m \in (0, \infty)$ $n \in (0, \infty)$		$\frac{n}{n-2}$ if $n > 2$	$\frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}$ if $n > 4$
(9) Beta (a,b) $\beta(a,b)$	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$	$x \in (0, 1)$ $a, b \in (0, \infty)$		$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
(10) Lognormal	$(2\pi x^2)^{1/2} \exp\left[-(\ln x)^2/2\right]$	$x \in (0, \infty)$		$e^{1/2}$	$e^2 - e$
(11) Extreme value	$e^{-x-e^{-x}}$	$x \in (-\infty, \infty)$	$\Gamma(1-it)$	0.577	$\pi^2/6$
(12) Weibull	$cx^{c-1}e^{-x^c}$	$x \in (0, \infty)$		$\Gamma(c^{-1}+1)$	$\Gamma(2c^{-1}+1) - [\Gamma(c^{-1}+1)]^2$
(13) Bivariate normal	See (59)				

random variables, that they have the same probability function. Notice

$$
\begin{aligned}
&B(1) \sim U(1) \sim \text{Degenerate}(1) \sim \text{Geometric}(1)\\
&B(p) \sim B(1, p),\\
&B(n, 0) \sim P(0) \sim \mathit{Degenerate}(0),\\
&B(n, p) \sim n - B(n, 1-p).
\end{aligned}
$$

In Table 2

$$
B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{6}
$$

For the continuous distributions (Table 2), there are many interesting variations which arise from transformations (12). For example, $\sigma\, N(0, 1) + \mu$ with $\sigma \geq 0$ defines a normal random variable with expectation μ and variance σ^2, that is, $N(\mu, \sigma^2) \sim \sigma N(0, 1) + \mu$. The 3-parameter log-normal is obtained from Table 2(10) with the transformation $x = [(y-\alpha)/\beta]^{\gamma}, \gamma > \alpha, \beta > 0, \gamma > 0$ and the real root is used. Two useful connections between discrete and continuous random variables are

$$
P[B(n, p) \geq k] = P[\beta(k, n-k+1) \leq p] \quad \text{for } k = 0, \ldots, n, \quad 0 \leq p \leq 1, \tag{7}
$$

and

$$
\cdot P[P(\lambda) < a] = P[\chi^2(2a) \geq 2\lambda] \quad \text{for } a = 1, 2, \ldots \quad \text{and } \lambda > 0. \tag{8}
$$

Elementary Manipulations of Random Variables

An absolutely continuous distribution does not have a unique density; the value of an integral is not changed if the integrand is changed on a countable number of points or a set of Lebesgue measure 0.

For $0 < \alpha < 1$, the α-*percentile* of a random variable X, denoted by x_α, is defined by

$$x_\alpha = \inf\{x : F(x) \geq \alpha\} \tag{9}$$

Deciles, quartiles and medians (50-percentiles) are special cases.

If $X = X(\omega)$ is a random variable, then for any (measurable) function $g(\cdot)$ one obtains the random variable

$$Y = Y(\omega) = g[X(\omega)] = g(X)$$

with

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P[g(X) \leq y] = \int_{\{x;g(x)\leq y\}} dF_x(x). \end{aligned} \tag{10}$$

where F_X is the distribution function of X. In the discrete case, the $\int$ is a summation and $\mathrm{d}F(x)$ is the probability function. In the absolutely continuous case, $\mathrm{d}F(x)$ becomes $f(x)\mathrm{d}x$ where f is a density function.

If $g(\cdot)$ is a strictly increasing function,

$$F_Y(y) = F_X\left[g^{-1}(y)\right], \tag{11}$$

and further, in the absolutely continuous case when g', the derivative of g, is assumed to exist, then

$$f_Y(y) = f_x\left[g^{-1}(y)\right]\left[\mathrm{d}g^{-1}(y)/\mathrm{d}y\right]. \tag{12}$$

If $g(x) = \mu + \sigma x$, and $\sigma > 0$, then

$$f_Y(y) = \sigma^{-1} f_X[(y-\mu)/\sigma]. \tag{13}$$

In this monotone increasing case

$$y_\alpha = g^{-1}(x_\alpha) \tag{14}$$

Notice, if g is strictly decreasing, then in (12) replace $(g^{-1})'$ by $\left|(g^{-1})'\right|$ and (14) becomes

$$y_\alpha = g^{-1}(x_{1-\alpha}). \tag{15}$$

The *expected value* of X, denoted by EX, is defined by

$$\begin{aligned} EX = &\int_{-\infty}^{\infty} x\,\mathrm{d}F(x) \\ &\times < \infty\text{provided} \int_{-\infty}^{\infty} |x|\mathrm{d}F(x) < \infty; \end{aligned} \tag{16}$$

this last condition is sometimes relaxed to either

$$-\int_{\infty}^{0} x\mathrm{d}F(x) < \infty \quad \text{or} \quad \int_{0}^{\infty} x\mathrm{d}F(x) < \infty.$$

In the discrete case,

$$EX = \sum xf(x), \tag{17}$$

and in the absolutely continuous case,

$$EX = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x. \tag{18}$$

If $Y = g(X)$, then

$$EY = \int_{-\infty}^{\infty} g(x)\mathrm{d}F_X(x). \tag{19}$$

So we write $Eg(X)$ and it is not necessary to compute $F_Y(y)$.

Assume $g(x)$ is *convex* on an interval I which includes the support of X; that is, for x_1 and $x_2 \in I$ and $\alpha \in [0, 1]$,

$$\begin{aligned} g[\alpha x_1 + (1-\alpha)x_2] \leq \alpha g(x_1) \\ + (1-\alpha)g(x_2). \end{aligned} \tag{20}$$

(If $g''(x) \geq 0$ for $x \in I$, then g is convex.) Then *Jensen's inequality* is:

$$Eg(X) \geq g(EX), \tag{21}$$

provided both expected values exist. In particular,

$$EX^2 \geq (EX)^2, \tag{22}$$

so the *variance*, denoted by σ^2 and defined by

$$\sigma^2 = EX^2 - (EX)^2, \tag{23}$$

satisfies $\sigma^2 \geqslant 0$ An equivalent definition is

$$\sigma^2 = E(X - EX)^2. \tag{24}$$

A continuous distribution is *unimodal* with mode x_0 if $F(x)$ is convex for $x \in (-\infty, x_0)$, and $1 - F(x)$ is convex for $x_0 \in (x,\infty)$ The discrete X is *unimodal* if there is a unique x which maximizes its probability function. The discrete examples (except for the uniform), for most choices of the parameters are unimodal, and when the probability function has several maximizing values, they are contiguous.

Following are results and definitions for often-used expectations, if they exist (16):

$$\mathit{Mean\ of\ } X = \mu = EX. \tag{25}$$

$$k\mathit{th\ moment\ of\ } X = \alpha_k = EX^k, \text{ so } \mu = \alpha_1. \tag{26}$$

$$k\mathit{th\ absolute\ moment\ of\ } X = E|X|^k. \tag{27}$$

$$k\mathit{th\ central\ moment\ of\ } X = \mu_k = E(X-\mu)^k, \\ \text{so that } \sigma^2 = \mu_2, \text{ see } (24). \tag{28}$$

$$E\{N(0,1)\}^{2k+1} = 0,$$

and

$$E\{N(0,1)\}^{2k} = (2k)!/k!2^k \text{ for } k, \\ \text{a non} - \text{negative integer} \tag{29}$$

Characteristic function of X, $\phi_X(t)$ or $\phi(t)$ is defined by

$$\phi(t) = Ee^{itX}. \tag{30}$$

Always $\phi(0) = 1$, $|\phi(t)| \leq 1$, ϕ is uniformly continuous in t, and there is a one-to-one correspondence between distribution functions and characteristic functions. This correspondence is made explicit by

The *inversion formula*: If ϕ is the characteristic function of X, then

$$F(x+\Delta) - F(x-\Delta) \\ = \lim_{T\to\infty} \int_{-T}^{T} \frac{1}{\pi} \frac{\sin t\Delta}{t} \mathrm{e}^{-itx} \phi(t) \mathrm{d}t \tag{31}$$

provided $x - \Delta$ and $x + \Delta$ are continuity points of F.

$\phi(t)$ is a characteristic function if and only if ϕ is continuous, $\phi(0) = 1$ and ϕ is non-negative definite; that is, for all $n(\geq 1), t_1, \ldots, t_n$, and $h_1, \ldots, h_n$

$$\sum_{j=1}^{n} \sum_{k=1}^{n} h_j \phi(t_j - t_k) \overline{h}_k \geq 0, \tag{32}$$

where $\overline{h}$ is the complex conjugate of h.

The characteristic function of X is real if and only if

$$X \sim -X; \tag{33}$$

that is, X has a symmetric (about 0) distribution.

If $Y = aX + b$, then

$$\phi_Y(t) = e^{ibt} \phi_X(at). \tag{34}$$

If X and Y are independent (50), then

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t). \tag{35}$$

The *moment generating function of X* is $m(t) = Ee^{tX}$ and

$$\ln m(t) \mathit{\ is\ the\ cumulant\ generating\ function}. \tag{36}$$

If X has support on the non-negative integers, then the *probability generating function of X* is

$$\theta(t) = Et^X, \qquad \text{for } 0 \leq t \leq 1,$$

and

$$P(X=k) = \frac{1}{k!} \frac{\mathrm{d}^k \theta(t)}{dt^k}\bigg|_{t-0}, \qquad \text{for } k = 1, 2, \ldots \tag{37}$$

If the kth moment of X exists, then

$$EX^k = d^k m(t)/dt^k\big|_{t=0} = i^{-k} d^k \phi(t)/dt^k\big|_{t=0}, \quad \text{for } k = 1, 2, \ldots \tag{38}$$

The *kth cumulant* is the kth derivative of the cumulant generating function evaluated at 0. For $P(\lambda)$ the cumulant generating function is $\lambda(e^t - 1)$ so all the $P(\lambda)$ cumulants are λ. In general, the cumulants can be expressed in terms of the central moments and vice versa.

Many inequalities can be obtained from the *Markov inequality*:

$$\text{If } P[g(X) \geq 0] = 1 \text{ and } A > 0, \quad \text{then } P[g(X) \geq A] \leq Eg(X)/A. \tag{39}$$

The *Chebychev inequality* is one consequence of

$$P(|X - \mu| \geq \lambda) \leq \frac{E|X - \mu|^p}{\lambda^p}, \quad \lambda > 0 \text{ and } \quad p > 0, \tag{40}$$

and another is the *Bernstein inequality*: If $x \geq 0$ then

$$P(X \geq x) \leq \inf_{t \geq 0} [\mathrm{e}^{-xt} m(t)]. \tag{41}$$

Several Random Variables

To this point we have mentioned the possibility of several random variables being defined on the same Ω. Now the discussion will focus on two random variables, say $X_1 = X_1\ (\omega)$ and $X_2 = X_2\ (\omega)$, for example, X_1 could be Income and X_2 could be Savings. So if ω_0 is an individual, then $[X_1\ (\omega_0), X_2\ (\omega_0)]$ is the income and savings of ω_0.

The *distribution function* is defined by

$$F(x_1, x_2) = F_{x_1, x_2}(x_1, x_2) = P(X_1 \leq x_1 \text{ and } X_2 \leq x_2) \tag{42}$$

Notice,

$$\underset{x_1 \to -\infty}{\text{limit}} F(x_1, x_2) = \underset{x_2 \to -\infty}{\text{limit}} F(x_1, x_2) = 0, \tag{43}$$

$$\begin{aligned} F_{X_1}(x_1) &= \underset{x_2 \to \infty}{\text{limit}} F(x_1, x_2), \\ F_{X_2}(x_2) &= \underset{x_1 \to \infty}{\text{limit}} F(x_1, x_2), \end{aligned} \tag{44}$$

$$\begin{aligned} &F(x_1 + \Delta, x_2 + \varepsilon) - F(x_1, x_2 + \varepsilon) \\ &-F(x_1 + \Delta, x_2) + F(x_1, x_2) \\ &= P(x_1 < X_1 \leq x_1 + \Delta \text{ and } x_2 < X_2 \leq x_2 + \varepsilon) \\ &\geq 0, \end{aligned} \tag{45}$$

when $\Delta \geq 0$ and $\varepsilon \geq 0$, and

$$F \text{ is right continuous in each of its arguments.} \tag{46}$$

Conditions (43), (44), (45) and (46) are necessary and sufficient for F to be a distribution function; (45) is the analogue of $F(x)$ being an increasing function in the univariate case (1). The discrete bivariate distributions offer no surprises when compared to the univariate case, but there are many special cases of mixed continuous and discrete situations. And there are simple continuous examples where nothing like a density could exist, such as

$$F(x, y) = \begin{cases} 0 & \text{for } x + y \leq 1 \\ \min(x, 1) + \min(y, 1) - 1, & \text{for } x + y \geq 1 \end{cases} \tag{47}$$

The random variables X and Y, are said to have the *joint density, f(x, y)*, provided

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \mathrm{d}s\, \mathrm{d}t, \quad \text{for} -\infty < x < \infty \quad \text{and} \quad -\infty < y < \infty. \tag{48}$$

When continuous random variables are discussed, the usual meaning will be this absolutely continuous situation with a (joint) density.

The random variables X and Y are *independent* if and only if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \quad (49)$$

for all measurable sets A and B. A necessary and sufficient condition for the independence of X and Y, is

$$F(x,y) = F_X(x)F_Y(y)$$
$$\text{for } -\infty < x < \infty \quad \text{and} \quad -\infty < y < \infty; \quad (50)$$

when there is a density or probability function, (50) is equivalent to

$$f(x,y) = f_X(x)f_Y(y)$$
$$\text{for } -\infty < x < \infty \quad \text{and} \quad -\infty < y < \infty. \quad (51)$$

The expected value of a function of several random variables, say $Z = g(X, Y)$, can be found from

$$EZ = Eg(X,Y) = \int\int g(x,y)\mathrm{d}F(x,y). \quad (52)$$

Always,

$$E(aX + bY) = aEX + bEY, \quad (53)$$

and if X and Y are independent,

$$EXY = (EX)(EY), \quad (54)$$

provided both EX and EY exist. In these formulas one can replace X by $r(X)$ and Y by $s(Y)$ to obtain results, such as, (35) and (24). The *covariance* is defined by

$$\sigma_{X,Y} = E(X - \mu_X)(Y - \mu_Y) = EXY - \mu_X\mu_Y, \quad (55)$$

and the Pearson product moment *correlation* is defined by,

$$\rho_{X,Y} = \sigma_{X,Y}/\sigma_X\sigma_Y \quad (56)$$

The Cauchy–Schwarz inequality is equivalent to $-1 \le \rho \le 1$. If X and Y are independent, then $\rho = 0$, but not conversely.

Assume the equations

$$u = u(x,y) \quad \text{and} \quad \upsilon = \upsilon(x,y) \quad (57)$$

have a unique inverse; that is,

$$x = x(u,\upsilon) \quad \text{and} \quad y = y(u,\upsilon) \quad (58)$$

and the *Jacobian* determinant,

$$J(x,y;u,\upsilon) = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial \upsilon} & \frac{\partial y}{\partial \upsilon} \end{pmatrix}, \quad (59)$$

exists, is continuous and is never equal to 0. Then

$$f_{U,V}(u,\upsilon) = |J(x,y;u,\upsilon)|f_{X,Y}[x(u,\upsilon),y(u,\upsilon)]. \quad (60)$$

The above conditions are sufficient and in applications, even if they fail, a little analysis might show that (60) still holds. The linear transformation

$$Y_1 = \mu_1 + \sigma_1 X_1 \text{and} Y_2 = \mu_2 + \sigma_2\left[\rho X_1 + (q - \rho^2)^{1/2} X_2\right]$$

of X_1 and X_2, two independent $N(0, 1)$ variables, yields

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi}\frac{1}{\sigma_1\sigma_2(1-\rho^2)^{1/2}}\exp\left\{\frac{-1}{2(1-\rho^2)\sigma_1^2\sigma_2^2}\left[\sigma_2^2(y_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(y_1-\mu_1)(y_2-\mu_2) + \sigma_1^2(y_2-\mu_2)^2\right]\right\}, \quad (61)$$

or

$$(Y_1, Y_2) \sim N\left(\mu, \sum\right). \tag{62}$$

In (62) $N(\mu, \Sigma)$ is a *bivariate normal* random variable with mean vector μ and variance–covariance matrix Σ where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \sum = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

From the general bivariate normal density (61), one can show $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$. The random variables Y_1 and Y_2 are independent if and only if $\rho = 0$.

To finish this introduction to bivariate distributions, we introduce the conditional probability of A given B,

$$P(A|B) = \frac{P(A \,\text{and}\, B)}{P(B)}, \tag{63}$$

provided $P(B) > 0$. When X and Y are discrete random variables (defined on the same sample space), the conditional probability function of X given Y, is defined accordingly

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \tag{64}$$

provided $f_Y(y) > 0$. For each y-value, $f_{X|Y}$ will be a probability function, and conditional expectations are defined by

$$E(X|Y = y) = \sum x f_{X|Y}(x,y). \tag{65}$$

The function in (65) has y as its argument. If $y = Y(\omega)$ then the conditional expectation becomes a random variable denoted by

$EX|Y$. The conditional density may be defined as in (64) for the continuous case as well. For the bivariate normal (61), the conditional distribution of Y_1 given $Y_2 = y_2$ is

$$N\left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y_2 - \mu_2), \left(1 - \rho^2\right)\sigma_1^2\right] \tag{66}$$

Conditioning on Y_2 reduces the variance of Y_1 and makes the expected value of Y_1 depend linearly on the specified value of Y_2 when $\rho \neq 0$.

Asymptotic Theorems for Sums of Independent Random Variables

A random variable X with expectation μ is often observed near μ. And if there are several random variables, $X_1, \ldots, X_n$ each with expectation μ, then their average, $\overline{X}_n = \sum X_i/n$, should even be closer to μ. Now consider a sequence of random variables $X_1, \ldots, X_n, \ldots$, which are independent each with the same distribution and $EX_i = \mu$.

Weak Law of Large Numbers:

$$\operatorname*{limit}_{n\to\infty} p\left(|\overline{X} - \mu| < \varepsilon\right) = 1, \text{for every } \varepsilon > 0. \tag{67}$$

Strong Law of Large Numbers:

$$p\left(\operatorname*{limit}_{n\to\infty} \overline{X}_n = \mu\right) = 1 \tag{68}$$

The mode of convergence in (67) is *in probability* or *weak*, and the mode of convergence in (68) is *with probability one* or *strong*. The difference between weak and strong statements is shown in the following example, $\bar{A}$ is the complement of A, and $\text{mod}_1\, a = r$ where $a = a' + r$ with a' an integer and $0 \leq r < 1$. Let:

$$Z_n(\omega) = \begin{cases} 0, \text{if } \omega \in \overline{A}_n \\ 1, \text{if } \omega \in A_n \end{cases}. \tag{69}$$

where $\Omega = (0,1)$, $A_n = \left(\text{mod}_1 \sum_1^n 1/i, \text{mod}_1 \sum_1^{n+1} 1/i\right)$ and the probability of a set in Ω is the length of the set. Then $P(A_n) \to 0$ so that Z_n converges in probability to 0. Also, every ω is in an infinite number of A_n so that Z_n does *not* converge with probability one.

Strong convergence implies weak convergence; when the Strong Law of Large Numbers applies, so will the Weak Law of Large Numbers. A sequence of random variables Z_n is said to converge in *mean square* to μ if and only if

$$E(Z_n - \mu)^2 \to 0. \quad (70)$$

Mean square convergence implies weak convergence. A consequence of the Chebychev inequality is:

Assume for each i and j that $EX_i = \mu, \sigma_{X_i,X_j} = \sigma_{ij}$

$$\text{and} \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}/n^2 \to 0; \ \text{then} \overline{X}_n \to \mu \text{ in probability.} \quad (71)$$

Another type of convergence is *in distribution*. The sequence of random variables X_1, X_2, ...,X_n,... converge in distribution to X if and only if

$$F_X(x) = \operatorname*{limit}_{n\to\infty} F_{X_n}(x)$$

at all continuity points of F_X. It is expressed as $X_n \rightsquigarrow X$. Note X and X_n can be defined on different sample spaces Ω and Ω_n so $P(X \in A, X_n \in B)$ would not be defined. The symbol $X–X_n$ requires X and X_n to be defined on the same space. Thus the conclusion of (67) is $X_n \rightsquigarrow$ degenerate (μ). Note $B(n, \lambda/n) \rightsquigarrow P(\lambda)$ for fixed λ.

Also, $X_n a \rightsquigarrow Y$ means there are sequences $\{a_n\}$ and $\{b_n\}$ such that $a_n X_n + b_n \rightsquigarrow Y$.. The assumptions of (72) imply $\overline{X}_n a \rightsquigarrow N(0,1)$, $\overline{X}_n a \rightsquigarrow \mu$, $\overline{X}_n a \rightsquigarrow 0$. Read $X_n \rightsquigarrow Y$ as 'the limiting distribution of $\{X_n\}$ is the distribution of Y' or 'X_n converges to Y in distribution'. And read $X_n a \rightsquigarrow Y$ as 'the *asymptotic* distribution of $\{X_n\}$ is the distribution of Y'.

Central Limit Theorem. If each X_i is independent and has the same distribution with finite mean μ and finite variance σ^2, then

$$n^{1/2}(\overline{X}_n - \mu)/\sigma \rightsquigarrow N(0,1). \quad (72)$$

The Weak Law of Large Numbers probabilistically says that $(\overline{X}_n - \mu)$ is small, but the Central Limit Theorem gives a much stronger statement since it implies that $n^{1/2}(X_n - \mu)$ is not large in a probabilistic sense.

Berry–Essen Theorem. Assume and define:

(a) $X_1, \ldots, X_n$ are independent, and $EX_i = 0$.

(b) For each i write $\sigma^2_{Xi} = \sigma^2_i$.

(c) $$S_n^2 = \sum_{i=t}^{n} \sigma_i^2 > 0.$$

(d) For each i, $\gamma_i = E|X_i|^\Delta < \infty$, for some $2 < \Delta \leq 3$.

(e) $$\Gamma_n^\Delta = \sum_{i=1}^{n} \gamma_i. \quad (73)$$

Then there exists constants $C_\Delta (C_3 \leq 7.5)$ such that

$$\max_{-\infty<x<\infty} \left| P\left[n\overline{X}_n/s_n < x\right] - \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} \mathrm{e}^{-t^2/2} \mathrm{d}t \right| \leq C_\Delta (\Gamma_n/s_n)^\Delta.$$

When the random variables are identically distributed, $C_\Delta(\Gamma_n/s_n)^\Delta$ is proportional to $n^{[(2-\Delta)/2]}$ which might be small enough to give useful bounds.

Some tools useful in proving limit theorems are: (31) and (35).

Consider a sequence of distribution functions $\{F_i\}$ and the corresponding sequence of characteristic functions $\{\phi_i\}$. Assume there exists a function $\phi(t)$ which is continuous at 0 and

$$\phi(t) = \operatorname*{limit}_{i\to\infty} \phi_i(t) \text{ for every } t.$$

Then

(a) $\phi(t)$ is a characteristic function, and if $F_i(x)$ is the associated distribution function,

(b) $F(x) = \operatorname*{limit}_{i\to\infty} F_i(x)$ at every x which is a continuity point of F. (74)

$$(X_{n1}, X_{n2}) \rightsquigarrow (Y_1, Y_2) \text{ if and only if } t_1 X_{n1} + t_2 X_{n2} \rightsquigarrow t_1 Y_1 + t_2 Y_2 \tag{75}$$

for each pair of real numbers (t_1, t_2).

Slutsky's Theorem. If for each n, the random variables X_n, Y_n, Z_n are defined on the same sample space, and

$$\begin{aligned} &(a)\ X_n \rightsquigarrow X, \\ &(b)\ Y_n \rightsquigarrow a, \\ &(c)\ Z_n \rightsquigarrow b; \quad \text{then} \quad X_n Y_n + Z_n \rightsquigarrow aX + b. \end{aligned} \tag{76}$$

Propagation of Error. Assume $\left[n^{1/2}(X_n - \mu_x), n^{1/2}(Y_n - \mu_y)\right]$ has a bivariate normal limiting distribution with mean vector (0, 0) and variance–covariance matrix

$$\begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY} \end{pmatrix},$$

and $H(x, y)$ has continuous first derivatives,

$$\left[H_x(x,y) = \frac{\partial H(x,y)}{\partial x}, H_y(x,y) = \frac{\partial H(x,y)}{\partial y}\right]$$

in the neighbourhood of (μ_X, μ_Y); then

$$\begin{aligned} {}^{n1/2}[H(X_n, Y_n) - H(\mu_X, \mu_Y)] &\rightsquigarrow N\left[0, H_x^2(\mu_X, \mu_Y)\sigma_X^2\right. \\ &+ 2H_x(\mu_X, \mu_Y)H_Y(\mu_X, \mu_Y)\sigma_{XY} \\ &\left.+ H_Y^2(\mu_X, \mu_Y)\sigma_Y^2\right], \end{aligned} \tag{77}$$

provided this variance is > 0.

Although the limiting distribution of $n^{1/2}[H(X_n, Y_n) - H(\mu_X, \mu_Y)]$ exists and has finite moments, the moments of H (X_n, Y_n) may not exist; for example, let $H(x) = 1/x$ and X_n be the average of n independent $P(\lambda)$ variables. Then $H\ (EX_n) = 1/\ \lambda$ and $EH\ (X_n) = \infty$ for every n.

Example: if

$$X_n \sim B(n,p) \text{then } n1/2(X_n/n - p) \rightsquigarrow N\big(0, p(1-p)\big];$$

from (72) where $X_n = \sum_{i=1}^{n} Y_i$, with the Y_i independent Bernoulli (p) random variables. Now consider $H(x) = \arcsin\,(x)^{1/2}$ so that $H_x(x) = 1/\left\{2[x(1-x)]^{1/2}\right\}$, and thus

$$n^{1/2}\left[\arcsin(X_n/n)^{1/2} - \arcsin(p)\right] \rightsquigarrow N(0, 1/4).$$

The transformation is *variance stabilizing*.

There are many other classes of limit theorems, such as,

Law of the Iterated Logarithm: Assume

(a) $X_1, \ldots, X_n, \ldots$ are mutually independent with the same distribution.
(b) $EX_1 = 0$.
(c) $EX_1^2 = 1$. (79)

Then

$$P\left\{\overline{\lim}\, n \to \infty \frac{\sum_{i=1}^{n} X_i}{[n\ln(\ln n)]^{1/2}} = (2)^{1/2}\right\} = 1.$$

Distributions Related to the Normal

Assume X, $X_1, \ldots X_2, \ldots, X_i, \ldots$ are independent $N(0, 1)$; then the following definitions, examples and theorems apply:

$$\overline{X}_n = \sum_1^n X_i/n \text{ and}$$

$$S_n^2 = \sum_1^n (X_i - \overline{X}_n)^2/(n-1).$$

are independent, and

$$\overline{X}_n \sim N(0, 1/n), \quad S_n^2 \sim (n-1)^{-1}\chi^2(n-1). \tag{80}$$

A more general result is: $\overline{X}_n$ and D_0 ($X_1, \ldots, X_n$) are independent provided D_n is translation invariant; that is, $D_n(X_1, \ldots, X_n) = D_n(X_1 + a, \ldots, X_n + a)$ for every a.

R

$$\sum_{i=1}^{r} X_i^2 \sim \chi^2(r) \text{ with } r \text{ a positive integer.} \quad (81)$$

If $\chi^2(r)$ and $\chi(s)^2$ are independent, then

$$\chi^2(r) + \chi^2(s) \sim \chi^2(r+s). \quad (82)$$

If X and $\chi^2(r)$ are independent, then

$$t = X/\left[\chi^2(r)/r\right]^{1/2} \quad (83)$$

has a t-distribution with r degrees of freedom, $t(r)$.

Student's Theorem: If $Y_1,\ldots,Y_n$ are independent $N(\mu, \sigma^2)$ then

$$\frac{n^{1/2}\left(\overline{Y}_n - \mu\right)^{1/2}}{\left[\frac{\sum\left(Y_i - \overline{Y}_n\right)^2}{n-1}\right]} \sim t(n-1). \quad (84)$$

$$\frac{\chi^2(n) - n}{(2n)^{1/2}} \rightsquigarrow N(0,1). \quad (85)$$

$$t(n) \rightsquigarrow N(0,1). \quad (86)$$

If $\chi^2(r)$ and $\chi^2(s)$ are independent, then

(a) $\beta = \frac{\chi^2(r)}{\chi^2(r)+\chi^2(s)}$ has a $\beta(r,s)$ distribution, and

(b)

$$F = \frac{[\chi^2(r)/r]}{\chi^2(s)/s} \text{ has an } F - \text{distribution with } r \text{ and } s \text{ degrees of freedom, } F(r,s). \quad (87)$$

$$t^2(n) \sim F(1,n). \quad (88)$$

$$[F(r,s)]^{-1} \sim F(s,r). \quad (89)$$

$$rF(r,s) \underset{s\to\infty}{\rightsquigarrow} \chi^2(r). \quad (90)$$

Under the conditions of (83), if $n = n_1 + n_2$, then

$$S_1^2/S_2^2 \sim F(n_1 - 1, n_2 - 1),$$

where

$$S_1^2 = \sum_1^{n_1}\left[Y_i - \left(\sum_1^{n_1} Y_i/n_1\right)\right]^2/(n_1 - 1).$$

and

$$S_2^2 = \sum_{n+1}^{n}\left[Y_i - \left(\sum_{n+1}^{n} Y_i/n_2\right)\right]^2/(n_2 - 1). \quad (91)$$

Order Statistics

In this section assume X, $X_1,\ldots, X_n$ are independent real-valued random variables; each X has density f and distribution function F. Further, let $\{X_i\}$, when arranged in order from smallest to largest, be denoted by $X_{(1)},\ldots,X_{(n)}$.

For $1 \leq r \leq n$, the density of $X(r)$ is

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}F^{r-1}(x) \times [1 - F(x)]^{n-r}f(x). \quad (92)$$

$$F_{X_{(n)}}(x) = F^n(x). \quad (93)$$

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n. \quad (94)$$

If $1 \leq r < s \leq n$, then

$$f_{X_{(r)},X_{(s)}}(x,y) = \frac{n!}{(r-1)!(r-s-1)!(n-s)!}F^{r-1}(x) \times [F(y) - F(x)]^{s-r-1}[1 - F(y)]^{n-s}f(x)f(y), \quad -\infty < x < y < \infty. \quad (95)$$

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(y_1, y_2, \ldots, y_n) = n!\prod_1^n f(y_i), \quad -\infty < y_1 < y_2\cdots < y_n < \infty. \quad (96)$$

Theorem: If

(a) f is continuous where $f = F'$.
(b) For specified α and β satisfying $0 < \alpha < \beta < 1$, define, x_α and x_β as the unique solutions of $F(x) = \alpha$ and $F(x) = \beta$, respectively.
(c) $0 < f(x_\alpha) < \infty$ and $0 < f(x_\beta) < \infty$
(d) $[m]$ is the largest integer not exceeding m.

Then,

$$n^{1/2}\left(X_{[\alpha n]} - x_\alpha, x_{[\beta n]} - x_\beta\right)$$
$$\rightsquigarrow N\left\{\begin{pmatrix}0\\0\end{pmatrix}, \begin{bmatrix}\dfrac{\alpha(1-\alpha)}{f^2(x_\alpha)} & \dfrac{\alpha(1-\beta)}{f(x_\alpha)f(x_\beta)}\\ \dfrac{\alpha(1-\beta)}{f(x_\alpha)f(x_\beta)} & \dfrac{\beta(1-\beta)}{f^2(x_\beta)}\end{bmatrix}\right\}. \tag{97}$$

Not all limit theorems involve sums of random variables (97). Theorem (95) appears to be an example, but consider the following:

$$P\left[X_{(k)} < x\right] = P\left\{\sum_1^n B_i[F(x)] \geq k\right\},$$

where $B_1, \ldots, B_n$ are independent, Bernoulli $[F(x)]$ and $k = 1, 2, \ldots, n$.

Renyi Representation. If f is exponential [Table 2(1)], then

$$X_{(i)} = \sum_{j=1}^{i} Y_j/(n - j + 1), \tag{99}$$

where the $\{y_j\}$ satisfy the same conditions as the $\{X_i\}$.

Extreme Value Theorem. If $a_n X_{(n)} + b_n a \rightsquigarrow Z$, then Z is one of three types, the principal ones being the Extreme Value [Table 2(11)] and Weibull [Table 2(12)]. In particular, if F is normal, then there exist sequences of constants $\{A_n\}$ and $\{B_n\}$ such that

$$A_n X_{(n)} + B_{(n)} \rightsquigarrow \text{extreme value [Table 2(11)]}. \tag{100}$$

Failure Rate (Hazard Rate, Force of Mortality, Intensity Rate)

Starting with Karl Pearson, families of distributions have been introduced to unify theory or applications. The exponential family (107) plays a central role in theoretical statistics, and the characterization of failure rates is central to the description of lifetimes of organizations, animals, equipment, etc.

Assume F satisfies:

$$\text{(a) } F(0) = 0;\ \text{(b)} f(x) = F'(x) \text{ exists for all } x \geq 0.$$

Then define the *failure rate*, $r(t)$, by

$$r(t) = \frac{f(t)}{1 - F(t)} \text{ for all } t \text{ such that } 1 - F(t) > 0. \tag{101}$$

The following classes of distributions are non-empty since the exponential [Table 2(1)] belongs to each of them. The names of the classes are suggestive of their applied interest.

IFR (DFR) – Increasing (Decreasing) Failure Rate:

$$r(t) \text{ increases (decreases) for } t \geq 0. \tag{102}$$

IFRA(DFRA) – Increasing (Decreasing) Failure Rate Average:

$$\frac{1}{t}\int_0^t r(u)\mathrm{d}u \text{ increases (decreases) for all } t > 0. \tag{103}$$

NBU (NWU) – New Better (Worse) than Used:

$$[1 - F(x+y)] \leq (\geq)[1 - F(x)] \times [1 - F(y)] \text{ for all x} \geq 0, y \geq 0. \tag{104}$$

NBUE (NWUE) – New Better (Worse) than Used Expectation:

(a) μ
$$= \int_0^\infty xf(x)\mathrm{d}x = \int_0^\infty [1 - F(x)]\mathrm{d}x < \infty(\le \infty),$$
(b) $\int_t^\infty [1 - F(x)]\mathrm{d}x \le (\ge)\mu[1 - F(t)]\,\mathrm{for}\,t \ge 0.$
(105)

The classes obey the following inclusion relations:

$$\mathrm{IFR} \subset \mathrm{IFRA} \subset \mathrm{NBU} \subset \mathrm{NBUE},$$

and

$$\mathrm{DFR} \subset \mathrm{DFRA} \subset \mathrm{NWUC} \subset \mathrm{NWUE}. \quad (106)$$

A Weibull random variable [Table 2(12)] is IFR if $c \ge 1,$, and is DFR if $c \ge 1$ There are many analytic results associated with these classes, for example, if X and Y are independent and NBUE, then $Z \sim X + Y$ is NBUE; however, the analogous implication fails for NWU.

$$\textit{Exponential Family}. \quad (107)$$

Many of the examples of densities (or probability functions) are of the form:

$$\exp[a(x)b(\theta) + c(x) + d(\theta)]. \quad (108)$$

where x is a possible value of the random variable, and θ is a parameter.

Statistical Sufficiency. If $X_1, \ldots, X_n$ are independent each with the same density of form (106), and $T = \sum a(X_i)$, then

$$f_{X_1,\ldots,X_n\backslash T}(x_1, \ldots, X_n|T, \theta) \quad (109)$$

does not depend on θ that is, for purposes of making inferences about θ, all of the information in the sample, $\{X_i\}$, is in the sufficient statistic, $\sum a(X_i)$.

Theorem: If X has density (106), $b(\cdot)$ is one to one, $b(\theta)$ is an interior point of $\{b(\theta)$: $d(\theta)$ is finite$\}$ and the derivatives denoted by $'$ and $''$ below exist. Then

(a) $Ea(X) = -d'(\theta)/b'(\theta)$.

(b) The Fisher information, I_θ, for a family of densities, $f(, \theta)$, is defined by

$$I_\theta = E\left[\frac{\partial \ln f(X_1, \theta)}{\partial \theta}\right]^2.$$

When X has a density in the form of (106),

$$I_\theta = -d''(\theta) + d'(\theta)[b''(\theta)/b'(\theta)] \quad (110)$$

When, as often happens, $\mathrm{b}(\theta) = \theta$, the result becomes

$$I_\theta = -d''(\theta) = V[a(X)].$$
$$\textit{Infinitely Divisible and Stable Random Variables}. \quad (111)$$

A random variable, X, is *infinitely divisible* if and only if for every positive integer, n, there exists independent and identically distributed random variables, $\{X_{ni}\}$, such that $X \sim X_{n1} + X_{n2} + \ldots + X_{nn}$ Normal, Poisson, Cauchy and exponential random variables are infinitely divisible. A random variable, X, is *stable* if and only if for each choice of $a_1 \ge 0$ and $a_2 \ge 0$ there exists $a > 0$ and b such that $aX + b \sim a_1X_1 + a_2X_2$, where X_1 and X_2 are independent and $X \sim X_1 \sim X_2$. Stable random variables are unimodal, absolutely continuous and infinitely divisible. Normal random variables are the only stable random variables with finite variance; Cauchy random variables are stable.

The function, $\phi(t)$ is the characteristic function of an infinitely divisible random variable if and only if it has the form

$$\log\,\phi(t) = itA - \frac{\sigma^2}{2}t^2 + \int_{-\infty}^{0-}\left(\mathrm{e}^{itu} - 1 - \frac{itu}{1 + u^2}\right)$$
$$\mathrm{d}M(u) + \int_{0+}^\infty \left(\mathrm{e}^{itu} - 1 - \frac{itu}{1 + u^2}\right)\mathrm{d}N(u),$$
(112)

where $M\,(u)$, $N\,(u)$ and σ^2 satisfy:

(a) $M(u)$ and $N(u)$ are non-decreasing on $(-\infty, 0)$ and $(0, \infty)$.

(b) $M(-\infty) = N(\infty) = 0$.

(c) $\int_{-\varepsilon}^{0} u^2 \mathrm{d}M(u)$ and $\int_{0}^{\varepsilon} u^2 \mathrm{d}N(u)$ are finite for every $\varepsilon > 0$.

(d) $\sigma^2 \geq 0$.

The representation is unique. [Define log $\phi(t)$ so that log $\phi(0) = 0$]

For stable random variables (111) must have either

$$\begin{aligned}&\text{(a)}\quad \sigma^2 > 0, M(u) \equiv 0, \text{ and } N \equiv 0, \text{ or}\\ &\text{(b)}\quad \sigma^2 = 0, M(u) = C_1 |u|^{-\alpha} \text{ for } u < 0,\\ &N(u) = -C_2 u^{-\alpha} \text{ for } u > 0 \text{ where } 0 < \alpha < 2,\\ &C_1 \geq 0, C_2 \geq 0 \text{ and } C_1 + C_2 > 0.\end{aligned} \tag{113}$$

Conversely, functions satisfying these conditions are characteristic functions of stable random variables.

Mixtures and Exchangeability

Assume the distribution of X depends on the random parameter θ. When $\theta = \theta_0$ let the conditional density of X be $f(x\,|\theta_0)$. Now assume θ is a random variable with density $f(\theta)$, and the marginal density of X is

$$f_X(x) = \int f(x|\theta)\mathrm{d}F(\theta). \tag{114}$$

Since $f_x\ (x)$ is a weighted combination of the conditional densities of X, it is called a *mixture*.

Example: Assume X is Poisson (λ) and $\lambda\sim$ exponential, then

$$f_X(x) - \int_0^\infty \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}\mathrm{e}^{-\lambda} = 2^{-x},\ x = 0, 1, \ldots \tag{115}$$

An interpretation of this example is that the λ associated with an individual reflects his accident-proneness, while X is the number of accidents.

Example: Let I be income and θ be sex. Then it is plausible that I given θ has a log-normal distribution with parameters dependent onθ. The marginal distribution of income would be of interest. The $P(\theta =$ female) would have a simple sampling interpretation. The components of these examples appear in Bayesian models.

A condition, not as restrictive as independence and identical distribution of random variables, is

$$f_{X_1,\ldots X_n}(x_1, \ldots, x_n) \equiv f_{X_1,\ldots,X_n}\left[x_{\pi(1)}, \ldots, x_{\pi(n)}\right], \tag{117}$$

where the identity holds for all $n \geq 1$ amd all $n!$ permutations [$\pi(1), \ldots, \pi(n)$] of ($1, \ldots, n$). Random variables for which condition (118) holds are called *exchangeable random variables*. Note (118) is equivalent to the following generalization of (115):

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \int \prod_{i=1}^{n} f(x_i|\theta)\mathrm{d}f(\theta) \tag{118}$$

Now consider sequences of random variables, $\{X_i\}$, where in contrast to exchangeability the labels on the random variables, $\{i\}$, are important. Much of applied statistics is concerned with

$$X_a \sim \sum_{m=1}^{M} t_{am}\beta_m + \varepsilon_a,\ a = 1, \ldots, N, \tag{119}$$

where the $\{\varepsilon_a\}$ are independent and identically distributed. This is the regression model with M independent variables, $\{t_{am}\}$, and parameters $\{\beta_m\}$ Further complications involve structure on the errors.

(*Strictly*) *Stationary*: For every choice of k distinct integers $\{a_i\}$ and every integer t:

$$(X_{a1}, X_{a2}, \ldots, X_{ak}) \sim (X_{a1+t}, X_{a2+t}, \ldots, X_{ak+1}) \tag{121}$$

This implies, for every *a*, $X_a \sim X_0$ so the random variables are identically distributed but not necessarily independent.

Weakly Stationary:

(a) $EX_a = \mu$, and

(b) EX_a^2 is finite

$$\text{(c)} \quad \sigma_{a,b} = C(|a-b|). \tag{122}$$

where, specifically, the function C will be of the form

$$C(k) = \sigma^2 \int_{-\pi}^{\pi} \cos k\omega \, \mathrm{d}G(\omega), \quad \mathrm{k} = 1, 2, \ldots \tag{123}$$

In (124) G is a distribution function with support in $(-\pi, \pi)$, and $G(\omega) + G(-\omega) = 1$ at continuity points of G. If a sequence is weakly stationary, then there is a random variable X such that

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - X\right| > \varepsilon\right) = 0,$$

where $\varepsilon > 0$ and $\overline{X}_n = (X_1 + \ldots + X_n)/n$.

Brownian Motion or Wiener Processes

Now consider a random variable for each $t \in [0,\infty]$ that is, an uncountably infinite number of random variables. The notation will be $X(t, \omega)$ or $X(t)$ In the above discussion of n random variables, each ω gives an n-dimensional vector $[X_1(\omega), \ldots, X_n(\omega)]$. In the current situation, each ω gives a curve $X(t, \omega)$, $0 \le t$.

We will not consider continuous time stochastic processes, $X(t)$ for $0 \le t$, in general, but we will report results for *Brownian motion* or *Wiener processes*.

Definition of (standard) Brownian motion:

(a) $X(0) = 0$.
(b) The sample paths, $X(s, \omega)$ for $s \in [0, \infty]$ are continuous except for a set $A \in \Omega$ with $P(A) = 0$.
(c) If $0 \le t_1 < t_2 < \cdots < t_k$ then the random variables $X(t_i) - X(t_{i-1})$ are independent, and

$$X(t_i) - X(t_{i-1}) \sim N(0, t_i - t_{i-1}) \text{ for } i = 2, \ldots, k \tag{124}$$

Some properties of Brownian motion are:

$$\text{If } 0 \le s \le t, \text{then } [X(s), \ X(t)] \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} s & s \\ s & t \end{pmatrix}\right] \tag{125}$$

The sample paths are differentiable on a set of Lebesgue measure 0 with probability one.

If T replaces n and

$$X(T) \text{ replaces} \sum_{i=1}^{n} X_i, \text{ then (79) holds.} \tag{127}$$

For $a > 0$ define T_a as the least t with $X(t) \ge a$. Then

$$\begin{aligned} P(T_a \le t) &= 2P[X(t) \ge a] \\ &= P\left[t^{1/2}|N(0,1)| \ge a\right]. \end{aligned} \tag{128}$$

Assume $\mu < 0$ and define $W = \max_{0 \le t}[X(t) + \mu t]$ Then,

$$P(W \ge w) = \mathrm{e}^{2\mu w}, \quad w \ge 0. \tag{129}$$

Let T^* be the largest s such that $X(s) = 0$ for $0 \le t_0 \le^*$. Then,

$$P(T^* \le t_0) = (2/\pi) \arcsin(t_0/t^*)^{1/2} \text{ for } 0 \le t_0 \le t^*. \tag{130}$$

Assume $X_1, \ldots, X_n$ are independent, and each X_i has the same distribution, F. Define the empirical distribution, F_n, by

$$F_n(x) = (\text{number of } X_i \le x)/n \text{ for } -\infty < x < \infty. \tag{131}$$

The Glivenko–Cantelli Lemma asserts that

$$P\left[\underset{n\to\infty}{\text{limit}} \ \underset{-\infty<x<\infty}{\text{maximum}} |F_n(x) - F(x)| = 0\right] = 1. \tag{132}$$

Further, if F is continuous, then for large values of n,

$$B_n(t) = n^{1/2}\left\{F_n\left[F^{-1}(t)\right] - t\right\} \text{for } 0 \leq t \leq 1 \tag{133}$$

behaves asymptotically like a *Brownian bridge*,

$$B(t) = X(t) - tX(1), \quad 0 \leq t \leq 1, \tag{134}$$

where X is Brownian motion.

A consequence is

$$\underset{n\to\infty}{\text{limit}}\, P\left[\underset{-\infty<x<\infty}{\text{maximum}}\,(n)^{1/2}|F_n(x) - F(x)| \geq t\right] = 2\sum_{m=1}^{\infty}(-1)^{m+1}\exp\left[-2m^2t^2\right]. \tag{135}$$

The Glivenko–Cantelli Lemma (132) is a generalization of the strong law (68), and (133, 134) is a generalization of the central limit theorem (72).

Brownian motion also provides limit theorems for *random walks*. Let $X_1,\ldots,X_n,\ldots$ be independently and identically distributed random variables with mean μ and variance σ^2 Then

$$S_{tn} = \sum_{i=1}^{[tn]} X_i + (nt - [nt])X_{[nt]+1}, \quad n \geq 1, 0 < t,$$

particularly when X is a lattice random variable, is called a *random walk*. Its large sample properties, n large, can be found by treating

$$\frac{S_{tn} - \mu tn}{\left(n\sigma^2\right)^{1/2}} \tag{136}$$

as a sequence of stochastic processes that behaves asymptotically as a Brownian motion, $X(t)$. Some properties of $X(t)$ that are of interest in this regard are:

$$\text{With probability } 1, X(t) = 0 \text{ infinitely often.} \tag{137}$$

Assume T is a *stopping time*; that is, the event $T \leq t$ depends on the values of $X(s)$ only for $s \leq t$ and assume $ET < \infty$ Then

$$EX(T) = 0. \tag{138}$$

Notice this implies the stopping time in (128) does not have finite expectation.

The nonlimiting case yields Wald's equation,

$$ES_N = \mu EN, \tag{139}$$

where $\{N \leq n\}$ depends on $\{X_i\}$ for $i = 1,\ldots,n$.

Assume $a < 0 < b$ and T is the least t such that $X(T) \leq a$ or $X(T) \geq b$ Then

$$P[X(T) = b] = \frac{|a|}{|a| + b} \quad \text{and} \quad ET = |a|b. \tag{140}$$

Bibliography

Note. Some of the criteria used in selecting the references were clarity, availability and completeness. The Johnson and Kotz volumes are a storehouse of information about specific random variables, and Greenwood and Hartley gives extensive detail on available printed tables. The best entry to the current literature on random variables or other statistical-probabilistic topics is the *Current Index* to *Statistics* (published by the American Statistical Association and the Institute of Mathematical Statistics, 1984, volume 10, also available electronically as *MathScience* produced by the American Mathematics Society) which is a key-word, permuted-title index. Barlow and Proschan, David, Lukacs and Pollard are monographs on specialized topics. Comprehensive views of broad areas are given by Anderson, Chow and Teicher, Rao and Serfling. Ash gives a detailed mathematical setting for probability theory, and Lamperti quickly shows the power of the theory.

Anderson, T.W. 1984. *An introduction to multivariate statistical analysis*, 2nd ed. NewYork: Wiley. 1972.

Ash, R.B. 1972. *Real analysis and probability*. New York: Academic.

Barlow, R.E., and F. Proschan. 1975. *Statistical theory of reliability and life testing probability models*. New York: Holt, Rinehart and Winston.

Chow, Y.S., and H. Teicher. 1978. *Probability theory: Independence interchangeability, martingales*. New York: Springer.

David, H.A. 1981. *Order statistics*, 2nd ed. New York: Wiley.

Greenwood, J.A., and H.O. Hartley. 1962. *Guide to tables in mathematical statistics*. Princeton: Princeton University Press.

Johnson, N.L. 1969. *Distributions in statistics: Discrete distributions*. Boston: Houghton Mifflin, chs 1–11.

Johnson, N.L. 1970a. *Continuous distributions*, vol. 1. Boston: Houghton Mifflin, chs 12–24.

Johnson, N.L. 1970b. *Continuous distributions*, vol. 2. Boston: Houghton Mifflin, chs 22–33.
Johnson, N.L. 1972. *Continuous multivariate distributions*. New York: Wiley, chs 34–42.
Karlin, S., and H.M. Taylor. 1975. *A first course in stochastic processes*, 2nd ed. New York: Academic.
Lamperti, J. 1966. *Probability: A survey of mathematical theory*. New York: W.A. Benjamin.
Lukacs, E. 1970. *Characteristic functions*, 2nd ed. New York: Hafner Publishing.
Pollard, D. 1984. *Convergence of stochastic processes*. New York: Springer.
Rao, C. 1973. *Linear statistical inference and its applications*, 2nd ed. New York: Wiley.
Serfling, R.J. 1980. *Approximation theorems of mathematical statistics*. New York: Wiley.

Randomization

James O. Berger

Randomization refers to the selection of an element a, from a set A, according to some probability distribution P on A.

Example 1

In the 1970 United States draft lottery it was necessary to order eligible males randomly for possible later induction into the armed services. In an attempt to do this fairly, capsules representing each day of the year were mixed in a large drum and selected by drawing. Those individuals with birthdays on the day corresponding to the first capsule drawn would be drafted first; those with birthdays corresponding to the second capsule drawn would be drafted second, and so on. The set A was thus the set of all sequences of capsules corresponding to days 1 through 366 of the year (1970 was a leap year). The sequence of capsules that was actually drawn began $a = (258, 115, 365, 45, 292, 250,\ldots)$.

The goal of the randomization was to be 'fair', so that any such sequence had the same chance of occurring. Choosing a according to the uniform probability distribution on A would have achieved this fairness, the uniform distribution being that which assigns equal probability to each a in A. Interestingly, the mixing process used with the capsules was not very good, and the capsules with large numbers ended up being drawn sooner than capsules with small numbers (on the average). Thus the actual randomization used was not the uniform distribution, and resulted in bias against individuals with late birthdays (see Rosenblatt and Filliben 1971).

Randomization is very commonly used to select winners (or losers) as in Example 1. Lotteries are the most common examples. There are also technical roles for randomization in such fields as statistics and game theory, and it is to these roles we now turn.

The use of randomization in statistics is very widespread, particularly in experimental design.

Example 2

Two medical treatments A and B are to be tested, and 20 patients are available for the experiment. From the 20 patients, 10 are randomly selected using *simple random sampling* (i.e., the selection is done in such a way that any 10 people would have the same chance of being chosen). These 10 are given treatment A, with the remaining 10 being given treatment B.

The major reason for use of randomization in Example 2 is to help prevent possible (unintentional) experimental bias. For instance, the doctors administering the treatments might well have feelings as to which treatment is better for a patient with given characteristics, and could (perhaps subconsciously) allow these feelings to affect the assignment of patients to treatment, if given that responsibility. Historical examples of (unintentional) experimenter-induced bias abound, to the extent that randomization of treatment assignment is now standard practice in most statistical experimentation. The statistician most responsible for the widespread adoption of randomization was R.A. Fisher (see Fisher 1966).

Modes of randomization, considerably more complicated than that in Example 2, are used in sophisticated experimental designs. The major reason for such sophistication is that, while random assignment of treatments can help prevent experimenter-induced bias, it can result in 'unlucky bias'. In Example 2, for instance, the sickest people could (by bad luck) all end up in the group chosen to receive treatment A. To help prevent such an eventuality, and to reduce variance, randomization is often combined with use of *control* in experimental design. (See Cox 1958; Fisher 1966; and Anderson and McLean 1974, for general discussions. Moore 1979, gives an excellent nontechnical introduction to the subject.)

Another use of randomization in statistics, and also in game theory, is to choose an action or statistical answer randomly. The motivation in game theory is easiest to perceive.

Example 3

Each of two players in a game is to choose 'odd' or 'even'. If their choices match, player I wins; otherwise player II wins. This game is to be played repeatedly.

It is clear that if either player falls into a recognizable pattern of choosing 'odd' or 'even', the other player can adapt his strategy to this pattern and win repeatedly. Thus it might be wise for the players to adopt *random* strategies, whereby their choice of 'odd' or 'even' is determined by a chance mechanism. For instance, a simple random strategy is to flip a fair coin, choosing 'odd' if a head occurs and 'even' if a tail occurs (and, of course, keeping the coin flip secret). This would correspond to choosing 'odd' and 'even' with probability 0.5 each.

In a general game having a set A of available strategies, a *randomized strategy* is simply a choice from A according to a probability distribution, P, on A. Each P corresponds to a different randomized strategy. (Some of these strategies can, of course, be quite bad.) Randomized strategies play a crucial role in game theory (cf. Thomas 1984; Berger 1985).

Some proponents of the frequentist approach to statistics advocate use of randomized statistical strategies. The reason is that one could be in a situation where it is impossible, say, to find a statistical test having type I error probability of 0.05, unless one is willing, for certain data, to allow the possibility of deciding at random whether to accept or reject the hypothesis. This can put the statistician in the rather untenable position of having to flip a coin at the end of the analysis, with heads leading to 'rejection at the 0.05 level' and tails leading to acceptance. The careful experimenter, seeing the statistician draw conclusions from his data in this fashion, will not be thrilled. Use of randomized statistical strategies has thus never been very widespread.

Implementing a desired randomization is not as easy as one might expect; witness the fiasco described in Example 1. The most common method used today is based on random number tables or random number generators in computers.

Example 1 (Cont.)

The most direct method of generating a uniform random sequence would be to label the days as 001, 002, 003, . . ., 366, and use a random number table or generator to obtain a sequence of three-digit random numbers. Simply list the three-digit random numbers in the order they occur (ignoring any three-digit numbers, other than those above, which happen to be generated). Note that it is necessary to label day 1 as 001, day 15 as 015, and so on; if the labels of the days were allowed to have different numbers of digits, they would have different probabilities of being generated. (Any one-digit number has three times the chance of being generated, by a uniform random number generator, as does a three-digit number.) See Moore (1979) for further discussion at an introductory level. Note that computers also have available software for generating probability randomizations other than the uniform.

See Also

► Likelihood
► Probability

R

Bibliography

Anderson, V.L., and R.A. McLean. 1974. *Design of experiments*. New York: Marcel Dekker.

Berger, J. 1985. *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.

Cox, D.R. 1958. *Planning of experiments*. New York: Wiley.

Fisher, R.A. 1966. *The design of experiments*, 8th ed. New York: Hafner.

Moore, D.S. 1979. *Statistics, concepts and controversies*. San Francisco: Freeman.

Rosenblatt, J.R., and J.J. Filliben. 1971. Randomization and the draft lottery. *Science* 171: 306–308.

Thomas, L.C. 1984. *Games, theory and applications*. New York: Wiley.

Rank

M. J. Beckmann

All men are equal, but members of an organization are not. Organizations are associations of persons for the achievement of tasks that exceed the capacity of an individual. These tasks must, therefore, be subdivided and the subtasks assigned to certain individual organization members, the operatives.

Supervision

Coordination of subtasks in an organization is (normally) not done through markets but is made a subtask in itself. A simple organization has but one coordinator, the boss, who supervises all operatives directly. Large organizations become possible if *supervision*, too, can be subdivided, supervised and coordinated. This cascading of supervision can, in principle, be carried to any length and no technical limits exist to the size of organizations. (Whether their effectiveness declines with size is another question, see below).

In small organizations the necessary subdivision and coordination can, in principle, be achieved also through mutual consultation and voluntary cooperation: partnerships and communes. Alternatively supervision may be rotated in order to restore equality, but at the expense of stripping the supervision of much of his/her power (as the example of Deans in European universities shows).

Organizations that rely on professional supervisors, or ‘managers’ for the discharging of the coordinator’s job, are called *hierarchical* organizations. In their purest form, they are composed of operatives who do not supervise and supervisors who do no operative work. A ranking system is the outward manifestation of the supervisory structure that is the essential feature of hierarchical organizations, as will be shown now. The coordination of subtasks or of the division of labour in organizations thus relies on supervision as its key element: organizations coordinate through supervision. In practice, supervision can mean many things such as selection, induction, training, setting goals, assisting, monitoring, checking, correcting, evaluating, rewarding and firing.

The remarkable similarities between organizations in the face of the great differences in their tasks results, however, from the fact that formally and structurally supervision has the same properties in all hierarchical organizations.

Formally, supervision is a binary relationship which is acyclic. (This rules out ‘mutual supervision’ or any closed chain of supervision.) A single individual, the president, is distinguished by having no supervisor. All others report to one and only one organization member, their assigned supervisor. This principle of ‘unity of command’ is occasionally replaced by multiple supervision in the form of a functional division of supervision (as between administrative and research director) or of temporary assignment to task forces with special supervisors as in ‘matrix management’.

The supervisory relationships may be pieced together in a ‘directed graph’ known as an *organizational chart*. On this chart every organization member is connected to the president through a unique *line of command*, and to every other organization member by official channels of communication coming together in some manager who directly or indirectly supervises both.

The organizational chart shows that there are at least three types of positions in any complex (i.e. not simple) organization: president (no supervisor), operatives (no subordinates) and intermediate supervisors. Further classifications may exist according too such functions as support and advisory (staff), but these, too, much be fitted into the supervisory structure that holds the organization together.

Control

An organization member j is said to be *controlled* by another organization member i if i supervises j directly or indirectly through a chain of intermediate supervisors. Every supervisor controls a subset of the organization, a department, division, etc.

The president, in control of the entire organization, is usually made responsible for achieving the organization's task to an outside board of trustees, although he/she is exempt from day-today supervision.

The relationship $\rightarrow$, control, is by construction *transitive* (if supervisory chains exist from i to j and from j to k then this defines a supervisory chain from i to k). Control is *irreflexive*: not $i \rightarrow i$ because the relationship of supervision is irreflexive. Control is also *asymmetric*: if $i \rightarrow j$ then not $j \rightarrow i$, because the relationship of supervision is acyclic.

Any relationship with these properties defines a *strict partial ordering* in the set of organization members, the organization. It is partial because in any complex organization there exist two members such that neither controls the other. For purposes of supervision alone, and also for setting salary schedules that do not conflict with supervision and control relationships, this partial ordering would suffice. But interpersonal relationships in large organizations become more transparent when this partial ordering is extended to a complete ordering. In a complete ordering the following is true for any two organization members, i, j:i precedes j, j precedes i, or i and j are equivalent.

Rank

A complete ordering is another word for ranking. Such a ranking is *ordinal* and makes no statement about the 'degree' to which i precedes j.

If the complete ordering is to be made the basis of the salary or compensations structure, then the ranking system must be converted to a *cardinal* one. This is usually achieved by mapping the equivalence classes under the ordinal ranking onto the positive integers. (Example: executive ranks I–V, and GS ranks 1–18 in the Federal government.)

Even when ranks are not labelled numerically, but as titles arranged in well-defined sequences (secretary, under secretary, deputy under secretary, assistant secretary, deputy assistant secretary . . .) this is equivalent to a numerical rank system.

Assignment of Rank

A number system $1, \ldots, r, \ldots, R$ generates unique ranks for all positions on the longest line of command (supervisory chain from the president to operatives). If all lines are equally long, the organization is 'balanced'. In nonbalanced organizations there is some choice in assigning ranks. The choice is constrained by the following *assignment principle*: a supervisor must have a strictly greater rank than any subordinate. Rank jumps may occur. Two simple methods of rank assignment are the following: in *counting up* all operatives are assigned rank 1 and each supervisor ranks one above the highest rank of any subordinate. Rank is then measured by the distance from the president. In *counting down*, each person is assigned a rank just one below his/her supervisor, possibly resulting in a rank for operatives above one. Counting up may be shown to result in the highest and counting down in the lowest ranks consistent with the assignment principle. While organization members may pressure for counting down, organizational efficiency – achieving a given task at minimum cost – is best served by counting up.

In some organizations a person's *permanent rank* ('rank in the job') assigned to the position on the organizational chart, is distinguished from his/her *temporary rank* ('rank in the man') which reflects a person's station in a career (see below) although ideally and normally the two ranks coincide.

Although a rank system is ultimately founded on supervision, ranks above the minimum may be assigned also to operatives (or staff or support persons) even in a system of counting up, to fit them into the salary structure.

Attributes of Rank

The main attributes of rank are: *power, prestige*, and *money income*.

Power is directly exercised over the organization members in the control set of the rank holder, the set of persons reached through a chain of supervision or command. Supervisory power in a hierarchical organization is, however, restricted by the formalities first spelled out by Max Weber in his definition and analysis of bureaucracy (1925).

For rank to be economically meaningful, compensation must increase with rank including: salary, bonus, stock options and the tax-exempt income generated by expense accounts, insurance, use of cars, paid leisure and recreation and other perquisites of rank. Utility is often a linear function of rank. Diminishing marginal utility and the progressive income tax imply that income should then be a convex function of rank. The simplest scheme is one where income grows at a constant rate with rank. In practice, salary schedules tend to combine an exponential with a linear function of rank. The growth factor differs widely between organizations and countries: from 5 per cent in the Chinese civil service to approximately 50 per cent in some American corporations. (Presidential compensation of $2 million compared to an operative's salary of $20,000 in the presence of twelve intervening ranks implies an average salary growth rate of 47 per cent.)

The question of the economic determinants of salary schemes is answered differently in the case of organizations that have no interaction with labour markets except at the entry level of rank one (church, military) and those that operate in competitive labour markets by hiring and firing personnel at all levels (see below).

Prestige, not properly an economic variable, is a residual category that reflects all attributes of rank other than power and money income. The prestige aspect may be expressed by such marks of rank as titles and the trappings of the office (size, location, furnishings, and number of windows).

Span of Control

The workload of a supervisor is best described by the number of subordinates under his/her immediate supervision: the *span of control*. This span may vary between departments and ranks. It is larger in supervising operatives than in supervising other managers. The literature on business organization recommends a lighter span for the president. The smaller the span, the more effective is supervision, *ceteris paribus*. The longer the chains of command, the less effective is top managements' control over operatives. But organizations with a given task size and given labour force (number of operatives) face a tradeoff between average span of control and number of ranks (Simon 1948). The optimal span of control and hence the optimal number of ranks is thus the result of an economic choice rather than of purely technical considerations (such as complexity of the job), and depends also on relative wages. Given the span of control s and the number Q of operatives required to handle the organization's task, the number of managers M to perform the required supervision (including that of their own) is determined as $M = Q - 1/\ S - 1$. Efficient organizational design can then aim at keeping the ranks of management as low as possible. This may be constrained by the presence of rigid departmental boundaries which cannot be crossed by supervisory relationships.

Efficient organizational design is further complicated by the fact that rank express not only a supervisory structure but also serve as an instrument of motivation when rank is considered a stage in a member's career. An analysis of optimal organizational structures, taking account of both the structural and the motivational side of ranks is part of ongoing research.

Careers

Ranks in an organization are not assigned once and for all. Beginners start in rank one and attain higher ranks in this or other organizations only through a process of screening, training, rotation, monitoring, and selection, designed to protect the organization from gross mismanagement and to discover the persons most productive in the higher ranks. Only when no significant differences of qualification persist after initial selection can promotions be reduced to a matter of seniority, which has been called the only truly objective method for advancement (Chapman 1970).

Normal career expectations exist in many organizations. This means that promotion to the top rank in a given rank class is practically a matter of certainty, although the timing is not. Transition from one class to the next (lower to middle management, middle management to top management), is, however, rare and based on rigorous selection.

Expected Utility of a Career

Attainment of high rank is thus always beset by uncertainties, but probabilities of reaching the various levels may be calculated, given the number of ranking positions and the timing of promotions on the average, or conditional on the present age distribution of incumbents. A career is then a lottery in which the end positions may be considered the prizes. The economic value of a prospective career in an organization is then the expected utility of this lottery. Competition for personnel at the entry point to organizations must then equalize the expected utility of careers for qualified personnel in all organizations. This is confirmed by the fact that the calculated utility of careers in various federal agencies, the military, and universities in the USA all lie close together, and apart from the expected utility of government careers for candidates without college degrees (Beckmann 1978, p. 109).

The economic value of a rank, that is, of a continuing career when a certain rank has been reached is then the expected utility of attaining various end positions conditional on the present rank. The incentive of a promotion depends both on the utility difference between the ranks and the probability of this promotion. This incentive is maximal for an intermediate value of the promotion probability.

When ranks are standardized between comparable organizations, for example, universities in the same prestige class, then the expected utility must be equal among organizations at all rank levels for which mobility exists. In the absence of prestige differences salary becomes a proxy variable or sufficient statistic for equivalent rank and this comparability carries over to markets for managers moving between organizations with different rank structures. Prestige differences between organizations imply now that the equivalent rank carries a lower salary in a more prestigious organization (as exemplified by the Ivy League universities).

Salary levels are determined by supply and demand at all levels. But in the long run both are dependent on the frequency distribution of ability. The unravelling of this complicated relationship constitutes another problem area of current research.

Alternatives to Hierarchy

Are there alternatives to hierarchical organization? Some organizations perform tasks that in principle could also be carried out by individuals on their own (artists, artisans, teachers, lawyers, scholars, inventors), but in fact organizations compete with individuals in almost every field or have even replaced them completely. Why and when are organizations superior to individual effort? Clearly when there are *increasing returns* to scale which can be reaped by organizations but not as well by individuals even with access to specialized capital goods through a rental market. Organizations are the best-known method of capturing returns to scale.

A more subtle and difficult question is whether and when subordination can be avoided by sharing the monitoring and coordinating work among the organizations members in a more equitable way. Partnerships exist and communes exemplify alternatives, but only to simple organizations. Their main attractions are freedom from bosses, and a more even distribution of rewards. But mutual supervision and the expected conformity to a high code of ethics may turn out to be even more oppressive. According to Samuel Johnson

> Subordination tends greatly to the happiness of men.… Were we all upon an equality, none of us would be happy, any more than single animals who enjoyed mere animal pleasure. (Boswell, *London Journal*, 20 July 1763.

Modern democratic society, professing the principle of equality, considers rank to be a necessary evil, a means without substitute for running the complex organizations that play such an important part in modern life.

See Also

▶ Hierarchy

Bibliography

Alchian, A.A., and H. Demsetz. 1972. Production, information costs, and economic organization. *American Economic Review* 62: 777–795.

Beckmann, M.J. 1978. Rank in organizations. In *Lecture notes in economics and mathematical systems*, no. 161. Berlin/New York: Springer.

Beckmann, M.J. 1983. *Tinbergen lectures on organization theory*. Berlin/New York: Springer.

Calvo, G., and S. Wellisz. 1978. Supervision, loss of control and the optimum size of the firm. *Journal of Political Economy* 86(5): 943–952.

Chapman, B. 1970. *The profession of government*, 4th ed. London: Unwin University Books.

Crémer, J. 1980. A partial theory of the optimal organization of bureaucracy. *Bell Journal of Economics* 11(2): 683–693.

Hess, J.D. 1983. *The economics of organization*. Amsterdam: North-Holland.

March, J., and H.A. Simon. 1958. *Organizations*. New York: Wiley.

Marschak, J. 1975. Economics of organizational systems. In *Man and computer*, ed. T. Marvis. Amsterdam: North-Holland.

Mirrlees, J. 1976. The optimal structure of incentives and authority within an organization. *Bell Journal of Economics* 7: 105–131.

Radner, R. 1986. Decentralization and incentives. In *Information, incentives, and economic mechanisms: Essays in honor of Leonid Hurwicz*, ed. T. Groves, R. Radner, and S. Reiter. Minneapolis: University of Minnesota Press.

Rosen, S. 1982. Authority, control and the distribution of earnings. *Bell Journal of Economics* 13(2): 311–323.

Simon, H. 1948. *Administrative behavior*. New York: Macmillan.

Simon, H. 1951. A formal theory of the employment relationship. *Econometrica* 19: 293–305.

Spence, A.M. 1975. Incentives, risk, and information: notes towards a theory of hierarchy. *Bell Journal of Economics* 6(2): 552–579.

Tuck, R.H. 1954. *An essay on the economic theory of rank*. Oxford: Basil Blackwell.

Weber, M. 1925. *The Theory of Social and Economic Organization*. Trans. A.M. Henderson and T. Parsons, ed. T. Parsons. New York: Oxford University Press, 1947.

Williamson, O.E. 1975. *Markets and hierarchies: Analysis and antitrust implications*. New York: Free Press.

Williamson, O.E. 1985. *The economic institutions of capitalism*. New York: Free Press.

Rate of Exploitation

Fabio Petri

According to Karl Marx, the proletariat, i.e. wage labourers, is exploited by the capitalists: behind the apparent freedom and equality of the partners in the wage contract, Marx sees a power inequality which results in the workers being exploited by the capitalists in the same sense in which the serfs were exploited by their feudal landlords, or slaves by their masters. The capitalists are able to compel the workers to produce a surplus product, which they appropriate as profits, not by virtue of any productive contribution of theirs, but simply owing to their superior bargaining position vis-à-vis the workers, deriving from their collective monopoly of the means of production. Much the same (although without using the term ‘exploitation’) had already been said by Adam Smith, who also anticipated Marx on the importance of the repressive state apparatus's support for the institution of private property.

This general perspective explains Marx's occasional use of the term 'rate of exploitation' as synonymous with 'rate of surplus value', the latter being the more frequently used term, whose meaning will now be clarified. The labour value of, or labour embodied in, a commodity is defined by Marx as the sum of the direct and indirect labour necessary to its production, i.e. of the live labour expended in its direct process of production plus the labour embodied in the means of production used up (according to the socially necessary conditions of production) in that same process. If the socially necessary live labour performed in the whole economy is L, and the labour embodied in the means of production used up to produce the total social product is C, then the labour value of the total social product is $L + C$, and of the net social product is again L (because the net social product is defined as the total social product minus that part of it which replaces the means of production used up, a part whose labour value is clearly C). If now V is the labour embodied in the part of the net social product going to the workers, then $S \equiv L - V$, the surplus labour, or surplus value, is the labour embodied in the surplus product. Under constant returns to scale, only V, instead of L, would be necessary to produce a net product equal to the workers' share only; hence Marx calls V the 'necessary' or 'paid' labour, and S the surplus or 'unpaid' labour, and divides in the same proportions the average working day in a 'paid' and an 'unpaid' part. The ratio S/V is what he calls 'rate of surplus value' or 'rate of exploitation'.

Given the techniques in use, S/V depends on the average wage basket, and its changes reflect changes in the balance of power between classes. Its importance for Marx lies in its being one of the two proximate determinants of the rate of profits, the other one being the average 'organic composition of capital', i.e. the ratio of what Marx called 'constant capital' (the labour value of the capital goods employed in the production process) to what Marx called 'variable capital' (the labour value of the wage goods, which for Marx are part of capital because he considered wages to be advanced, rather than paid at the end of the production period as is usually assumed nowadays), in other words the ratio (assuming for simplicity that all the capital goods utilized in the economy are circulating capital) C/V. The rate of exploitation and the organic composition of capital can also be defined for each industry: then $s + v$ is the live labour performed in that industry; s/v, the rate of exploitation, is the ratio of the surplus or 'unpaid' labour to the labour value of the real wages obtained by the workers in that industry, c the value of the capital goods employed; c/v the organic composition of capital; and the rate of profits is given by $r = s/(c + v)$ which can also be re-written as $r = (s/v)[(c/v) + 1]$. If – as Marx assumes in Volume I of *Capital* – commodities exchanged at prices proportional to labour values, then the rate of profits (assuming prices proportional to labour values) could be uniform across the different industries only if – what observation shows not to be true – c/v were uniform (s/v is, on the other hand, uniform if the hourly wage is uniform or, as Marx assumes, heterogeneous or differently paid labour is reduced to homogeneity on the basis of relative wages). Marx was thus able to understand, more clearly than anyone before him, why the tendency of profit rates towards uniformity will cause relative prices to deviate from relative labour values. He nonetheless thought that in the economy as a whole the deviations cancel out, and that the uniform rate of profits is therefore the same as the average rate of profits which would obtain if commodities did exchange at labour values, i.e. $r = S/(C + V)$, or $r = (S/V)/[(C/V) + 1]$.

Thus, he thought, the influences on the rate of profits can be better understood by studying the way they affect the two ratios S/V and C/V. This he thought to be a useful distinction because it allowed one better to separate the effects on the rate of profits of various types of technological change (effects which could be seen to be important in so far as they affected C/V or – e.g. speedups – S/V) from the effects of the workers' struggles over the wage level or, given the *daily* wage, over the length of the working day (affecting S/V).

This role of the rate of exploitation as defined by Marx has been undermined by the subsequent analytical advances in the theory of prices of production, associated with the names of Dmitriev, Bortkiewicz, Sraffa and now many others. It has been seen that Marx's basic insight was correct in that the *data* (the technological conditions, i.e. the matrix of physical and labour inputs, and the average wage basket), from which individual labour values and the aggregate magnitudes *S, V, C* are derived, do suffice to determine the rate of profits and relative prices; but it has also been seen that Marx's formula $r = S/(C+V)$ is incorrect except in very special cases, and that, although it would be possible to find algorithms to determine the correct rate of profits and prices from individual labour values, the calculation of labour values is anyway superfluous, a direct determination of the rate of profits and prices from those data being possible and easier. New analytical instruments, e.g. the wage-profit frontier, allow a more rigorous study of the effects of changes in technology or in the real wage on the rate of profits than *S/V* and *C/V* (e.g. it has been seen that technical change may in some cases cause r to move in a direction opposite to what Marx's formula would lead one to expect), relegating – for the study of these problems – labour value magnitudes to historical importance only, in that they allowed Marx to determine prices and the rate of profits, and the effects of the main forces acting on them, in the only (imperfect) way concretely possible at the time (Garegnani 1984).

Many marxists (e.g. Sweezy, Hunt, Nuti) defend the importance of labour values by arguing that these allow one to show that workers are exploited. It is often claimed, in this connection, that central to Marx's analysis was the so-called Fundamental Marxian Theorem, stating that the rate of profits is positive if and only if the rate of exploitation is positive (Morishima 1974). This is a doubtful claim, since the theorem re-states, in terms of labour embodied, the obvious fact – accepted by all critics of Marx as well – that profits can only be positive if wages do not absorb the entire net product. To call the *S/V* ratio 'rate of exploitation' is not a *demonstration* that workers are exploited: e.g. the marginalist, or neoclassical, approach would have no quarrel with the Fundamental Marxian Theorem and yet would argue that workers are not exploited, because they receive their marginal products, i.e. as much as each of them is contributing to production, and in the same way a positive rate of profits does not emerge from domination but rather corresponds to the marginal product of capital, and is therefore a just reward to the sacrifice of postponed consumption which, through savings, creates the capital: the marginalist explanation of distribution thus implies that capitalists (i.e., in the marginalist approach, savers) do contribute to production. The required demonstration of the existence of exploitation appears rather to lie in the validity of Marx's different explanation of why the surplus product does not go to the workers, referred to above, now supported by the criticisms directed at the marginalist theory of distribution (Eatwell and Milgate 1983).

The existence of exploitation is therefore not endangered by the demonstration, due to Steedman (1975), that the Fundamental Marxian Theorem cannot be generalized to the case of joint production, so long at least as labour values are defined as usual, i.e. as the prices (in terms of the wage) at a zero rate of profits (if A and B are the square matrices of input and output coefficients respectively, and l the labour input vector, then the vector of labour values k is determined by $kA + l = kB$; this expression is what the price equations $(pA + wl)(1 + r) = pB$ collapse to if $r = 0$ and $w = 1$; without joint production one has $B = I$, the identity matrix, and hence $kA + l = k$). With joint production, some labour values may be negative, and the surplus product may then have a negative labour value, implying a negative rate of exploitation. An intuitive explanation is as follows. The labour value of a commodity is an employment multiplier, indicating by what amount total employment would change if (with constant returns) the net product of that commodity increased by 1 unit, the other net products remaining constant. If several commodities are jointly produced by several processes, an increase in the net product of only one commodity may require expanding some processes but contracting some others (no contraction could be necessary in the absence of joint products): the resulting total variation in employment need not be positive. If the rate of exploitation is negative, total employment would have to increase in order not to produce at all the surplus product. But, it would seem, there still is exploitation, because the surplus product is not going to the workers, while it would if the capitalists' domination were not preventing the wage from rising.

Morishima and others have counter-argued that the idea of a negative labour embodied in a (single or composite) commodity makes no sense, and have proposed to re-define (via linear programming) the labour embodied in a commodity as the minimum labour time necessary, with the known techniques, to produce a net product containing at least that commodity (but possibly other commodities as well; individual labour values are then no longer additive, the labour value of a bundle of wool and mutton is no longer the sum of the labour values of the wool and of the mutton). The surplus labour S^* is then the difference between L and the minimum labour V^* necessary to produce, with the available technical knowledge, a net product containing at least the total wage basket. The rate of exploitation is then re-defined as S^*/V^*: a notion, it would seem, only interesting for purposes of comparison of reality with possible utopias ('how much less workers could afford to work if the social goal were the minimization of their working time, given their consumption'). It is not impossible, anyway, that in extreme cases S^*/V^*: be zero in spite of a positive surplus product, as shown by the following example: the economy produces only, and jointly, wool and mutton from sheep, the surplus product consists of all the wool and the real wages of all the mutton; the rate of profits might be positive too (Petri 1980).

This and other recent attempts at re-defining labour values and the rate of exploitation cannot, it would seem, find support in Marx, where the role of labour values appears to have been only the determination of prices and of the rate of profits, as shown for example by the way labour values are determined: Marx, like Ricardo, determines labour values on the no-rent land, and reduces heterogeneous labour to homogeneity on the basis of relative wages (implying a rate of exploitation uniform by assumption for all kinds of labour; see Steedman 1985): which is what he must do in order to argue that prices would be proportional to labour values were it not for the non-uniformity of the organic composition of capital. Nowadays, ethical aims, for example some measurement of the degree of suffering imposed upon workers by capitalism, are often implicit in the search for re-definitions of the rate of exploitation. This is not necessarily illegitimate, but should be clearly stated and distinguished from Marx's own project.

See Also

► Surplus Value

Bibliography

Dmitriev, V.K. 1974. *Economic essays on value, competition and utility*, ed. D.M. Nuti. Cambridge: Cambridge University Press.

Eatwell, J., and M. Milgate (eds.). 1983. *Keynes's economics and the theory of value and distribution*. London: Duckworth.

Garegnani, P. 1984. Value and distribution in the classical economists and Marx. *Oxford Economic Papers* 36(2): 291–325.

Hunt, E.K., and J.G. Schwartz (eds.). 1972. *A critique of economic theory*. Harmondsworth: Penguin.

Mainwaring, L. 1984. *Value and distribution in capitalist economies. An introduction to Sraffian economics*. Cambridge: Cambridge University Press.

Marx, K. 1867–94. *Capital*, vols. I–III. Moscow: Progress Publishers, 1965–6.

Morishima, M. 1974. Marx in the light of modern economic theory. *Econometrica* 42(4): 611–632.

Morishima, M., and G. Catephores. 1978. *Value, exploitation and growth*. London: McGraw-Hill.

Petri, F. 1980. Positive profits without exploitation: A note on the generalized fundamental marxian theorem. *Econometrica* 48(2): 531–533.

Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.

Steedman, I. 1975. Positive profits with negative surplus value. *Economic Journal* 85: 114–123.

Steedman, I. 1985. Heterogeneous labour, money wages, and Marx's theory. *History of Political Economy* 17(4): 551–574.

Steedman, I., P. Sweezy, et al. 1981. *The value controversy*. London: New Left Books.

von Bortkiewicz, L. 1907. Value and price in the Marxian system. Trans. *International Economic Papers,* No. 2, 5–60, 1952.

R

Rational Behaviour

Amartya Sen

Abstract

A clear distinction must be drawn between (*a*) the type of behaviour that might be described as *rational*, and (*b*) rational behaviour models that might be useful in making predictions about *actual* behaviour. Neither of the two standard views of rational behaviour – as 'consistent choice' or as 'self-interest maximization' – has emerged as an adequate representation of rationality or of actuality. The difficulties that these views encounter carry over to rational behaviour models accommodating uncertainty.

Keywords

Adaptive expectations; Binariness; Bounded rationality; Cardinal utility; Choice; Completeness; Enforced maximization; Exchange; Expectation formation; Expected utility; Externalities; Instrumental rationality; Natural selection; Pareto efficiency; Preferences; Prisoner's Dilemma; Probability; Rational behaviour; Rational expectations; Revealed preference theory; Risk; Satisficing; Self-interest; Sen, A. K.; Social norms; Strong independence; Substantive rationality; Sure thing principle; Transitivity; Uncertainty; Utilitarianism

JEL Classifications

D8

The concept of rational behaviour is frequently used in economic theory. The interest in this concept springs from two quite distinct motivations. First, in so far as economic exercises often take a prescriptive form, it is interesting to know how one could behave rationally in a given situation. This may be called the 'prescriptive motivation'. It should be warned that the prescription need not be necessarily of an ethical kind. Indeed, the prescriptive motivation is sometimes described in clearly non-ethical terms, involving the pursuit of self-interest only. In a classic presentation of this position, Harsanyi (1977) describes 'perfectly rational behaviour' in the context of game theory in the following terms:

> . . . our theory is a *normative* (prescriptive) theory rather than a *positive* (descriptive) theory. At least formally and explicitly it deals with the question of how each player *should* act in order to promote his own interests most effectively in the game and not with the question of how he (or persons like him) *will* actually act in a game of this particular type. (Harsanyi 1977, p. 16)

The second motivation concerns the possible use of models of rational behaviour in explaining and predicting *actual* behaviour. This exercise is done, as it were, in two steps. The first step consists in characterizing rational behaviour and the second, following that, bases actual behaviour on rational behaviour. In this way the characterization of rational behaviour may end up specifying the predicted actual behaviour as well. This motivation underlies much of the theory of general equilibrium (see, for example, Edgeworth 1881; Arrow 1951; Debreu 1959; Arrow and Hahn 1971). The argument is that while actual behaviour can, in principle, take any form, it is reasonable to assume that much of the time it will, in fact, be of the kind that can be described as 'rational'.

In reviewing the theory of rational behaviour, this duality of motivations has to be borne in mind. Even though the primary concern of this essay is with the way rational behaviour has been characterized, the nature of the second motivation makes it imperative that the possible use of rational behaviour models for explaining and predicting actual behaviour must not be overlooked.

Rationalizability, Binariness and Self-Interest

In the presence of uncertainty, rational behaviour requires an appreciation of possible variations in the outcome of any chosen action, and such behaviour must, therefore, be based on systematic reading of uncertainties regarding the outcome and ways of dealing with them. Rational behaviour under uncertainty will be presently taken up, but before that the more elementary case when there is no uncertainty has to be dealt with. In fact, behaviour under certainty can be formally seen as an extreme case of behaviour under uncertainty when the uncertainty in question is not only small but simply absent. In this sense, rational behaviour under certainty must be subsumed by any theory that deals with rational behaviour in the presence of uncertainty.

Although there are many different approaches to rational behaviour under certainty, it is fair to say that there are two *main* approaches to this question. The first emphasizes *internal consistency*: rationality of behaviour is identified with a requirement that choices from different subsets should correspond to each other in a cogent and systematic way. Various conditions of internal consistency have been proposed in the literature, but the one which seems to command most attention in formal economic theory is *binariness,* which requires that the choices from different subsets can be seen as maximizing solutions from the respective subsets according to some binary relation R (often interpreted as 'preference', for example, xRy standing for 'x being preferred or indifferent to y'). Or, to put it another way, rational behaviour, in this interpretation, amounts to our ability to find a binary relation R over the universal set of alternatives such that the choice from any particular subset of that universal set consists of exactly the R-maximal elements of that subset. Richter (1971) calls this 'rationalizability'.

In other formulations – still within the general approach of internal consistency – the condition of rationalizability has been relaxed, demanding only a part of the kind of consistency that binary maximization must entail. On the other hand, in some other formulations, the demands have been made stronger than that of maximization according to a binary relation by requiring further that the binary relation in question be an ordering, satisfying both completeness and transitivity.

An enormous variety of conditions of internal consistency have been proposed in the literature, but it can be shown that many of them are equivalent to each other, and indeed altogether they fall into a number of classes, with each class containing different, but essentially equivalent, demands. Such reductionist analyses can be found, for example, in Houthakker (1956), Uzawa (1956), Arrow (1959), Richter (1971), Sen (1971), Herzberger (1973), Suzumura (1983). For critiques (and arguments for the rejection of) the binary approach to rationality, see Kanger (1976), Gauthier (1985), Sen (1985a, 1986b) and Sugden (1985).

The second common approach to rational behaviour under certainty sees it in terms of reasoned pursuit of self-interest. The origins of this approach are often traced to Adam Smith, and it is frequently asserted that the father of modern economics saw human beings as tirelessly fostering their respective self-interests. As a piece of history of economic thought, this is, to say the least, dubious, since Adam Smith's (1776, 1790) belief in the hold of self-interest in some spheres of activity (for example, exchange) was qualified by his conviction that many other motivations are important in human behaviour in general (on this see Winch 1978; Brennan and Lomasky 1985; and Sen 1987). But it is certainly true that the assumption of the 'economic man' relentlessly pursuing self-interest in a fairly narrowly defined form has played a major part in the characterization of individual behaviour in economics for a very long time.

Self-Interest and Consistency

Rational behaviour in the form of maximization in pursuit of self-interest makes the analysis of individual behaviour a good deal more tractable than a less structured assumption would permit. This is certainly one of its appeals. In addition this behavioural assumption is also quite crucial for the derivation of certain central results in traditional and modern economic theory, for example, Pareto optimality of competitive equilibria and vice versa (Arrow 1951; Debreu 1959; Arrow and Hahn 1971). This is sometimes called the 'Fundamental Theorem of Welfare Economics'. Roughly stated, it claims, first, that every perfectly competitive equilibrium (with each person maximizing utility, given the prices) under certain assumptions (such as no externalities) achieves Pareto optimality, and second, under a slightly different set of assumptions (including the requirement of no externalities, but also some additional requirements, such as the absence of increasing returns to scale), every Pareto optimal state is a perfectly competitive equilibrium with respect to some set of prices and some initial distribution of resources. This correspondence between Pareto optimality and competitive equilibria works neatly given individual self-interested behaviour precisely because Pareto optimality is one characteristic of self-interest maximization of a group, in the sense that in such a situation no one's self-interest can be further enhanced without hurting the self-interest of somebody else. It is the assumption of rational behaviour in the form of the pursuit of self-interest that established the close relationship between competitive equilibria and Pareto optimality (with price-taking behaviour and absence of externalities preventing people from getting in each other's way in their respective pursuit of self-interest). In this result and in many other similar ones, the particular characterization of rational behaviour chosen plays a strategically crucial role.

It can be argued that rational behaviour under the self-interest approach is a special case of that under the consistency approach. If a person does pursue self-interest, it may follow that his or her behaviour will have the consistency needed for maximization of a cogent function. On the other hand, a person can be consistent without necessarily maximizing self-interest, since the maximizing function may have a different interpretation altogether (for example, the pursuit of some moral values or political goals). Thus internal consistency of choice may be taken to be necessary but not sufficient for self-interested behaviour. There is undoubtedly something in this way of seeing the correspondence between the two common approaches to rational behaviour.

However, that alleged correspondence is also somewhat misleading, since the nature of self-interest need not necessarily take the uncomplicated form of being binary in character. Strictly speaking, neither does the self-interest thesis entail the consistency thesis, nor of course the other way round. While this must, in general, be correct, nevertheless the way self-interest has been actually viewed in standard economic theory has made it clearly binary and more typically an ordering (and often seen as being numerically representable). If self-interest must take this form, then it would indeed be the case that the self-interest approach is just a special case of the consistency approach.

In some treatises on rational behaviour, the distance between the self-interest approach and the consistency approach is bridged by some careful definitions. For example, in the 'revealed preference theory', pioneered by Samuelson (1938), consistency is demanded in the form of the 'Weak Axiom of Revealed Preference', to wit: if x is chosen from a set containing y, then y will not be chosen from any set containing x. This type of consistency is, on its own, without a particular substantive interpretation, except that it corresponds generally to some kind of maximization. However, the term 'revealed preference' might indicate that the chosen alternative is always also the preferred one. In so far as preference reflects self-interest (as is typically assumed to be the case), this established, through the terminology of 'revealed preference', what looks like a congruence of choice and self-interest.

The consistency entailed by the Weak Axiom of Revealed Preference does not, in general, entail transitivity, which is a property that might be thought to be a natural one to impose on the relation of self-interest. But that hole can be plugged *either* by demanding stronger conditions (such as Houthakker's 1950, 'Strong Axiom of Revealed Preference') *or* by demanding that the consistency of the Weak Axiom be satisfied over all finite subsets, which makes the strong axiom equivalent to the weak (on this see Arrow 1959; Sen 1971). One way or another, the consistency imposed by revealed preference axioms can lead to a 'preference' relation that has the regularity properties normally associated with the concept of self-interest, and then the gap between the two could be seen as fully bridged.

However, that entire bridging exercise is based on *defining* the relation of choice as a relation of 'preference' which happens to be 'revealed' by the act of choice. But that terminology is arbitrarily imposed, and it is possible that the binary relation of choice, even when fully transitive and complete, may in fact reflect neither the person's preference, nor his or her self-interest. There is, obviously, scope for methodological arguments on this point, and these issues have often been joined.

In the philosophical literature, it is common to distinguish between 'instrumental rationality' and 'substantive rationality' (see Latsis 1976). It is clear that the self-interest view of rational behaviour is one of substantive rationality requiring that rational behaviour must take the form of pursuing some independently defined self-interest. Obviously, this characteristic of substantiveness is not satisfied by the theory of revealed preference, since there the identification of choice with preference or self-interest takes the form of *defining* the relation of choice as a relation of preference, which is not an independent way of characterizing preference or self-interest. But in other theories, the substantive exercise is carefully done, for example, in the typical general equilibrium theory (see Arrow 1951; Debreu 1959; Arrow and Hahn 1971). The starting point of individual behaviour is, then, not a choice function but a utility function, representing the self-interest of the person in question. Choices follow from constrained maximization of that utility function. In this form, the substantive nature of the characterized rationality is strongly asserted, in the shape of pursuit of self-interest.

A number of criticisms have been recently made about the special nature of the assumption of self-interest maximization. Human beings may well have other motivations, and self-interest is just one of various things that a person might wish to pursue. Different types of criticisms of this substantive assumption have been made by such authors as Nagel (1970), Kornai (1971), Sen (1973, 1977, 1987), Scitovsky (1976), Leibenstein (1976), Schelling (1978), Wong (1978). Elster (1979, 1983), Hirschman (1982, 1983), McPherson (1982), Margolis (1982), Akerlof (1984), Schick (1984) and others.

If the assumption of self-interest maximization is seen as too narrow, it can be argued that merely requiring internal consistency is much too permissive. Indeed, it is tempting to think of the consistency approach as belonging to the 'instrumental' view of rationality. But this is not quite so, since the instrumental view requires that the person pursues some independently defined objective (even though the objective need not be based on self-interest only). In the consistency view there is no such independently defined function at all, and the binary relation that is precipitated by the choice function is a *reflection* of choice rather than a *determinant* of it. It is rather that the consistency approach opens the way to some instrumental view of rationality, involving the maximization of some objective function. Indeed, in this sense, the consistency approach can be seen as permissively admitting the approach of instrumental rationality implicit in the self-interest approach, where the objective function maximized happens to be the self-interest of the person in question.

The consistency approach can be criticized on grounds of inadequacy in characterizing rationality of behaviour. A person's choice function may be internally consistent in the sense that the different things chosen from different subsets correspond to each other in an apparently cogent and coherent way, but this does not in itself indicate that the person's behaviour is consistent with his or her aims or objectives. Indeed, a person who systematically does exactly the *opposite* of what has to be done for the pursuit of his or her objective function may end up producing a consistent choice behaviour, but the binary relation that will be revealed by the choices – the 'opposite' of the person's objective function – will be, clearly, at war with the goals and aims of that person. To describe such a person as behaving rationally would, obviously, lead to some interesting methodological difficulties.

Maximizing, Satisficing and Bounded Rationality

These problems with the standard views of rationality tend to undermine the very foundations of these approaches. Some other approaches have involved more qualified use of the standard presumptions. For example, Herbert Simon (1957, 1979) has argued powerfully that individuals may not actually *maximize* any function at all, and their behaviour may take the form of what has been called 'satisficing'. There are various ways of characterizing satisficing, but it can be thought of in terms of a person having a certain target level of achievement, which he or she will try to reach, but beyond which he or she may not try to improve the achievement any further.

There is a genuine problem of interpretation involved in analysing satisficing, and it can be argued that satisficing behaviour really is maximization according to an effectively incomplete relation, such that the states satisfying the target level of achievement are all put in a non-comparable class as far as choice behaviour is concerned. Maximization can indeed be defined in terms of such incomplete relations (see, for example, Debreu's 1959 analysis of 'maximal' sets based on 'pre-orderings'), and, if it is seen in these terms, the gap between satisficing and maximizing may be, at least formally, reduced. However, the content of the claim of satisficing is that the person in question *can* tell between the different levels of achievement which are all beyond the target level required, and, despite this discernibility, choice behaviour departs from relentless maximization of the level of achievement. In this version of the story, a substantial difference is indeed made by the notion of satisficing, and the implications of satisficing behaviour may, in this interpretation, be quite different from those of maximization.

Variations of the maximization assumption and the related consistency conditions can be justified by seeing the use of reason in human affairs in terms of what has been called 'bounded rationality'. In this structure human choice is seen not in terms of grand maximizing behaviour, but as a series of particular decisions, not fully integrated with each other, taken in situations of partial information and based on limited reflection. This approach has been developed by Herbert Simon (1957, 1979, 1983) both at a theoretical level and in the context of specific empirical applications. The results differ quite substantially from that of rational behaviour seen in terms of consistency, or in terms of optimization according to self-interest. As Simon puts it:

> Rationality of the sort described by the behavioural model [of bounded rationality] doesn't optimize, of course. Nor does it even guarantee that our decisions will be consistent. As a matter of fact, it is very easy to show that choices made by an organism having these characteristics will often depend on the order in which alternatives are presented. (Simon 1983, p. 23)

Natural Selection and Motives

Supporters of optimizing models have typically used two different types of arguments to defend the practice, against models of the kind characterized by 'bounded rationality' and other behavioural departures. One argument takes the direct form of arguing that human beings do optimize and take care to do so. The second argument suggests that natural selection will lead in this result: those who optimize do better, and those who do not get eliminated by natural selection. For example, non-profit-maximizing firms may go to the wall, so that only the profit-maximizing ones may survive (see Friedman 1953). This type of indirect justification of what has been called 'enforced maximization' has many pitfalls, since the analogy with natural selection in biology is at best tenuous (see Helm 1984; Matthews 1984), and the biological story itself is far from straightforward (Dawkins 1982; Maynard Smith 1982).

It is by no means clear that individual self-interest-maximizers will typically do relatively better in a group of people with diverse motivations. More importantly, when it comes to comparisons of survival of different *groups*, it can easily be the case that groups that emphasize values other than pure self-interest maximization might actually do better (see Sen 1973; Sen 1974; Sen 1985b; Akerlof 1984). It has been argued that economic success has often come more plentifully in cultures that emphasize norms of conduct quite different from that of persistent maximization of individual self-interest, focusing on other values (for example, what Morishima 1982, calls 'the Japanese ethos'; see also Dore 1983). The relation between social norms and individual conduct is an enormously complex field, and the simple assumptions of self-interest maximization, of straightforward models of apparent 'consistency', may overlook important aspects of the individual–society relationships (see, for example, Hirschman 1970, 1982). This is not to argue that 'natural selection' arguments are worthless in economics – they may be far from that – but the results of the selection may lack the simplicity demanded by supporters of simple optimization and may take a more complex form (see Hirshleifer 1977; Helm 1984; Matthews 1984).

In assessing the overall value of standard models of rational behaviour, it is important to pay attention to the distinction made earlier between the value of these structures as representations of *rationality* and their usefulness in terms of predicting *actual* behaviour. Some of the deficiencies of the optimizing 'structure apply specifically to the latter. For example, models of 'bounded rationality' are often defended by claims of greater plausibility in explaining actual human conduct.

In fact, the entire enterprise of getting to actual behaviour via models of rationality may itself be seen us methodologically quite dubious. There is scope for argument here on both sides, since the unrealism of rational behaviour may be large, but the unrealism of any *specific kind* of 'irrational' behaviour could be larger still. Whether 'bounded rationality' is the right kind of compromise in getting a grip on actuality via limited use of rationality remains an interesting question.

Reason and Rationality

As far as the other objective of rational behaviour models is concerned, that is, the ability of these models to capture the essence of rationality (no matter how people do actually behave), there are a number of complex philosophical issues underlying the question. It is easy enough to argue that mere internal consistency of choice cannot be adequate for rationality, nor can self-interest maximization be seen as uniquely rational in a way that pursuing other kinds of objectives (such as altruism, public spirit, class consciousness, group solidarity) must fail to be. What is much harder to do is to develop an alternative structure for rationality that would be regarded as satisfactory for the purpose of capturing what can be demanded of reason in human choice (whether or not it also serves the second purpose of giving us a good guess regarding actual behaviour). This question remains, to a great extent, an open one, which has been as yet rather inadequately explored.

Two difficulties, in particular, may be worth mentioning in this context. First, while 'instrumental rationality' must have some place in economics, and the role of reasoned choice of means for serving *given* ends cannot be dismissed, it is hard to believe that any kind of objectives no matter how bizarre – must be seen as okay, that is, not compromising the rationality of the person pursuing it. The need for rational assessment of objectives and preferences have been analysed by John Broome (1978), Derek Parfit (1984) and others, and both the procedural and substantive features of this type of assessment do deserve serious attention.

Second, even when goals are clearly given, the translation of these into actions depends on the pattern of social interdependence assumed in group behaviour, with members having partly divergent goals. As the discussions on the so-called 'Newcomb's problem' and other complex cases have brought out, the correct individual decision may not be entirely unproblematic even when there appears to exist a strictly dominant strategy (see Nozick 1969; Brams 1975; Levi 1975: Gibbard and Harper 1978; Jeffrey 1983, among others). The nature of beliefs permits alternative interpretations of the nature of the decision problem, and this philosophical question is of relevance to decision problems in economics as much as it is in other fields of human choice.

The Prisoner's Dilemma has been frequently used in economic arguments to illustrate the nature of inefficiencies of atomistic non-cooperative behaviour when the interdependence incorporates both congruence and conflict of interests in such a way that the combination of each person's dominant strategies produces an outcome that is inferior in terms of the goals of everyone in the group (see Luce and Raiffa 1957). Attempts to resolve the problem by assuming temporal repetition of the game have not been easy, since it can be demonstrated that with complete knowledge and standard optimizing behaviour, a finitely repeated Prisoner's Dilemma will continue to produce the inferior outcome throughout (Luce and Raiffa 1957, pp. 97–101).

Such non-cooperative behaviour is, however, violated in many experimental games as well as in the usual readings of many real-life situations. The apparent dissonace between received theory and observed behaviour has been explained in a variety of ways in the large literature that has developed on the Prisoner's Dilemma. The 'ways out' have included relaxing the assumption of mutual knowledge, for example, introducing uncertainty about the number of times for which the game will be played, admitting ignorance of the players about other people's knowledge and motivation, limiting the range of alternative strategies that can be considered, and other relaxations (see Howard 1971; Basu 1977; Davis 1977; Radner 1980; Smale 1980; Kreps et al. 1982; Axelrod 1984). Other analyses have emphasized more complex features of 'practical reasoning' involving various types of action ethics, sensitive beliefs, behavioural commitments, and instrumental use of reciprocity; see Sen (1974, 1985b), Watkins (1974, 1985). Levi (1975), Gauthier (1985), McClennen (1985). If it has done nothing else, the literature has at least brought out sharply the complexity of the nature of rationality in situations of interdependence as well as various conceptual and logistic difficulties in using models of rationality to understand the nature of actual behaviour.

It stems easy to accept that rationality involves many features that cannot be summarized in terms of some straightforward formula, such as binary consistency. But this recognition does not immediately lead to alternative characterizations that might be regarded as satisfactory, even though the inadequacies of the traditional assumptions of rational behaviour standardly used in economic theory have become hard to deny. It will not be an easy task to find replacements for the standard assumptions of rational behaviour – and related to it of actual behaviour – that can be found in the traditional economic literature, both because the identified deficiencies have been seen as calling for rather divergent remedies, and also because there is little hope of finding an alternative assumption structure that will be as simple and usable as the traditional assumptions of self-interest maximization, or of consistency of choice.

Uncertainty and Expected Utility

The extension of the modelling of rational behaviour from certainly to uncertainty involves both (*a*) the characterization of uncertainty, and (*b*) taking note of uncertainty thus characterized in making actual decisions over alternative courses of actions. The model that has been most extensively used in this context is that of 'expected utility'. This takes the form of weighing the value of each of the outcomes by the respective probabilities of the different outcomes. The probability-weighted overall 'expected value', thus derived, is then maximized in this approach to rational choice under uncertainty.

The use of probability calculus involves interpretational problems as to what the probabilities stand for. While the view of probability as a measure of relative frequency is a natural one to consider, there is clearly much cogency in interpreting probability as a measure of the degree of belief (as argued by Fisher 1921, and Keynes 1921).

Actual decision-taking operations involve a reading of the likelihood of different outcomes and an assessment of the different outcomes in the light of the respective likelihoods. In a pioneering contribution in axiomatizing conjointly characterized probabilities and utilities, Frank Ramsey (1931) provided the structure (and a possible derivation) of the expected utility calculus. Another major contribution in this area came from von Neumann and Morgenstern (1947). Given the probabilities of different outcomes, consistent and complete rankings of the possible lotteries over the outcomes (including lotteries of lotteries and so forth) permit the construction of cardinal utility functions for the respective rankings associated with the outcomes, provided the rankings in question satisfy certain regularity properties which were specified by von Neumann and Morgenstern (see also Marschak 1946). The assigned cardinal utility numbers of the respective outcomes, weighted by the respective probabilities, when summed together, yield the expected values of the lotteries, and provide numerical representations of the overall goodness of the respective lotteries. Rational behaviour under expected utility maximization takes the form of choosing that lottery which has the highest overall value, thus calculated. The expected utility approach can be and has been used extensively both in economic theory and in applied economics (see, for example, Friedman and Savage 1948; Arrow 1971).

Independence and Consistency

The axioms underlying the derivation of expected utility maximization have been subjected to a good deal of examination and scrutiny. There is scope for disputation about both the exact content and the plausibility of the expected utility axioms (for a very helpful introduction see Luce and Raiffa 1957; see also Fishburn 1970, 1981).

The axiom that has perhaps attracted the most criticism is the so-called 'strong independence'. This independence condition can be stated in several different ways, but a rather immediate one is the following. If in a combined lottery over, say, lotteries L^1 and L^2, the latter L^2 is replaced by another lottery L^3 which is preferred to L^2 (leaving the probabilities and L^1 unchanged), then the modified combined lottery (over L^1 and L^3) would be preferred to the original one (over L^1 and L^2). And vice versa.

Another axiom, related to this one, is sometimes called 'the sure thing principle', which, in one version, requires that anything that raises the probability of the preferred component in a two-alternative lottery would improve the lottery. These axioms are implicit in expected utility maximization, even though the 'independence' condition can be dispensed with in a more limited ('locally' valid) version of expected utility behaviour (as has been shown by Mark Machina 1982).

Various 'counter-examples' to expected utility maximization have been proposed in the literature, often on the basis of considering interesting 'hypothetical' cases, but sometimes on the basis of experimental observations as well. In assessing these objections, we must distinguish, once again, between the claims *to* rationality of this model, and the claims of the model to explain actual behaviour *via* rationality.

It is certainly clear that very often people do act in a way that cannot be made consistent with expected utility maximization. (An early critique, with an alternative framework for choice behaviour, came from Shackle 1938, 1952.) Observations of behaviour and articulated judgements under uncertainty have indicated different types of violations of expected utility behaviour (see, for example, Kahneman et al. 1982). There seem to be problems both in risk perception as well as in the utilization of probability information in making actual decisions. These departures from rational behaviour in the form of expected utility maximization have considerable implications on the way economic models may have to be constructed involving uncertainty (on this see Arrow 1982, 1983). As a framework for understanding actual behaviour, the merits and demerits of the expected utility model are certainly becoming clearer on the basis of recent work. But the 'bottom line' of overall judgement continues to vary. While some have been extremely sceptical, others (such as Harsanyi) continue to emphasize, with some justice, the usefulness of this model in 'explaining or predicting real-life human behaviour' (1977, p. 16).

The need for departures – small or great – from the expected utility model in explaining *actual* behaviour does not, of course, settle the question of the rationality or irrationality of maximization of expected utility. However, a number of telling and powerful arguments have also been presented in the literature giving reasons for departing from 'consistency' of the kind demanded by the expected utility model (for arguments on both sides, see the collection of papers in Daboni et al. 1986). Allais (1953) has followed up his empirical critique of expected utility model as representation of actual behaviour by arguments in favour of the reasonableness of the departures, and more arguments on this have been outlined in recent years (see Allais and Hugen 1979; Stigum and Wenstop 1983; and Daboni et al. 1986). Also, the possibility of 'state-dependent utilities' has raised questions of a different sort, requiring reformulation of the original model (see Drèze 1974).

One of the important considerations that the expected utility model may leave out consists of 'counterfactual' information. One's 'disappointment', 'regret', and so on may well depend on what one anticipated and what did not occur. Earlier discussions of such criteria as 'minimax regret' (see Savage 1954) have been followed in recent years by various models of disappointment and regret (see, for example, Bell 1982; Loomes and Sugden 1982).

It is arguable that something which has not happened, but could have, should not really affect one's decision, and in particular, it is irrational to regret and sigh about what could have happened. But while it is indeed possible to argue that it is irrational to regret a past decision on the ground of what could have happened in the light of later information, nevertheless, if it is the case that one would willy-nilly regret the past decision if it turns out to be unfortunate, then it is not in any sense obviously irrational to *recognize* that fact and take that inescapable feeling into account. Clarity of analysis requires that we distinguish between (*a*) the rationality of what psychology we ought to have, and (*b*) the rationality of decisions, taking note of what psychology we might not be able to escape. Many counter-examples to expected utility behaviour presented in the literature relate – directly or indirectly – to mental-state considerations, for example, Allais (1953), MacCrimmon (1968), Bernard (1974), Drèze (1974), Tversky (1975), Machina (1981), McClennen (1983) and others.

One reason why the inclusion of mental states among the influences on choice is resisted is the idea that mental state is a particular interpretation of *utility* of which another – alternative – interpretation is given by the numerical representation of choice, with which the expected utility model is concerned. In the context of utilitarianism, the mental-state utility and the numerical representation of choice can indeed be seen as *alternatives,* as they have been viewed in the ethical literature. However, in terms of the description of the world, both mental states and choices are distinct parts of the reality, and the acknowledgement of the existence of one does not deny the existence of the other. Indeed, it is not unreasonable to ask how each might relate to the other. The states of affairs over which choices may be considered (including choices over lotteries of those states) may, quite importantly, include the mental states of the parties involved.

On the other hand, including such mental states in the description of states of affairs makes the scope of such conditions as 'strong independence' rather limited. Varying an *alternative* lottery (for example, L^3 vis-à-vis L^2) might affect the description of the 'prize' of a given lottery (L^1) through variations of mental states (now included in the outcome of L^1) related to considering and reflecting on the nature of the alternative (L^1 vis-à-vis L^2) and the corresponding disappointment, regret, and so on. If L^1 is no longer 'the same' in the two cases, then 'strong independence' would make no demand. Thus 'strong independence' may be saved only at the cost of making it often trivially fulfilled (see Sen 1985a). The same difficulty applies if strong independence is 'rescued' by including counterfactual information in describing states of affairs.

The basis of rationality implicit in expected utility calculation does, however, require descriptions of states of affairs in sufficient detail such that choices can be made taking all the relevant considerations into account. It can be argued, as indeed Peter Hammond (1986) has, that 'consequential' reasoning, taking into account all the relevant considerations, will push us in the direction of expected utility maximization. The important question is whether the relevant considerations would include either counterfactuals or mental states, and, if they do so, whether enough scope for the use of such conditions as 'strong independence' can be found to build up utility numbering in a way that would make the expected utility model work in practice. This is not a matter, obviously, of pure theory only, and much depends on the nature of people's psychology and what considerations might be regarded as rational, in taking note of the complexities of our psychology.

Concluding Remarks

Attempts at constructing models of rational behaviour have certainly played a creative part in reducing the intractability of unstructured assessment of (*a*) the demands of rationality, and (*b*) facts of actual behaviour. On the other hand, models of rational behaviour actually presented have tended to ignore some of the complexities that have to be faced. This problem arises even when no uncertainty is introduced into the picture.

Neither of the two standard views of rational behaviour – as 'consistent choice' or as 'self-interest maximization' – has emerged as being really adequate as representations of rationality or of actuality. Various suggestions as to the directions in which we might go were reviewed earlier. Although none of the suggestions are unproblematic, many fruitful avenues of investigation have certainly been identified in the critical literature.

These difficulties carry over to rational behaviour models accommodating uncertainty. The limitations of characterizing rational behaviour in terms of just internal consistency, as discussed in the context of choice under certainty, obviously would apply to the modelling of choice under uncertainty as well. Similarly, pursuit of self-interest cannot be seen us being *uniquely* rational in models of uncertainty, any more than they can be so seen when everything is certain. However, it is not really necessary that expected utility models be seen in terms of self-interest maximization, and indeed some writers, for example, Ramsey (1931), have explicitly repudiated that interpretation. In fact, what the expected utility models do concentrate on is 'consistency' in a very demanding sense, and in this context objections similar to the ones raised in models of choice *without* uncertainty can be raised a fortiori with uncertainty.

Rationality may be seen as demanding something other than just consistency of choices from different subsets. It must, at least, demand cogent relations between aims and objectives actually entertained by the person and the choices that the person makes. This problem is not eliminated by the terminological procedure of describing the cardinal representation of choices as the 'utility' of the person, since this does not give any independent evidence on what the person is aiming to do or trying to achieve.

A more difficult issue, as discussed in the context of certainty, concerns the *assessment* of aims and objectives pursued by a person, even if they are fully reflected in the choices actually made. As Patrick Suppes has put it, the standard normative model of expected utility 'can be satisfied by cognitive and moral idiots . . . Put another way, the consistency of computations required by the expected-utility model does not guarantee the exercise of judgement and wisdom in the traditional sense' (1984, pp. 207–8). Suppes argues in favour of moving to the Aristotelian view that the rational person acts 'in accordance with good reasons', and is not embarrassed by the fact that this leaves a certain amount of 'pluralism' in the possible approach to rationality.

In addition to those problems of rationality that are shared by models of certainty as well as uncertainty, there are some special problems that apply particularly to considerations of uncertain outcomes. The status of counterfactuals, and their influences on mental states, raise interesting and important questions as to what may or may not be relevant to take into account in rationally assessing alternative courses action.

While these problems were addressed earlier on in this paper, one issue that has not yet received much attention here concerns the nature of uncertainty itself. Reference was made earlier to the distinction between interpreting probabilities as degrees of belief, and interpreting them as frequencies. There are also other issues (see, for example, Levi 1982, 1987). Even the very idea of having beliefs about possible outcomes in the form of probabilities in a situation of partial ignorance raises some interesting philosophical questions. At the very least, it is possible to make a distinction that was made by Frank Knight (1921) between 'risk' and 'uncertainty', with probability distributions being specified in the case of the former but not in the latter case. Whether arguments such as 'insufficient reason' can permit one to *construct* probability distributions even when we do not start with them remains a hard question to settle.

The area of expectation formation is also one in which the demands of rationality are not easy to specify. In some models of rational behaviour, no requirements of rationality are imposed on expectations at all, and the problem of rationality arises only in taking note of the actual expectations in arriving at decisions regarding action. In models of 'adaptive expectations' a step is taken in the direction of making expectations responsive – in an intelligent way – on experience. What goes very much further than this is the assumption of 'rational expectation' by which each person anticipates what can, in some sense, be described as objective probabilities; see Muth (1961) and Lucas and Sargent (1982).

This approach not only raises the question as to what the philosophical status of objective probabilities might be, but also whether it is really a matter of *rationality* as such whether one is successful in guessing what the objective probabilities are. It is fair to say that the assessment of models of 'rational expectation' cannot be bused on the idea of rationality alone, since the demands of such a theory go well beyond the requirements of the use of reason, especially in a situation of ignorance. It is sensible enough to think that there are problems in models of behaviour in which people's expectations are systematically wrong, but to try to move from that recognition to one in which everyone manages to take note of objective probabilities fully is quite a dramatic step. Whether that step is worth taking in predicting actual behaviour might well be discussed and assessed in the light of the ability of such a theory to explain actual behaviour, but that, as we have already discussed, is a rather different problem from assessing the *rationality* us such of that behaviour.

In addition to the issue of the role of rationality involved in 'rational expectation' models, even the basic rational behaviour model (without such expectational assumptions), widely used in economics, raises, as we have seen, difficult – sometimes perplexing – questions. It is not hard to see the merit of trying to reduce a complex reality by characterizing rationality in rather narrow terms, but nor is it hard to fathom that such a narrowing might do grave injustice to the notion of rationality, which is, after all, one of the central concerns of human life.

We have to make a clear distinction between (*a*) what type of behaviour might be described as *rational,* and (*b*) what rational behaviour models might be useful in making predictions about *actual* behaviour. These different questions are not, of course, independent of each other. But the first step in pursuing their interrelations is to recognize the distinction between the two questions. What issues respectively arise in facing these distinct questions, and how they might possibly be related, were discussed earlier on in this article in the light of the existing literature. There was, however, no escape from noting the fact that the existing literature is indeed deeply incomplete in that real difficulties have been identified without providing an adequate structure for solutions. The need to go beyond the existing literature is apparent enough, but where to go is less clear.

See Also

▶ Philosophy and Economics
▶ Social Choice
▶ Welfare Economics

Bibliography

Akerlof, G.A. 1984. *An economic theorist's book of tales*. Cambridge: Cambridge University Press.

Allais, M. 1953. Le comportment de l'homme rational deviant to risque: Critique de postulates et axiomes de l'école Américaine. *Econometrica* 21: 503–546.

Allais, M., and O. Hugen, eds. 1979. *Expected utility hypotheses and the Allais paradox: Contemporary discussions of decisions under uncertainty with Allais' rejoinder*. Dordrecht: Reidel.
Arrow, K.J. 1951. An extension of the basic theorems of classical welfare economics. In *Proceedings of the second Berkeley symposium of mathematical statistics*, ed. J. Neyman. Berkeley: University of California Press.
Arrow, K.J. 1959. Rational choice functions and orderings. *Economics* 26: 121–127.
Arrow, K.J. 1971. *Essays in the theory of risk-bearing*. Amsterdam: North-Holland.
Arrow, K.J. 1982. Risk perception in psychology and economics. *Economic Inquiry* 20: 1–9.
Arrow, K.J. 1983. Behaviour under uncertainty and its implications for policy. In Stigum and Wenstop (1983).
Arrow, K.J. and F.H. Hahn. 1971. *General competitive analysis.* San Francisco: Holden-Day. Republished, Amsterdam: North-Holland, 1979.
Axelrod, R. 1984. *The evolution of cooperation*. New York: Academic Press.
Basu, K. 1977. Information and strategy in iterated prisoners' dilemma. *Theory and Decision* 8: 293–298.
Bell, D.E. 1982. Regret in decision making under uncertainty. *Operations Research* 30: 961–981.
Bernard, G. 1974. On utility functions. *Theory and Decision* 5: 205–242.
Brams, S.J. 1975. *Game theory and politics*. New York: Free Press.
Brennan, G., and L. Lomasky. 1985. The impartial spectator goes to Washington: Toward a Smithian theory of economic behavior. *Economics and Philosophy* 1: 189–211.
Broome, J. 1978. Choice and value in economics. *Oxford Economic Papers* 30: 313–333.
Broome, J. 1984. Uncertainty and fairness. *Economic Journal* 94: 624–632.
Campbell, R., and L. Sowden, eds. 1985. *Paradoxes of rationality and cooperation*. Vancouver: University of British Columbia Press.
Chipman, J.S., L. Hurwicz, M.K. Richter, and H.F. Sonnenschein, eds. 1971. *Preferences utility and demand*. Harcourt, Brace: New York.
Daboni, L., A. Montesano, and M. Lines, eds. 1986. *Recent developments in the foundations of utility theory and risk*. Dordrecht: Reidel.
Davidson, D., P. Suppes, and S. Siegel. 1957. *Decision making: An experimental approach*. Stanford: Stanford University Press.
Davis, L.M. 1977. Prisoners, paradox and rationality. *American Philosophical Quarterly* 14: 319–27. Also in Campbell and Sowden (1985).
Dawkins, R. 1982. *The extended phenotype*. Oxford: Clarendon Press.
Debreu, G. 1959. *A theory of value*. New York: Wiley.
Dore, R. 1983. Goodwill and the spirit of market capitalism. *British Journal of Sociology* 34: 459–482.
Drèze, J.H. 1974. Axiomatic theories of choice, cardinal utility and subjective probability: A review. In *Allocation under uncertainty: Equilibrium and optimality*, ed. J.H. Drèze. London: Macmillan.
Edgeworth, F. 1881. *Mathematical psychics*. London: Kegun Paul.
Elster, J. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
Elster, J. 1983. *Sour grapes*. Cambridge: Cambridge University Press.
Fishburn, P.C. 1970. *Utility theory and decision making*. New York: Wiley.
Fishburn, P.C. 1981. Subjective expected utility: A review of normative theories. *Theory and Decision* 31: 139–199.
Fisher, R.A. 1921. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, Series A 222: 309–68.
Friedman, M. 1953. *Essays in positive economics*. Chicago: Chicago University Press.
Friedman, M., and L.J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56: 279–304.
Gauthier, D. 1985. Maximization constrained: The rationality of cooperation. In Campbell and Sowden (1985).
Gibbard, A. and W.L. Harper. 1978. Counterfactual and two kinds of expected utility. In Hooker, Leach and McClennen (1978).
Hammond, P.J. 1976. Changing basics and coherent dynamic choice. *Review of Economic Studies* 43: 159–173.
Hammond, P.J. 1986. Consequentialism and rationality in dynamic choice under uncertainty. In *Social choice and public decision making: Essays in honor of K.J. Arrow*, vol. 1, ed. W. Heller, D. Starrell and R. Starr. Cambridge: Cambridge University Press.
Harsanyi, J.C. 1977. *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.
Helm, D. 1984. Predictions and causes: A comparison of Friedman and Hicks on method. *Oxford Economic Papers* 36(Supplement), 118–34.
Herzberger, H. 1973. Ordinal preference and rational choice. *Econometrica* 41: 187–237.
Hirschman, A.O. 1970. *Exit, voice, and loyalty.* Cambridge, MA: Harvard University Press.
Hirschman, A.O. 1982. *Shifting involvements*. Princeton: Princeton University Press.
Hirschman, A.O. 1983. Against parsimony: Three easy ways of complicating some categories of economic discourse. *American Economic Review* 74 (May): 89–96.
Hirshleifer, J. 1977. Economics from a biological viewpoint. *Journal of Law and Economics* 20: 1–52.
Hooker, C.A., J.J. Leach, and E.F. McClennen, eds. 1978. *Foundations and applications of decision theory.* Dordrechi: Reidel.
Houthakker, H.S. 1950. Revealed preference and the utility function. *Economica* 15: 159–174.
Houthakker, H.S. 1956. On the logic of preference and choice. In *Contributions to logic and methodology in honor of J.J. Bochensk*, ed. A. Tymieniecke. Amsterdam: North-Holland.

Howard, N. 1971. *Paradoxes of rationality*. Cambridge, MA: MIT Press.
Jeffrey, R.C. 1965. *The logic of decision*. New York: McGraw Hill.
Jeffrey, R.C. 1983. *The logic of decision*. 2nd ed. Chicago: University of Chicago Press.
Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47: 263–291.
Kahneman, D., P. Slovik, and A. Tversky. 1982. *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
Kanger, S. 1976. *Preference based on choice*. Mimeo: Uppsala University.
Keynes, J.M. 1921. *A treatise on probability*. London: Macmillan.
Knight, F. 1921. *Risk, uncertainty and profit*. New York: Houghton Mifflin.
Kornai, J. 1971. *Anti-equilibrium*. Amsterdam: North-Holland.
Kreps, D.M., P. Milgrom, J. Roberts, and R. Wilson. 1982. Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory* 27: 245–252.
Latsis, S.J., ed. 1976. *Method and appraisal in economics*. Cambridge: Cambridge University Press.
Leibenstein, H. 1976. *Beyond economic man: A new foundation for microeconomics*. Cambridge, MA: Harvard University Press.
Levi, J. 1975. Newcomb's many problems. *Theory and Decision* 6: 161–175.
Levi, J. 1982. Ignorance, probability and rational choice. *Synthesis* 53: 287–417.
Levi, J. 1987. *Hard choices*. Cambridge: Cambridge University Press.
Loomes, G., and R. Sugden. 1982. Regret theory: An alternative theory of rational choice. *Economic Journal* 92: 805–824.
Lucas, R.E., and T.J. Sargent. 1982. *Rational expectation and econometric practice*. London: Allen & Unwin.
Luce, R.D., and H. Raiffa. 1957. *Games and decisions*. New York: Wiley.
MacCrimmon, K.R. 1968. Descriptive and normative implications of decision theory postulates. In *Risk and uncertainty*, ed. K. Borch and J. Mossin. London: Macmillan.
Machina, M. 1981. 'Rational' decision making vs. 'rational' decision modeling? *Journal of Mathematical Psychology* 24: 163–175.
Machina, M. 1982. 'Expected utility' analysis without the independence axiom. *Econometrica* 50: 277–323.
McClennen, E.F. 1983. Some-thing doubts. In Stigum and Wenstop (1983).
McClennen, E.F. 1985. Prisoner's dilemma and resolute choice. In Campbell and Sowden (1985).
McPherson, M.S. 1982. Mill's moral theory and the problem of preference change. *Ethics* 92: 252–273.
Margolis, H. 1982. *Selfishness, altruism and rationality*. Cambridge: Cambridge University Press.
Marschak, J. 1946. Von Neumann's and Morgenstern's new approach to static economics. *Journal of Political Economy* 54: 91–115.
Matthews, R.C.O. 1984. Darwinism and economic change. *Oxford Economic Papers* 36(Supplement), 91–117.
Maynard Smith, J. 1982. *Evolution and the theory of games*. Cambridge: Cambridge University Press.
Morishima, M. 1982. *Why has Japan 'Succeeded'? Western technology and Japanese ethos*. Cambridge: Cambridge University Press.
Muth, J.F. 1961. Rational expectations and the theory of price movements. *Econometrica* 29: 315–335.
Nagel, T. 1970. *The possibility of altruism*. Oxford: Clarendon Press.
Nozick, R. 1969. Newcomb's problem and two principles of choice. In *Essays in honor of Carl G. Hempel*, ed. N. Rescher. Dordrecht: Reidel.
Parfit, D. 1984. *Reasons and persons*. Oxford: Clarendon Press.
Radner, R. 1980. Collusive behaviour in non-cooperative epsilon-equilibrium of oligopolies with long but finite lives. *Journal of Economic Theory* 22: 136–154.
Ramsey, P.P. 1931. Truth and probability. In *The foundations of mathematics and other logical essays*, ed. F.P. Ramsey. London: Kegan Paul.
Richter, M.K. 1971. Rational choice. In Chipman et al. (1971).
Samuelson, P.A. 1938. A note on the pure theory of consumers' behaviour. *Economica* 5: 61–71.
Savage, L.J. 1954. *The foundations of statistics*. New York: Wiley.
Schelling, T.G. 1978. *Micromotives and macrobehavior*. New York: Norton.
Schelling, T.C. 1984. Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review* 74: 1–11.
Schick, F. 1984. *Having reasons: An essay and rationality and sociality*. Princeton: Princeton University Pess.
Scitovsky, T. 1976. *The joyless economy*. London: Oxford University Press.
Sen, A.K. 1971. Choice functions and revealed preference. *Review of Economic Studies* 38: 307–317.
Sen, A.K. 1973. Behaviour and the concept of preference. *Economica* 40: 241–59. Repr. in Sen (1982).
Sen, A.K. 1974. Choice ordering and morality. In *Practical reason*, ed. S. Körner. Oxford: Blackwell. Repr. in Sen (1982).
Sen, A.K. 1977. Rational fools: A critique of the behavioural foundations of economic theory. *Philosophy and Public Affairs* 6: 317–44. Repr. in Sen (1982).
Sen, A.K. 1982. *Choice, welfare and measurement*. Oxford: Blackwell/Cambridge, MA: MIT Press.
Sen, A.K. 1985a. Rationality and uncertainty. *Theory and Decision* 18: 109–27. Repr. in Daboni, Montesano and Lines (1986).
Sen, A.K. 1985b. Goals, commitment and identity. *Journal of Law Economics and Organization* 1: 341–355.
Sen, A.K. 1987. *On ethics and economics*. Oxford: Blackwell.
Shackle, G.L.S. 1938. *Expectations. investment and income*. Cambridge: Cambridge University Press.

Shackle, G.L.S. 1952. *Expectations in economics*. 2nd ed. Cambridge: Cambridge University Press.
Simon, H.A. 1957. *Models of man*. New York: Wiley.
Simon, H.A. 1979. *Models of thought*. New Haven: Yale University Press.
Simon, H.A. 1983. *Reason in human affairs*. Oxford: Blackwell.
Smale, S. 1980. The prisoner's dilemma and dynamic systems associated to non-cooperative games. *Econometrica* 48: 1617–1634.
Smith, A. 1776. In *An inquiry into the nature and causes of the wealth of nations*, ed. R.H. Campbell and A.S. Skinner. Oxford: Clarendon Press, 1976.
Smith, A. 1790. In *The theory of moral sentiments*, ed. D.D. Raphael and A.L. Macfie. Oxford: Clarendon Press, 1974.
Stigum, B.P., and F. Wenstop, eds. 1983. *Foundations of utility and risk theory with applications*. Dordecht: Reidel.
Sugden, R. 1985. Why be consistent? A critical analysis of consistency requirements in choice theory. *Economica* 52: 167–183.
Suppes, P. 1984. *Probabilistic metaphysics*. Oxford: Blackwell.
Suzumura, K. 1983. *Rational choice, collective decisions and social welfare*. Cambridge: Cambridge University Press.
Tversky, A. 1975. A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis* 9: 163–173.
Uzawa, H. 1956. A note on preference and axioms of choice. *Annals of the Institute of Statistical Mathematics* 8: 35–40.
von Neumann, J., and O. Morgenstern. 1947. *Theory of games and economic behavior*. Princeton: Princeton University Press.
Watkins, J. 1974. Comment: Self-interest and morality. In *Practical reason*, ed. S. Körner. Oxford: Blackwell.
Watkins, J. 1985. Second thoughts on self-interest and morality. In Campbell and Sowden (1985).
Winch, D. 1978. *Adam Smith's politics*. Cambridge: Cambridge University Press.
Wong, S. 1978. *Foundations of Paul Samuelson's revealed preference theory*. London: Routledge.

Rational Choice and Political Science

Susanne Lohmann

Abstract

'Rational choice in political science' stands for the application of the economics approach in the study of political phenomena. The research program is to *rationalize* collective behaviour that comes across as stupid or counterproductive. In its highbrow (esoteric) variant, rational choice is on the way out in political science. In its low-brow (sensible) variant, rational choice is here to stay, not as the dominant approach, but as one of three equal, and complementary, approaches: the rationalist approach, which focuses on individual agency; the culturalist approach, which centres on collective identities; and the structuralist approach, which emphasizes historical institutionalism.

Keywords

Arrow, K. J.; Asymmetric information; Behavioural economics; Buchanan, James; Tullock, Gordon; Olson, Mancur; Hayek, F. A.; Aristotle; Hobbes, T.; Coase Theorem; Collective action; Complex systems approach in social science; Complexity; Congressional committees; Culturalist approach in political science; Historical institutionalism; Collective identities; Individual agency; Institutionalism; Agency; Public choice; Externalities; Government intervention; Politics of monetary policy; Logic of Collective Action; Supply of and demand for collective action; Greed, rationality, and equilibrium; Keynesian approach; Hayekian approach; Developing countries; Social complexity; Post-autistic economics movement; Political science; Deadweight loss; Development economics; Experimental Economics; Free-rider problem; Game theory; Government failure; Independence of irrelevant alternatives; Market failure; Political economy; Positive political theory; Preference aggregation; Rational choice; Rational choice in political science; Representative agent; Self-interest; Social choice; Social welfare function; Special interests; Structuralist approach in political science

JEL Classifications

D01; Z13; D7

Chemical plants are vulnerable to terrorist attacks. Two of the most dangerous facilities are located in

R

Dallas, in the US state of Texas, right next to Joe Barton's Congressional district. They constitute a risk for more than one million people. In 2005, Barton used his clout as chair of the House Energy and Commerce Committee to block chemical plant security legislation. A Republican, Barton served as a consultant for an oil and gas company before he was elected to Congress in 1984, and in the following decades he received more than $1.8 million in campaign contributions from the energy and chemical industries. Barton routinely sides with the energy industry at the expense of his constituents, and you can forget about the welfare of the people represented by his colleagues in Congress (Cohen 2005).

For Barton to be in the position to benefit a special interest at the expense of a multitude, he must enjoy majority support in both his district and Congress. What is the logic whereby he gains such support?

'Rational choice in political science' stands for the import of the economics paradigm into the political science discipline, or the application of the economics approach in the study of political phenomena. The research programme is to *rationalize* collective behaviour that comes across as stupid or counterproductive. My purpose here is to spell out how this research programme is playing out in political science – or rather how it played out, for the research programme has recently lost much of its vitality (in its extreme form it is dead).

Economists and political scientists often use the same labels to denote different things and different labels to denote the same thing. For this reason, it is useful to start with some definitions: what do social choice, public choice, political economy, and positive political theory stand for, and how do they relate to 'rational choice in political science'?

I shall illustrate these labels in the context of the scientific life cycle of rational choice in political science, which describes the usual arc of fringe, vibrancy, maturity, ossification, and renewal. Over time, the research programme cycled in its emphasis on external and internal scientific progress, as it moved from solving real-world puzzles to theoretically refining the solutions and back again; the cycle includes a forward movement.

In the economics discipline, behavioural and experimental economics have recently relaxed some of the more extreme greed-and-rationality assumptions. In political science, rational choice took a different turn. Today, rational choice in its high-brow (esoteric) variant is on the way out, in part because leading rational choice theorists are 'holier than the Pope' in their refusal to join the rather more relaxed approach to economics. In its low-brow (sensible) variant, rational choice is here to stay, but it has largely shed its imperialist ambitions. Instead of emerging as the dominant approach, rational choice coexists more or less peacefully with one of three complementary approaches: the rationalist approach, which focuses on individual agency; the culturalist approach, which centers on collective identities; and the structuralist approach, which emphasizes historical institutionalism. (My description of rational choice scholarship as 'high-brow' and 'low-brow,' or 'esoteric' and 'sensible,' is not meant to express approval or disapproval; the two types of scholarship complement each other as they contribute to internal and external scientific progress.)

The Scientific Life Cycle of Social Choice, Public Choice, and Political Economy in the Economics Discipline

Because the scientific life cycle of rational choice in political science is an offshoot of the scientific life cycle of social choice, public choice, and political economy in the economics discipline, it is useful to start with an account of the latter.

Social choice was research-active from the 1950s through the 1970s; public choice, from the 1970s through the 1990s; political economy, from the 1980s through the 2000s. Dennis Mueller's textbook *Public Choice III* (2003) covers social choice and public choice. *(Public Choice III* is the third, and most comprehensive, edition of *Public Choice,* which was published in 1979.) Torsten Persson and Guido Tabellini's textbook *Political Economics* (2000) lays out the political economy programme.

When social choice started out, the reigning practice in economics was to derive normative statements about economic policy by maximizing a social welfare function subject to a set of economic constraints. At the time, it was taken for granted that people's preferences could be summarized by a social welfare function. Kenneth Arrow demonstrated, to the contrary, that it is impossible to represent people's preferences with a social welfare function that fulfills plausible criteria such as independence of irrelevant alternatives; this impossibility result holds if people's preferences are diverse. (The independence-of-irrelevant-alternatives criterion prohibits the social preference over two alternatives from switching places if the individuals' preferences over the two alternatives stay the same even as a third alternative is added to, or dropped from, the set of alternatives under consideration.) Social welfare functions subsequently went out of fashion among economists. (Actually, in many subfields of economics they returned through the side door; for example, in macroeconomics the illegitimacy of assuming a social welfare function was elegantly circumvented by assuming that the macroeconomy can be summarized by a representative agent.)

Social choice also demonstrated that voting rules affect voting outcomes. If people's preferences are sufficiently diverse, there exists no such thing as a neutral voting rule that will 'simply' aggregate people's preferences. At first blush, this result seems rather worrisome because of its potential to undercut the legitimacy of outcomes arrived at by democratic means. But we shall see how this insight would get picked up productively – after all, if institutions can warp democratic decision-making, this raises the possibility of designing political institutions to serve a corrective function.

Social choice consisted largely of mathematical exercises with little economic content; its concern with preference aggregation does not relate all that well to the standard economic concern with scarcity and constraint. Public choice, in comparison, employed microeconomic theory, and it was geared towards extending economic assumptions of self-interest and rationality to the political arena, with the idea of treating political and economic actors symmetrically. Before public choice entered the fray, economists were in the habit of spelling out what actions a benevolent dictator should take when he or she (or it?) encounters *market failure* due to externalities, information asymmetries and the like.

One early argument against government intervention consisted of the Coase Theorem, which implies that the system of economic actors will endogeneously adjust to internalize externalities (assuming that there exists a system of well-defined property rights and negligible transaction costs). The Coase argument can be – has been – exported to the political sector. For example, if the underlying problem in the political market consists of an information asymmetry between policymakers and voters, then information providers will have incentives to enter the political market, and voters will have incentives to take information cues from them (Wittman 1989).

Public choice theorists, most prominently among them James Buchanan and Gordon Tullock, refused to see market failure behind every bush – in economic markets, that is. At the same time, their minds would surely boggle at the idea of political markets being self-correcting. In their eyes, *government failure* loomed large. The public choice argument against government intervention is that government consists of self-serving politicians, bureaucrats, and special interests who are poorly held in check by ignorant voters. Public policy is thus riddled with biases and loopholes benefiting special interests at the expense of consumers and taxpayers.

Public choice theorists also proposed institutional solutions to government failure. Examples are the flat tax, the gold standard, and constitutional limits on government borrowing. The typical proposal seeks to tie the hands of politicians and bureaucrats and to create transparency vis-à-vis voters or other audiences. It forgoes the benefits of efficiency, equity and flexibility. That is, it does not offer the best solution for some idealized world lorded over by benevolent experts; it does not go out of its way to do good for the poor; and it disallows Keynesian-style economic stabilization and micro-intervention. It does, however, come with the potential to prevent self-serving politicians and bureaucrats from doling out goodies to special interests precisely because voters, or other audiences, can easily monitor slippages, make a public fuss, and 'vote the bastards out'. The simplicity and transparency of the proposed institutions create a political cost of defecting from them.

It is worthwhile appreciating public choice for driving home this important point: as we compare different institutional solutions, we must take into account their relative political corruptibility. Public choice spelled out how the policy process is warped by collective action and political institutions. For example, if small groups have an easier time solving the free-rider problem of collective action than do large groups, then policy will be biased in favour of special interests (Olson 1965); for example, too, the power of special interests is the result of Congressional committees being captured by high demanders, that is, members of Congress who represent constituencies (voters and campaign contributors) with a high demand for certain kinds of government handouts (Shepsle and Weingast 1987).

From the outset public choice thus stood on two legs: one leg was about inserting politics into apolitical models of economic policy, as in, 'the politics of monetary policy'; the other was about applying rational choice to political behaviour and institutions, as in, 'the logic of collective action'.

Where public choice applied microeconomic theory, as in, 'the supply of and demand for collective action', political economy made use of game theory, as in, 'greed, rationality, equilibrium'. The result was a higher standard of spelling out the rationality of political actors, including their informational states, and of making sure that all of their strategies and beliefs are consistent with each other so that the strategies and beliefs constitute an equilibrium.

For example, the story that policymakers pander to special interests at the expense of voters does not necessarily make sense if voters follow a voting strategy by which they vote for the incumbent when they are well off and for the challenger when they are hurting. To make this story fly, one has to specify how a policymaker can increase her re-election chances by taking something of value from the large mass of voters; losing a little of bit of it along the way (this is the deadweight loss created by redistribution, which generally distorts people's economic choices); and giving the remainder to special interests: why wouldn't the policymaker lose more votes among the large mass of voters than she would gain among the special interests (Lohmann 1998)? And if special interests are powerful because of campaign contributions, why wouldn't voters reject a policymaker who is loaded with campaign contributions – after all, campaign contributions are a sign that the policymaker is pandering to special interests at their expense? In the same vein, if special interest handouts are the result of high demanders hogging Congressional committees, why would a majority in Congress go along with bills that benefit the committee members' constituents at the expense of *their* constituents? And why does a Congressional majority allow high demanders to self-select onto Congressional committees in the first place?

Political economy also differed from public choice by taking a balanced view of market and government failures. For example, when economic markets fail to aggregate distributed information about the mapping of economic policy into policy outcomes, then special interests or high demanders on Congressional committees may well supply the requisite information, with the result that the quality of economic policy improves (because policymakers are well-informed) even though policy outcomes are biased (because policymakers pander to special interests or high demanders) (Gilligan and Krehbiel 1987). The implication is that we should not automatically assume that special interests and high demanders on Congressional committees are a Bad Thing; we need to consider the workings of the economic and political system as a whole, in which case a Political Bad might cancel out an Economic Bad, and the net effect is a Good Thing. Because political economy traded off the gains and losses of imperfect markets *and* imperfect politics, it came up with more complex and more flexible institutions than did public choice.

By the mid-1990s, the newly prominent subfield of experimental economics had amassed enough evidence to challenge the assumptions and predictions of standard economics models, and behavioural economics proceeded to explain the anomalies by relaxing the assumptions of greed and rationality (less so the assumption of equilibrium) in favour of a richer set of psychological motivations and cognitive limitations. All of this activity served to undercut the political economy programme of producing ever more refined rationality-and-equilibrium explanations of market-cum-government failure. Today, the extreme application of the game theory paradigm to political phenomena is passé. The cutting edge lies in employing richer models of human behaviour to understand the various forms of collective action we observe in reality, that is, in laboratory experiments and in the field.

Even as the one kind of political economy (the kind that applies old-style game theory to political phenomena) is intellectually stagnant, the other kind of political economy (the kind that inserts politics into models of economic policy) has been busy expanding into the political economy of development. Whereas public choice was largely focused on the developed countries, or the rich capitalist democracies, political economy increasingly included the developing countries, many of which were governed (some still are) by tin-pot dictators, military cliques, and the like. Whereas public choice was concerned about the discrepancy between economic theory and practice in developed democracies, the political economy of development worried about the disparities in economic performance across countries and sought to explain why and how some countries grew rich (why *these* countries, why *now*?) even as others remained poor.

Development economists who pushed this story, or variants of it, naturally appreciated the fact that well-functioning market economies rely on well-functioning governments, just as they naturally appreciated the fact that government in developed countries is functioning extremely well, both in historical and cross-country comparison. Such appreciation is a 180-degree reversal of the anti-government bias that permeated the public choice programme.

Whereas public choice had a Hayekian flavour to it (put into place simple institutions and let the economy do the rest), political economy took a rather more Keynesian approach (derive optimal institutions that will surgically correct the political economy). We are experiencing another reversal right about now, that is, a revival of the Hayekian approach. William Easterly's *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good* (2006) stands for the new bottom-up thinking (government is 'governance by the local people'), though it is clear from the popularity of Jeffrey Sachs's *The End of Poverty* (2005) that the old top-down thinking (government is 'management by benevolent experts') is not quite dead yet.

The emerging new approach – let us call it social complexity – stands in a tension with behavioural economics. The latter likes to make complicated assumptions about what is going on in people's heads even as it preserves the assumption of equilibrium, which implies that people have a complete and shared understanding of their environment (people might suffer from cognitive biases, but they best-respond to each other cognitive biases, and it all comes together very neatly). In contrast, social complexity likes to make simple assumptions about what going on in people's heads: people are relatively fixed in their behaviours, and their 'ways of seeing' the world are incomplete and diverse and partially inconsistent with each other (Hayek 1945). Social complexity is actually closer to what the economics discipline used to stand for historically:

> If social phenomena showed no order except insofar as they were consciously designed, there would indeed be no room for theoretical sciences of society and there would be, as is often argued, only problems of psychology. It is only insofar as some sort of order arises as a result of individual action but without being designed by any individual that a problem is raised which demands theoretical explanation . . . (Hayek 1955, p. 39)

R

The Scientific Life Cycle of Rational Choice in Political Science

Now that I have reviewed how rational choice evolved in the economics discipline, let me examine how it spilled over into the political science discipline.

In the political science discipline, social choice and public choice led a peripheral existence for decades; they still do. It was only in the late 1980s and early 1990s that political economy and positive political theory exploded onto the stage. Indeed, rational choice briefly looked as if it would take over political science, only to lose influence in the early 2000s, especially in its high-brow variant; the low-brow variant has been folded into political science for the long run. Low-brow rational choice theorists, including the political economy crowd, like to use the Persson and Tabellini textbook. The high-brow crowd prefers the two-volume effort by David Austen-Smith and Jeffrey Banks, *Positive Political Theory I: Collective Preference* (1999) and *Positive Political Theory II: Strategy and Structure* (2005).

The two legs of public choice and political economy – inserting politics into models of economic policy and using the economics paradigm to model political phenomena – can be found in political science, albeit with different labels assigned to them. The label 'political economy' has come to stand for inserting politics into models of economic policy, as in 'political economy of international trade'; this is typically done in a rational choice fashion, though there are some Marxist leftovers who call themselves political economists. The label 'positive political theory' denotes the use of rational choice to model political phenomena, as in 'positive political theory of Congressional committees'. Why positive? To distinguish positive political theory from political theory, a subfield of the political science discipline that corresponds to the subfield political philosophy in the philosophy discipline, which is concerned with, for example, interpreting Aristotle's *Politics* or Hobbes's *Leviathan* – and also with expanding the culturalist approach, which I shall describe later.

Why did political economy and positive political theory succeed in gaining significant market share in the political science market even as social choice and public choice were reduced to eking out a peripheral existence? Social choice modelled political behaviour and institutions in a way that was quite simply too abstract relative to the thick knowledge of political behaviour and institutions held by practising political scientists: at the time there was no research tradition in political science that could latch onto the idea that it might be interesting to prove the impossibility of a social welfare function; and as for the possibility of manipulating voting outcomes by fiddling with voting rules, what else is new?

Public choice was rather more practical in its orientation, but it was for the longest time rejected as a right-wing enterprise (Lowi 1992).

Political economy, with its more balanced take on market versus government failure, turned out to be politically more palatable. Perhaps more importantly, political economy – sailing under the flag of positive political theory – supported more refined models of political behaviour and institutions: from the perspective of a political scientist, it is not terribly interesting for a benevolent dictator to be replaced by a unitary-actor self-serving politician; it is exciting to explain why a majority in Congress would rationally constrain itself by applying closed rule to votes on proposals coming out of Congressional committees and to spell out the conditions under which the majority would allow for closed rule versus open rule (Gilligan and Krehbiel 1987). (Under closed rule, a committee proposal must be voted up or down, with amendments prohibited. Open rule allows for amendments.)

The mid-1990s saw the first stirrings of a counter-reaction to rational choice. Donald Green and Ian Shapiro's *Pathologies of Rational Choice Theory* (1994) took potshots at rational choice. Leading rational choice theorists fought back, their responses were collected in a special issue of *Critical Review* (Friedman (1995, 1996). The two sides pretty much talked past each other, with Green and Shapiro emphasizing the empirical silliness of many rational choice models (and getting some of the models wrong) and the rational choice elite celebrating the theoretical rigour of rational choice models (and rejecting Green and Shapiro for not understanding their models).

In the early 2000s, in parallel to the post-autistic economics movement in the economics discipline, the *Perestroika* movement (Monroe 2005) emerged seeking to liberate political science from rational choice. But if rational choice lost steam in political science, it was for a different reason than in the economics discipline. The behavioural and experimental economics revolution is not happening in political science because the leading rational choice theorists are all too heavily invested in the hyper-rational variant of the research programme. There exist political scientists who combine psychology and politics, but they are coming out of a different research tradition, one that is oblivious to economics: they rely on surveys to study mass opinion, and they are uninterested in running laboratory experiments on games that relax greed and rationality while preserving equilibrium for the simple reason that it never occurred to them that greed-rationality-equilibrium is an interesting benchmark in the first place.

Instead, people got fed up with rational choice for its esotericism. Its practitioners increasingly scored points by refining each others' theoretical models rather than by relating their models to urgent substantive problems in the real world.

Political science, as compared to the economics discipline, has a tendency to let a thousand flowers bloom: it has always supported a greater diversity of approaches in the leading doctoral programmes. Consistent with this diversity, rational choice is not actually going out of business. But it is the low-brow variant of rational choice that is surviving by combining 'sensible' rational choice arguments (such as the idea that collective action is subject to a free-rider problem) with an in-depth substantive understanding of the issues. In parallel to the emergence of the political economy of development in the economics discipline, the focus of attention in political science has shifted away from the kind of positive political theory that mostly made its living in the subfields of American politics (especially in the subfield of Congressional studies) and international relations, and towards applying rational choice in the field of comparative politics, with an emphasis on developing countries.

Rational choice in its extreme form came with imperialist ambitions. Today, there is a firm understanding, at least in political science if not in the economics discipline, that rational choice is just one approach, with strengths in some domains and weaknesses in other domains, which is where alternative approaches come to life. Complementing the rationalist approach are the culturalist and structuralist approaches (Lichbach and Zuckerman 1997; Lichbach 2003).

By way of illustrating the culturalist approach, consider the massive social change that occurred in the United States over the second half of the 20th century – consider the civil rights movement, the women's movement, and assorted sexual liberation movements (gay, lesbian, bisexual, transgender), and contemplate the prevailing attitudes 'then and now' towards assorted ethnic minorities (African-American, Jewish, Native-American, Asian-American, Hispanic). Economists will glibly talk about herding effects and information cascades. But a discipline that is (variously) defined as being about scarcity or individual agency or greed-rationality-equilibrium simply does not carry much purchase when it comes to explaining such dramatic changes in collective identities.

Today, society is mired in religious conflict, both domestically and internationally. Think of the divide between the red states and blue states in the United States, that is, the states located in the vast middle of the country, whose voters predominantly back the Republican Party, and the states located on the East and West Coasts, whose voters for the most part support the Democratic Party. Think also of the divide between Islam and the West. We can talk about 'the supply of and demand for religion' or 'the rational choice of religion', but the intellectual action clearly lies someplace else, where a culturalist approach gives us a better purchase on reality.

By way of illustrating the structuralist approach, let us take a look at the current popularity of exporting democracy and building democratic institutions. In the West, the emergence of democracy took a couple of centuries (more if you include Greece and Rome). Why so long? It turns out that institutions are not designed by experts and plunked down by bureaucrats in the same way that, say, bridges are conceived by engineers and built by construction workers. Institutions are the product of social conflict and social movements, and they come into existence by spreading across the minds of a people. Social movements resolve social conflicts by locking in structures which subsequently are taken for granted, as if they exist 'naturally'. In a well- functioning democracy, not everything is up for grabs all the time. There are huge swathes of society in which people simply play out the roles assigned to them by the structures they are embedded in. In these domains, 'rational choice' and 'individual agency' are grossly inadequate concepts. (Indeed some of the more disastrous interventions of academic economists in the real world – I have in mind the post-Communist transition to capitalism and democracy – derive from an impoverished understanding of historically evolved structures.)

We are coming full circle here, for the structuralist approach has a Hayekian touch to it. There are numerous indicators suggesting that a complex systems approach is on the rise in the social sciences. For example, as of 2006 UCLA supported a new undergraduate interdepartmental degree programme on Human Complex Systems, George Mason University, a new doctoral programme on Social Complexity.

Let me conclude. The strengths and weaknesses of rational choice in political science correspond to those of the economics approach. Rational choice promises to make political science more scientific, as in: universally applicable, theoretically rigorous, cumulatively progressive. Thanks to the scholarship in social choice and public choice and political economy, which spilled over into political science, our understanding of collective action and political institutions is light years ahead of where it was in the 1950s.

Just like economics, however, rational choice in its extreme form (greed, rationality, equilibrium *über Alles*) is a problem. It is blind to thick and local knowledge; it disdains culture and history; and it has a tendency to degenerate into internal scientific progress rather than producing external scientific progress. Rational choice deserves to survive in political science, but it is just as well that it no longer overshadows complementary 'ways of seeing' the political world.

See Also

▶ Collective Action
▶ Political Institutions, Economic Approaches to
▶ Public Choice
▶ Social Choice

Bibliography

Austen-Smith, D., and J. Banks. 1999. *Positive political theory I: Collective preference*. Ann Arbor: University of Michigan Press.

Austen-Smith, D., and J. Banks. 2005. *Positive political theory II: Strategy and structure*. Ann Arbor: University of Michigan Press.

Cohen, A. 2005. A lawmaker works, oddly enough, to keep his voters' backyards dangerous: 'Smokey Joe' and the politics of corporate profit. *The New York Times,* 26 May, A28.

Easterly, W. 2006. *The white man's burden: Why the West's efforts to aid the rest have done so much Ill and so little good*. New York: Penguin.

Friedman, J. (ed.). 1995. Rational choice theory and politics. Special issue of *Critical Review* (vol. 9, nos. 1–2).

Friedman, J. (ed.). 1996. *The rational choice controversy: Economic models of politics reconsidered.* New Haven: Yale University Press.

Gilligan, T., and K. Krehbiel. 1987. Collective decision-making and standing committees: an informational rationale for restrictive amendment procedures. *Journal of Law, Economics and Organization* 3: 287–335.

Green, D., and I. Shapiro. 1994. *Pathologies of rational choice theory: A critique of applications in political science*. New Haven: Yale University Press.

Hayek, F.A. 1945. The use of knowledge in society. *American Economic Review* 35: 519–530.

Hayek, F.A. 1955. *The counter-revolution of science: Studies on the abuse of reason*, 1964. New York: Glencoe Free Press.

Lichbach, M.I. 2003. *Is rational choice theory all of Social Science?* Ann Arbor: University of Michigan Press.

Lichbach, M.I., and A.S. Zuckerman, eds. 1997. *Comparative politics: Rationality, culture, and structure*. New York: Cambridge University Press.

Lohmann, S. 1998. An information rationale for the power of special interests. *American Political Science Review* 92: 809–827.

Lowi, T.J. 1992. The state in political science: How we become what we study. *American Political Science Review* 86: 1–7.
Monroe, K.R. 2005. *Perestroika! The raucous rebellion in political science*. New Haven: Yale University Press.
Mueller, D. 2003. *Public choice III*. 3rd ed. Cambridge, MA: Cambridge University Press.
Olson, M. 1965. *The logic of collective action*. Cambridge, MA: Harvard University Press.
Persson, T., and G. Tabellini. 2000. *Political economics: Explaining economic policy*. Cambridge, MA: MIT Press.
Post-Autistic Economics Network. Home Page. Available online at. http://www.paecon.net. Accessed 11 Feb 2007.
Sachs, J. 2005. *The end of poverty*. New York: Penguin.
Shepsle, K., and B. Weingast. 1987. The institutional foundations of committee power. *American Political Science Review* 81: 86–108.
Wittman, D. 1989. Why democracies produce efficient results. *Journal of Political Economy* 97: 1395–1424.

Rational Choice and Sociology

Peter Hedström and Charlotta Stern

Abstract

Rational-choice theorizing has a long tradition within sociology, but has always been controversial and contested. Yet it has influenced the theoretical vocabulary of the discipline at large and has made deep inroads into some important sociological areas such as social movements, social mobility, and religion. Most sociological rational-choice theories assume that actors act rationally in a broad sense, and focus on the aggregate outcomes that individual actors in interaction with one another are likely to bring about. This article reviews the most important contributions to the rational-choice tradition in sociology, and briefly discusses its historical past and its likely future.

Keywords

Assumptions; Behavioural economics; Blau, P; Boudon, R; Class; Coleman, J; Collective action; Economic sociology; Elster, J; Expected utility hypothesis; Experimental economics; Explanation; Free rider problem; General equilibrium; Homans, G; Implicit contracts; Methodological individualism; Olson, M; Pareto, V; Parsons, T; Preferences; Public choice; Public goods; Rational choice and political science; Rational-choice sociology; Reciprocity; Religion, sociology of; Social interactions; Social mobility; Social networks, sociology of; Social norms; Transaction costs; Trust; Weber, M

JEL Classification

A12; A14; Z12; Z13; D01

Rational-choice sociology is the branch of sociology which is most thoroughly influenced by economic theory. Yet it is not simply an application of economic theory to the explanation of social phenomena. Rational-choice sociology consists of a diverse set of theories only some of which can be said to have been imported from economics. The common denominator of rational-choice sociologists is that they use explanatory models in which actors are assumed to act rationally, in a wide sense of that term. Unlike in many other sociological theories, actors are not assumed to be governed by causal factors operating behind their backs, but are seen as conscious decision makers whose actions are significantly influenced by the costs and benefits of different action alternatives.

Most rational-choice sociologists do not seek to explain the actions of single individuals. The focus instead is on explaining macro-level or aggregate outcomes such as the emergence of norms, segregation patterns, or various forms of collective action. To make sense of outcomes like these, however, rational-choice sociologists focus on the actions and interactions that brought them about.

The Emergence of Rational-Choice Sociology

Rational choice-inspired theorizing has a long tradition within sociology. Max Weber, one of the founders of sociology, argued for the importance of basing sociological explanations on clearly articulated ideas about rational action (Weber 1922). Only since the 1980s, however, have we seen the emergence of a more clearly defined rational-choice approach within sociology. Given the constraints imposed by the format of this article, we are not able to give due attention to the range of work produced by rational-choice sociologists. We instead single out a few contributions that have been particularly important for the development of the approach.

Some of the contributions that proved important for the development of rational-choice sociology were not themselves based on rational-choice assumptions. One case in point is the work of George Homans (for example, 1958, 1964). At the height of his career Homans was a highly visible and influential sociologist who made many substantive and theoretical contributions to the discipline. Unlike many of his contemporaries he argued that sociological explanations should take the form of deductive arguments based on clearly explicated micro assumptions. In this respect he had much in common with current-day rational-choice sociologists. But unlike them he did not base his analyses on assumptions about rational actors.

Instead he maintained that sociological theories should be based on assumptions derived from behavioural psychology: ‘the principles of behavioral psychology are the general propositions we use, whether implicitly or explicitly, in explaining all social phenomena’ (Homans 1969, p. 204). Despite these differences between Homans’s type of sociology and contemporary rational-choice sociology, Homans’s emphasis on precise and deductive actor-based explanations meant that he paved the way for what later was to become rational-choice sociology (see Coleman 1990a).

Another early work which was important for the emergence of rational-choice sociology was Peter Blau’s *Power and Exchange in Social Life* (1964). The book covers a range of topics, but Blau was particularly interested in what we today would call implicit contract theory, and he focused in particular on the role of reciprocity in explaining the patterns of social interactions that are likely to emerge within a group of individuals. He also was interested in how differences in power and status emerge over time as the result of such exchanges (see also Emerson 1962; Cook and Emerson 1978).

Also of considerable importance was the economist Mancur Olson’s (1965) analysis of the logic of collective action. In the pre-Olson era, most sociological theories of social movements and collective action did not problematize the distinction between individual and collective interests. Using standard microeconomic theory to analyse individuals’ decisions whether or not to join an organization for collective action, Olson showed that one often should expect rational individuals to be free riders even when they would have been better off had they all joined the organization. In the light of Olson’s contributions, social movement researchers started to pay much more attention to the role of individual incentives, and as a consequence, rational-choice ideas came to have a great deal of influence. Hechter’s (1987) influential book on the principles of group solidarity exemplifies this trend.

In European sociology, one of the key contributors to the rational-choice tradition is Raymond Boudon (for example, 1981, 2000, 2003). In numerous publications he argued for the importance of explanations which assume that individuals act rationally. Boudon always has emphasized the importance of basing explanations on realistic theories of action, however. According to Boudon, it is important to recognize the cognitive limitations of real individuals. Individuals often act rationally in the sense of having good reasons for doing what they do, even if these actions may not necessarily be those prescribed by expected utility theory.

Other European sociologists who were important for the development of rational-choice sociology include Lindenberg (for example, 1985, 1990) and Opp (for example, 1986, 1989; see also Raub and Weesie 1990; Abell 1991). Lindenberg (a student of Homans) was one of the founders of the Interuniversity Center for Social Science Theory and Methodology (ICS), a Dutch graduate school built on the foundations of rational-choice theory, and he was also a driving force behind the establishment of rational-choice sections within the International Sociological Association (ISA) and the American Sociological Association (ASA).

Jon Elster is another social scientist who has been of considerable importance for rational-choice sociology. Elster's relation to rational-choice theory always has been somewhat ambivalent, however. On the one hand, he always has considered rational-choice theory to be the best available general theory of action (for example, Elster 1986); on the other hand, most of his writings have been concerned with the limitations of rational-choice explanations. Much of his work since around 1980 has been concerned with the relationship between rationality, social norms, and emotions (for example, Elster 1979, 1983, 1989, 1999). His writings in these areas have been widely read by sociologists and have established important links between sociological theory, the philosophy of action, and behavioural economics.

The single most important person to influence rational-choice sociology has been James Coleman. Coleman did early work on public choice theory (1966) and on the mathematics of collective action (1973), but his *Foundations of Social Theory* (1990b) is by far his most important contribution (see also Coleman 1986, which is an important programmatic statement of his rational-choice position). This treatise of nearly 2,000 pages summarizes and extends much of the work he did during the preceding two decades. In *Foundations*, he shows how a range of traditional sociological concerns such as norms, authority systems, trust, and collective action can be addressed from a rational-choice perspective. In the final third of the book he uses a slightly modified general equilibrium model borrowed from economics to formalize many of the ideas discussed in earlier parts of the book. It is often said that *Foundations* is a book admired by many but read by few, but to judge from Marsden's analyses of citation statistics we may not yet have seen its full impact: 'As of late 2004, more than 1850 indexed works have referenced it, the trend generally increasing over time' (Marsden 2005, p. 18).

Empirical Research

Sociology is an empirically oriented discipline in which the success of a theoretical approach ultimately depends upon its ability to inspire new empirical research and/or to explain important empirical observations. There is a long tradition of implicit use of rationality-like assumption in empirical research, but in some areas, most notably in those concerned with social movements, social mobility, and religion, explicit rational-choice theorizing is closely allied with empirical research, and in these areas rational-choice has become an important part of the intellectual agenda.

As mentioned above, sociological research on social movements was much influenced by the work of Olson (1965), and this is clearly the area of sociology in which rational-choice theories have made the deepest inroads. As a consequence, empirical research has paid a great deal of attention to the costs and benefits of participation when trying to explain the emergence and growth of social movements (see Udehn 1993, for an overview). In sociology, such costs are often understood as being social in the sense that they depend upon the actions of those with whom individuals interact (for example, Opp and Gern 1993; Hedström 1994; Sandell and Stern 1998).

An empirical regularity that has inspired a great deal of sociological research is the persistent influence of class background on educational choice. Boudon (1974) was an early attempt to use rational choice-inspired ideas to understand why this is so. Similarly, the educational choices of Italian youth was studied by Gambetta (1987) seeking to distinguish between the importance of choice-related factors and factors operating behind the back of the individuals.

Goldthorpe, one of the leading social mobility researchers of the last few decades, in an influential article argued for the importance of establishing closer ties between rational-choice theory and the type of statistical analyses that most social-mobility researchers were engaged in (Goldthorpe 1996; see also Goldthorpe 1998, and Blossfeld and Prein 1998). Many others have followed in his path and rational-choice theory is now fairly central to this research community (for example, Breen 1999; Jonsson 1999; Morgan 2002). Breen and Goldthorpe (1997), for example, developed a formal model aimed at explaining the class differential in educational attainment which assumed that families from different classes develop strategies which seek to minimize the risk of downwards mobility. This model has generated a great deal of empirical research (for example, Becker 2003; Davies et al. 2002; Need and de Jong 2001).

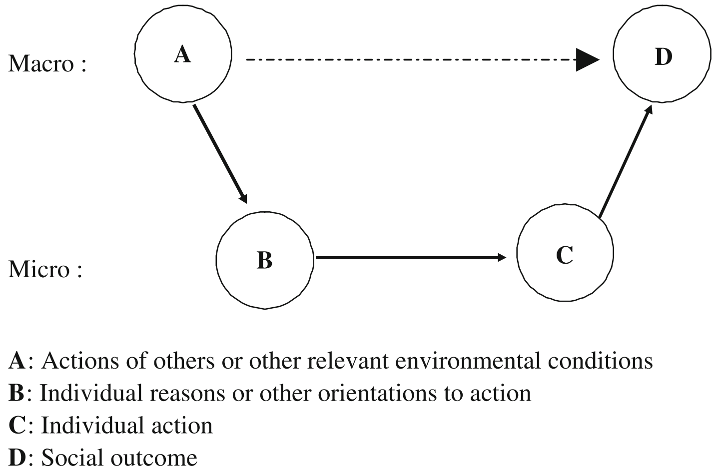
Perhaps somewhat surprisingly, sociology of religion is another area of sociology in which rational-choice theory has had a great deal of influence. For years, it was believed that modern, 'rational' thinking and exposure to alternative religious views would lead people to question the validity of religious belief systems and that religion would lose its foothold (for example, Berger 1969). The situation in Europe was cited as evidence. Rational-choice sociologists, however, pointed to the United States and suggested that pluralism of religious alternatives instead is likely to increase the appeal of religion. They assumed that there exists a market for religious goods which is similar to any other market in that competition can be expected to breed efficiency and entrepreneurial activity which in turn is likely to lead to a more attractive range of religious goods and to higher consumption levels. Rational choice theorists suggested that the European situation with low religious participation was due to state regulation and 'lazy' religious monopolists running the churches, and these ideas have inspired a great deal of empirical research (for example, Iannaccone et al. 1997; Finke and Stark 1992; Stark and Finke 2000).

Economic sociology is another increasingly important sociological area in which rational-choice theory plays a significant, although not dominant, role. The best work in this tradition has a strong empirical grounding and explains social and economic outcomes in terms of actions constrained by the normative, institutional, and structural contexts in which the actors are embedded (see, in particular, Granovetter 1985; Brinton and Nee 1998).

The Standing of Rational-Choice Sociology Within the Discipline

Although many well-known sociologists work within the rational-choice tradition, rational-choice sociology remains controversial. In part this is because rational choice raises important questions about the very identity of sociology as an academic discipline. Classic sociologists such as Pareto (1915–16), Weber (1922), and Parsons (1937) sought to define the core identity of the discipline by contrasting it with economic theory in general, and with the micro-level assumptions of economic theory in particular. From such a perspective rational-choice sociology may appear more like an example of economic imperialism than as 'real' sociology, and as a consequence many contemporary sociologists consider the use of rational-choice assumptions to be a violation of a 'disciplinary taboo' (Baron and Hannan 1994). The title of a recent book edited by Archer and Tritter, the former being an influential social theorist and past president of the International Sociological Association, describes the situation in a nutshell: *Rational Choice Theory: Resisting Colonisation* (Archer and Tritter 2000). These concerns about the discipline being 'colonized' by rational-choice theorists appear unfounded, however. Currently there are only about 200 members in the rational-choice sections of the ASA and the ISA. Although rational-choice sociology has attracted many visible and productive sociologists, these numbers suggest that rational-choice sociology is more of an endangered species than a species likely to invade the discipline at large. This is in sharp contrast to the situation in political science, where the reception of rational choice has been positive and this approach is now widespread especially in the United States.

A recurrent theme in the criticisms advanced against rational-choice sociology concerns the realism of its assumptions. Concerns for realism are also present among many of those close to the rational-choice tradition. As mentioned above, Boudon (for example, 2003) has always emphasized the importance of realistic assumptions about the individual's social situation, incentives, and cognitive abilities. Similarly, Hedström



Rational Choice and Sociology, Fig. 1 Coleman's (1986) micro–macro graph

(2005) has argued that knowingly accepting false assumptions because they lead to better predictions or to more elegant models threatens the explanatory value of the rational-choice approach because it gives incorrect answers to why we observe what we observe. Far from all sociologists are concerned about this, however. Some rational-choice sociologists take a similar position to that of Friedman (1953) and argue that the realism of the assumptions are rather irrelevant (for example, Jasso 1988), and others argue that deviations from rationality can be ignored because they tend to be like random error terms that cancel out in the aggregate (for example, Hernes 1992; Goldthorpe 1998).

Sociological and Economic Versions of Rational-Choice Theory

In an often-cited paper, Duesenberry (1960, p. 233) described the difference between sociology and economics as follows: 'Economics is all about how people make choices. Sociology is all about why they don't have any choices to make.' Although this is an obvious exaggeration of the differences between the disciplines, and particularly the differences between economists and rational-choice sociologists, it captures an important difference between the disciplines. This difference can be described using Coleman's (1986) so-called micro-macro graph (see Fig. 1).

As mentioned above, rational-choice sociologists are macro-oriented but they are methodological individualists in the same sense as economists are, that is, they seek to explain macro outcomes and correlations, such as outcome A or the relationship between A and D in Fig. 1, in terms of the intended and unintended outcomes of individuals' actions. Typically this entails explicating three causal links: (**a**) how individuals' orientations to action – their beliefs, preferences, and so on –are influenced by the social environments in which they are embedded (**A** → **B**); (**b**) how these orientations to action influence how they act (**B** → **C**); and (**c**) how these actions bring about the social outcomes to be explained (**C** → **D**)

As suggested by Duesenberry, sociologists tend to pay more attention to the macro-to-micro link (A → B) than to the latter two links. Sociologists tend to focus on how networks, social norms, socialization processes, and so on influence how individuals act by shaping their preferences, beliefs, opportunities, and so on (for example, Boudon 1988; Burt 1992; Coleman 1990b; Granovetter 1985; Hedström 2005; Raub and Weesie 1990). This choice of focus does not mean that sociologists believe that choices are unimportant, however; it simply is the result of an analytical focus on those aspects of the choice process which are closest to the intellectual heritage of the discipline, and therefore are perceived to be of particular sociological interest.

R

Another important difference between the disciplines concerns the ways in which one typically goes about analysing the type of processes described in Fig. 1. While economic theory is highly mathematized, sociological theory, including sociological rational-choice theory, tends to be much more inductive and empirically oriented. For example, while most economists would specify some mathematical model in order to analyse these types of processes, most rational-choice sociologists rather would take their point of departure in the results of an empirical study. In the sociological analysis, the role of the rational-choice assumption would not be that of an assumption or a postulate of a formal model, but it would be a guide to the type of narrative used for interpreting the empirical results (see Goldthorpe 1996, for a further discussion of this strategy).

These differences between the disciplines mean that rational-choice sociologists often use 'broader' notions of rational choice than economists typically do. As suggested by Camerer and Fehr (2006), the rationality assumption underlying most economic analyses consists of two components: (*a*) individuals are assumed to form, on average, correct beliefs about the world in which they are embedded, and (*b*) individuals are assumed to choose those actions that best satisfy their preferences, given these beliefs. In addition, it is typically assumed that the preferences are self-regarding, exogenously given, and stable through time. Given the sociological interest in how individuals' orientations to action – their beliefs, preferences, and so on – are influenced by the social environments in which they are embedded (the A → B link in Fig. 1), assumptions about stable preferences and non-biased beliefs appear empirically problematic and they would seem to remove from the analysis some of the most interesting and intriguing aspects of the social sciences.

From the economic side of the fence, the more empirically and verbally oriented sociological approach may appear lacking in rigour, while from the sociological side of the fence there is considerable scepticism about analytical results derived from models which, at least in part, are based on assumptions that lack firm behavioural foundations. It seems likely that these disciplinary differences will become less important in the years to come because of converging trends within each discipline. Within economics, there is a growing interest in traditional sociological concerns such as norms, social interactions, and social networks, and experimental approaches are becoming increasingly more important. And within sociology there is a growing recognition of the importance of the type of formal deductive modelling that currently characterizes so much of economic theory.

Concluding Remarks

At this point in time it is difficult to tell whether rational-choice sociology is destined to become an influential force within sociology. It has established itself within the discipline, and more so in Europe than in the United States, but there are no indications, to judge by the size of the rational-choice sections of the ASA and the ISA, that the number of rational-choice sociologists is increasing. Nevertheless, rational-choice theory has had and continues to have an important influence on the discipline, at least in forcing dissenters to clarify better their theoretical tools. One indication of this is that the mainstream sociological vocabulary now includes a range of concepts originating in rational-choice theory, such as free-riders, transaction costs, and collective goods. In addition, largely because of the influence of rational-choice theory, empirically oriented sociologists increasingly acknowledge the need for solid micro theories, and sociologists in general are increasingly concerned with the role of incentives in explaining actions and the collective outcomes that these actions bring about.

See Also

▶ Collective Action
▶ Public Choice
▶ Public Goods
▶ Religion, Economics of

Bibliography

Abell, P. (ed.). 1991. *Rational choice theory*. Aldershot: Edward Elgar.
Archer, M., and J. Tritter (eds.). 2000. *Rational choice theory: Resisting colonisation*. Florence: Routledge.
Baron, J., and M. Hannan. 1994. The impact of economics on contemporary sociology. *Journal of Economic Literature* 32: 1111–1146.
Becker, R. 2003. Educational expansion and persistent inequalities in Germany. *European Sociological Review* 19: 1–24.
Berger, P. 1969. *The sacred canopy: Elements of a sociological theory of religion*. New York: Anchor Books.
Blau, P. 1964. *Exchange and power in social life*. New York: Wiley.
Blossfeld, H.-P., and G. Prein (eds.). 1998. *Rational choice theory and large-scale data analysis*. Boulder: Westview.
Boudon, R. 1974. *Education, opportunity and social inequality*. New York: Wiley.
Boudon, R. 1981. *The logic of social action*. London: Routledge.
Boudon, R. 1988. The logic of relative frustration. In *Rationality and revolution*, ed. M. Taylor. Cambridge: Cambridge University Press.
Boudon, R. 2000. *The origin of values: Sociology and philosophy of beliefs*. New Brunswick: Transaction.
Boudon, R. 2003. Beyond rational choice theory. *Annual Review of Sociology* 29: 1–21.
Breen, R. 1999. Beliefs, rational choice and Bayesian learning. *Rationality and Society* 11: 463–480.
Breen, R., and J. Goldthorpe. 1997. Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society* 9: 275–305.
Brinton, M., and V. Nee (eds.). 1998. *The new institutionalism in sociology*. New York: Russel Sage Foundation.
Burt, R. 1992. *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
Camerer, C., and E. Fehr. 2006. When does ‘economic man’ dominate social behavior? *Science* 311: 7–52.
Coleman, J. 1966. Foundations for a theory of collective decisions. *American Journal of Sociology* 71: 615–627.
Coleman, J. 1973. *The mathematics of collective action*. Chicago: Aldine.
Coleman, J. 1986. Social theory, social research, and a theory of action. *American Journal of Sociology* 91: 309–1335.
Coleman, J. 1990a. Interview with James S. Coleman. In *Economics and sociology*, ed. R. Swedberg. Princeton: Princeton University Press.

We wish to thank Peter Abell, Filippo Barbera, Larry Blume, Richard Breen, Diego Gambetta, John Goldthorpe, Siegwart Lindenberg, Lars Udehn, Peyton Young, and Victor Nee for their useful comments.

Coleman, J. 1990b. *Foundations of social theory*. Cambridge, MA: Belknap Press of Harvard University Press.
Cook, K., and R. Emerson. 1978. Power, equity, and commitment in exchange networks. *American Sociological Review* 43: 721–739.
Davies, R., E. Heinesen, and A. Holm. 2002. The relative risk aversion hypothesis of educational choice. *Journal of Population Economics* 15: 683–714.
Duesenberry, J. 1960. Comment on ‘an economic analysis of fertility’. In *Demographic and economic change in developed Countries*, ed. NBER. Princeton: Princeton University Press.
Elster, J. 1979. *Ulysses and the Sirens: Studies in rationality and irrationality*. Cambridge, UK: Cambridge University Press.
Elster, J. 1983. *Sour grapes: Studies in the subversion of rationality*. Cambridge, UK: Cambridge University Press.
Elster, J. 1986. Introduction. In *Rational choice: Readings in social and political theory*, ed. J. Elster. New York: New York University Press.
Elster, J. 1989. *The cement of society: A study of social order*. Cambridge, UK: Cambridge University Press.
Elster, J. 1999. *Alchemies of the mind: Rationality and the emotions*. Cambridge, UK: Cambridge University Press.
Emerson, R. 1962. Power-dependence relations. *American Sociological Review* 27: 31–41.
Finke, R., and R. Stark. 1992. *The Churching of America, 1776–1990: Winners and losers in our religious economy*. New Brunswick: Rutgers University Press.
Friedman, M. 1953. The methodology of positive economics. In *Essays in positive economics*, ed. M. Friedman. Chicago: Chicago University Press.
Gambetta, D. 1987. *Were they pushed or did they jump? Individual decision mechanisms in education*. Cambridge, UK: Cambridge University Press.
Goldthorpe, J. 1996. The quantitative analysis of large-scale data-sets and rational action theory: For a sociological alliance. *European Sociological Review* 12: 109–126.
Goldthorpe, J. 1998. Rational action theory for sociology. *British Journal of Sociology* 49: 167–192.
Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91: 481–510.
Hechter, M. 1987. *Principles of group solidarity*. Berkeley: University of California Press.
Hedström, P. 1994. Contagious collectivities: On the spatial diffusion of Swedish trade unions, 1890–1940. *American Journal of Sociology* 99: 1157–1179.
Hedström, P. 2005. *Dissecting the social: On the principles of analytical sociology*. Cambridge, UK: Cambridge University Press.
Hernes, G. 1992. We are smarter than we think: A rejoinder to Smelser. *Rationality and Society* 4: 421–436.
Homans, G. 1958. Social behavior as exchange. *American Journal of Sociology* 63: 597–606.

R

Homans, G. 1964. Bringing men back in. *American Sociological Review* 29: 809–818.
Homans, G. 1969. The sociological relevance of behaviorism. In *Certainties and doubts: Collected papers, 1962–1985*. New Brunswick: Transaction Books, 1987.
Iannaccone, L., R. Finke, and R. Stark. 1997. Deregulating religion: The economics of church and state. *Economic Inquiry* 35: 350–364.
Jasso, G. 1988. Principles of theoretical analysis. *Sociological Theory* 6: 1–20.
Jonsson, J. 1999. Explaining sex differences in educational choice: An empirical assessment of a rational choice model. *European Sociological Review* 15: 391–404.
Lindenberg, S. 1985. Rational choice and sociological theory: New pressures on economics as a social science. *Journal of Institutional and Theoretical Economics* 141: 244–255.
Lindenberg, S. 1990. Homo socio-oeconomicus: The emergence of a general model of man in the social sciences. *Journal of Institutional and Theoretical Economics* 146: 727–748.
Marsden, P. 2005. The sociology of James S. Coleman. *Annual Review of Sociology* 31: 1–24.
Morgan, S. 2002. Modelling preparatory commitment and non-repeatable decisions. Information-processing, preference formation and educational attainment. *Rationality and Society* 14: 387–429.
Need, A., and U. de Jong. 2001. Educational differentials in the Netherlands: Testing rational action theory. *Rationality and Society* 13: 71–98.
Olson, M. 1965. *The logic of collective action*. Cambridge, MA: Harvard University Press.
Opp, K.-D. 1986. Soft incentives and collective action: Participation in the anti–nuclear movement. *British Journal of Political Science* 16: 87–112.
Opp, K.-D. 1989. *The rationality of political protest: A comparative analysis of rational choice theory.* Boulder: Westview press.
Opp, K.-D., and C. Gern. 1993. Dissident groups, personal networks, and spontaneous cooperation: The East German revolution of 1989. *American Sociological Review* 58: 659–680.
Pareto, V. 1915–1916. *Mind and society: A treatise on general sociology.* New York: Dover Publications, 1935.
Parsons, T. 1937. *The structure of social action*. New York: Free Press.
Raub, W., and J. Weesie. 1990. Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology* 96: 626–654.
Sandell, R., and C. Stern. 1998. Group size and the logic of collective action: A network analysis of a Swedish temperance movement 1896–1937. *Rationality and Society* 10: 327–345.
Simon, H. 1987. Behavioral economics. In *The new palgrave: A dictionary of economics*, vol. 1, ed. J. Eatwell, M. Milgate, and P. Newman. London: Macmillan.
Smelser, N. 1962. *Theory of collective behavior.* New York: Free Press.
Stark, R., and R. Finke. 2000. *Acts of faith. Explaining the human side of religion*. Berkeley: University of California Press.
Swedberg, R. 1990. *Economics and sociology.* Princeton: Princeton University Press.
Udehn, L. 1993. 25 years with the logic of collective action. *Acta Sociologica* 36: 239–261.
Weber, M. 1922. *Economy and society.* Berkeley: University of California Press, 1978.

Rational Expectations

Thomas J. Sargent

Abstract

Rational expectations is an equilibrium concept that attributes a common model (a joint probability distribution over exogenous variables and outcomes) to nature and to all agents in the model. The rational expectations equilibrium concept makes parameters describing agents' belief disappear as components of a model, giving rise to the cross-equation restrictions that offer rational expectations models their empirical power.

Keywords

Adaptive expectations; Cobweb theorem; Computatuion of equolibria; Consumer surplus; Corn-hog cycles; Curse of dimensionality; Distributed lags; Euler equations; Forecasting; Generalized method of moments; Identification; Lagrange multipliers; Linear prediction theory; Markov chain Monte Carlo methods; Markov processes; Maximum likelihood; Muth, J. F.; Ommitment; Ordinary differential equations; Producer surplus; Rational expectations; Rational expectations equilibrium; Real business cycles; Reputation; Self-confirming equilibria; Shadow pricing; Stochastic approximation; Stochastic optimal growth models; Time consistency

JEL Classifications

B4

'Rational expectations' is an equilibrium concept that can be applied to dynamic economic models that have elements of 'self-reference', that is, models in which the endogenous variables are influenced by the expectations about future values of those variables held by the agents in the model. The concept was introduced and applied by John F. Muth (1960, 1961) in two articles that interpreted econometric distributed lag models. Muth used explicitly stochastic dynamic models and brought to bear his extensive knowledge of classical linear prediction theory to interpret distributed lags in terms of economic parameters. For Muth, an econometric model with rational expectations possesses the defining property that the forecasts made by agents within the model are no worse than the forecasts that can be made by the economist who has the model.

Muth's first concrete application of rational expectations was to find restrictions on a stochastic process for income that would render Milton Friedman's (1957) geometric distributed lag formula for permanent income an optimal predictor for income. Muth showed that, if the first difference of income is a first-order moving average process, then Friedman's formula is optimal for forecasting income over any horizon. The independence of this formula from the horizon makes precise the sense in which Friedman's formula extracts from past income an estimator of 'permanent' income. In working backwards from Friedman's formula to a process for income in this way, Muth touched Lucas's critique (1976). Given any distributed lag for forecasting income, one can work backwards as Muth did and discover a stochastic process for income that makes that distributed lag an optimal predictor for income over some horizon. Similarly, Sargent (1977) reverse engineered a joint inflation- money creation process that makes Cagan's (1956) adaptive expectations scheme for forecasting inflation a linear least squares forecast.

Solving a few such inverse-optimal prediction problems in the fashion of Muth and Sargent quickly reveals the dependence of a distributed lag for forecasting the future on the form of the stochastic process that is being forecast. In 1963, Peter Whittle published a book that conveniently summarized and made more accessible to economists the classical linear prediction theory that Muth had used. That book repeatedly applies the Wiener–Kolmogorov formula for the optimal j-step ahead predictor of a covariance stationary stochastic process x_t with moving average representation $x_t = c(L)\ \varepsilon_t$. The Wiener–Kolmogorov formula displays the dependence of the optimal distributed lag for predicting future x on the form of $c(L)$. That dependence underlies Lucas's critique of econometric policy evaluation procedures that were common when Lucas composed his critique in 1973. Those procedures had assumed that distributed lags in behavioural relations would remain invariant with respect to alterations in government policy rules, alterations that took the form of changes in $c(L)$ for government policy instruments. Although the formulas in Whittle's book were used extensively by Nerlove (1967) to work out additional examples along the lines of Muth, it was not until the writing of Lucas's critique in 1973 and its publication in 1976 that the implications for econometric practice of Muth's ideas and the prediction formulas in Whittle began to be widely appreciated.

Lucas and Prescott (1971) clarified and extended rational expectations as an equilibrium concept and also pointed the way to connecting theory with observations. They described the partial equilibrium of an industry in which there exists a fixed number of identical firms, each subject to costs of adjustment for a single factor of production, capital. The industry faces a downward sloping demand curve for its output that shifts randomly due to a demand shock that follows a Markov process. The representative firm maximizes the expected present value of its profits by choosing a contingency plan for investment. To state the firm's optimum problem, it is necessary to describe what the firm believes about the motion of variables that influence its future returns even though they are beyond the firm's control. The price of output is such an uncontrollable variable, but the demand curve for output and the hypothesis of market clearing make price a function of the capital stock in the industry as a whole. It follows that to state the firm's decision problem requires the firm's view about the law of motion of the industry-wide capital stock be stated. The representative firm's optimum problem can then be solved, yielding a law of motion for the capital stock of the representative firm in which both the individual firm's capital stock and the market-wide capital stock are both state variables. Multiplying this law of motion by the number of firms then gives the actual law of motion for capital in the industry. In this way, the firm's optimization problem and the hypothesis of market clearing induce a mapping from a perceived law of motion to an actual law of motion for the industry's capital stock. A rational expectations equilibrium is a fixed point of this mapping. By studying an artificial planning problem that maximizes consumer plus producer surplus, Lucas and Prescott pursued an indirect approach to describing conditions under which a unique fixed point exists. In this way, they formulated a recursive competitive equilibrium.

From a practical perspective, an important property of a rational expectations model is that it imposes a communism of models and expectations. If we define a model as a probability distribution over a sequence of outcomes, possibly indexed by a parameter vector, a rational expectations equilibrium asserts that the same model is shared by (1) all of the agents within the model, (2) the econometrician estimating the model, and (3) nature, also known as the data generating mechanism. Different agents might have different information, but they form forecasts by computing conditional expectations with respect to a common joint density, that is, a common model. Communism of models gives rational expectations much of its empirical power and underlies the cross-equation restrictions that are used by rational expectations econometrics to identify and estimate parameters. A related perspective is that, within models that have unique rational expectations equilibria, the hypothesis of rational expectations makes agents' expectations disappear as objects to be specified by the model-builder or to be estimated by the econometrician. Instead, they are equilibrium outcomes.

The equilibrium law of motion for capital induces a stochastic process for capital that assumes the form of a Markov process. Lucas and Prescott showed that this Markov process converges in distribution to a unique invariant distribution. That justifies an asymptotic distribution theory adequate for doing time series econometrics, in particular, a mean ergodic theorem that guarantees that sample moments converge to the corresponding population moments. Lucas and Prescott's notion of a recursive competitive equilibrium thus takes a big step towards integrating dynamic theory and econometrics because it supplies an explicit mapping from economic parameters describing preferences, technology, and information sets to the population moments of observable sequences of economic time series. The task of econometrics under rational expectations is to 'invert' this mapping by using time series data to make inferences about economic parameters.

Hansen and Sargent (1980) used linear versions of Lucas–Prescott and Brock and Mirman (1972) models as laboratories for working out econometric techniques for estimating rational expectations models. They studied both generalized method of moments (GMM) and maximum likelihood approaches. They described how desirable statistical properties including consistency and asymptotic efficiency for estimators of the model's economic parameters induce a metric for measuring distance between the sample moments and the theoretical population moments implied by the equilibrium of the model at given parameter values. Typical metrics are those associated with the generalized method of moments, a special case of which is associated with the first-order conditions for maximizing a Gaussian likelihood function. Parameter estimates are obtained by minimizing the metric with respect to the parameter values, a nonlinear minimization problem.

Econometric identification of parameters means uniqueness of the minimizer of distance between the theory and the observations. Identification is partially achieved by the rich set of cross-equation restrictions that the hypothesis of rational expectations imposes (the same parameters appear in many equations, in highly nonlinear ways). These cross-equation restrictions achieve identification in a different manner from the

Cowles Commission's 'rank and order' conditions, which explicitly excluded cross-equation restrictions. Dynamic rational expectations models subvert such 'exclusion restrictions', and thereby destroy the neat division between 'supply' and 'demand' curves that underlay the 'exclusion' approach to identification.

Minimum distance estimation of a rational expectations model requires recomputing an equilibrium for each set of parameter values used during a descent with respect to the data-fitting metric. Except for linear models, Bellman's 'curse of dimensionality' makes it challenging to compute an equilibrium, so developing improved computational methods has become an important research area. Judd (1998) describes a variety of numerical approaches. Methods for computing equilibria are required not only for parameter estimation, but also for quantitatively evaluating the effects of proposed interventions, for example, new policies for setting government instruments. A new government policy implies, via the cross-equation restrictions, new laws of motion for all the endogenous variables in the models. It is no coincidence that full information estimation methods require calculations closely connected to those needed to evaluate policy.

Good computer programmes for solving and estimating complete rational expectations models have recently become available. A suite of Matlab programmes called Dynare was written by Michele Juilliard and colleagues and is available on the Internet. Dynare solves linear models as systems of expectational difference equations using methods originally described by Sargent (1979), Blanchard and Khan (1980), and Whiteman (1983). Dynare estimates models by either maximum likelihood or a Markov chain Monte Carlo procedure to construct a Bayesian posterior density over free parameters. Dynare also knows how to compute and estimate various linear and log-linear approximations to nonlinear models.

Hansen and Singleton (1982) suggested a short-cut estimation method capable of estimating the parameters of a subset of preference and technologies without computing or estimating a complete equilibrium. Their idea was to use back out parameter estimates from conditional moment restrictions implied by the first-order necessary conditions (Euler equations) for an agent's dynamic optimization problem. Hansen and Singleton pointed out that their GMM method requires special restrictions on the stochastic process of disturbances to the function being estimated, and that it typically fails to estimate enough parameters to permit evaluating many kinds of interventions. Nevertheless, its ease of use and presumed robustness to features of the environment that a researcher prefers not to specify have made it a very popular and fruitful approach.

As already mentioned, a rational expectations equilibrium is a fixed point from a perceived to an actual law of motion. It is tempting to hope that iterations on that mapping converge to a fixed point. But that is asking for too much because the mapping is not a contraction and it is easy to construct examples in which iterations diverge. Nevertheless, the mapping from a perceived to an actual law of motion plays an important role in studying how a rational expectations equilibrium can emerge as the limit point of a system of adaptive agents who use least squares on historical data to forecast the future, rather than the population moments from the equilibrium that are handed to them within a rational expectations equilibrium. By applying the theory of stochastic approximation, Marcet and Sargent (1989) and Woodford (1990) derived an ordinary differential equation (ODE) for beliefs that describe the limiting behaviour of such an adaptive system. That ODE expresses how the gap between the perceived and implied actual law of motion governs a limiting rate of change of beliefs. Necessary and sufficient conditions for convergence to a rational expectations equilibrium are stated in terms of the stability of the associated ode. These conditions have been dubbed the E-stability conditions by Evans and Honkapohja (2001) and are useful for constructing algorithms for computing rational expectations equilibria via least squares learning algorithms or direct attacks on the ordinary differential equation governing E-stability. This is in effect what Krusell and Smith (1998) do, though they do not connect their method to the learning literature.

The literature on least squares learning and adaptive learning in games (for example, Marcet and Sargent 1989; Woodford 1990; Fudenberg and Levine 1998) began partly as a response to a widespread scepticism about the plausibility of the communism of expectations imposed by rational expectations. How could people possibly come to learn to share a common model with each other, the econometrician, and nature? The learning literature offers an explanation. But the learning literature falls short of implying a communism of models as extensive as the one typically imposed in macroeconomics. A metatheorem is that, if a system of least squares agents converges, it converges to a self-confirming equilibrium (see Fudenberg and Levine 1998; Sargent 1999). In a self-confirming equilibrium, agents' models agree about events that occur frequently enough (infinitely often) within the equilibrium. But agents can have different subjective distributions about events that occur infrequently because they are off the equilibrium path. For those events, a law of numbers just doesn't have enough observations to work on. In a macro model, it is typically irrelevant that private agents' beliefs can be wrong off an equilibrium path because, being atomistic, all that matters for them are their conditional forecasts along an equilibrium path. But for the government, its beliefs about off-equilibrium paths events influence its choices in important ways: designing government policy is all about evaluating the effects of alternative hypothetical outcome paths, most of which will not be observed. Kreps (1998) defends the concept of self-confirming equilibrium.

Lucas and Prescott's model can be used to study aspects of the theory of policy. Their model generates a stochastic process for output, price and industry capital that exhibits recurrent but aperiodic 'cycles', as realizations of stochastic difference equations do. Thus, Lucas and Prescott's model is an alternative to the 'cobweb' mechanism for generating fluctuations in commodity markets. Two-industry versions of the model can readily be constructed to model 'corn-hog' cycles. Models along the lines of Lucas and Prescott's reveal a different perspective on these cycles than do cobweb models. Lucas and Prescott show that, despite cyclical fluctuations, the equilibrium of their model is optimal in the sense that it maximizes the expected present value of consumer surplus net of producer surplus. Therefore, unlike cobweb models, in which cycles partly reflect erroneous and readily improved upon perceptions of private agents, matters cannot be improved by government interventions designed to smooth out the cycles. Models of this kind have been calibrated to price and quantity data from markets for cattle, housing, and engineers by Rosen et al. (1994), Rosen and Topel (1988), and Ryoo and Rosen (2004).

For studying a variety of macroeconomic questions, researchers have used what can be interpreted as a version of Lucas and Prescott's model, suitably modified and reinterpreted to apply to an aggregative economy. Brock and Mirman (1972) analysed a centralized version of such an economy that took the form of a stochastic version of a one-sector optimal growth model. The planner in their model seeks to maximize the expected discounted value of utility of consumption subject to a technology for transforming consumption over time via investment in physical capital. Brock and Mirman gave conditions under which the optimal plan for capital and consumption induces a stochastic process that converges in distribution, so that, like Lucas and Prescott's model, theirs is prepared for rigorous treatment econometrically. It is possible to decentralize Brock and Mirman's model into an equivalent economy consisting of competitive firms and households who interact in markets for labour and capital and who have rational expectations about the evolution of the wages and interest rates that they face. Decentralized versions of Brock–Mirman models have been used to construct equilibrium theories of stock prices and interest rates, typically by computing particular shadow prices associated with the planning problem (Lucas 1978; Brock 1982). Decentralized versions of the Brock–Mirman model form the backbone of the modern version of 'real business cycle theory' that was initiated by Kydland and Prescott (1982). Since the stochastic optimal growth model has a stochastic difference equation for capital as its equilibrium, it shares with the

Lucas–Prescott model the property that it readily generates realizations for capital, output and consumption that display recurrent but aperiodic fluctuations of the kind observed in aggregate time series data. Kydland and Prescott embarked on the task of taking seriously the possibility that the preferences and technology of a small stochastic optimal growth model could be specified so that it would approximate closely the moments of a list of important aggregate economic time series for the United States. Kydland and Prescott have constructed several such models, each driven by a single unobserved shock, which they interpret as a disturbance to technology. This research strategy is charged with meaning, since it undertakes to explain aggregate time series data with a model whose equilibrium is optimal, and in which there is no government. The government is neither a contributing source to economic fluctuations nor a potential modifier of those fluctuations. Real business cycle models of this kind are capable of determining a long list of real variables, while remaining silent about all nominal variables.

But central banks are supposed to determine nominal variables, which has created an interest in adapting real business cycle models to include interactions among nominal and real variables. By directly imposing parameterized versions of wage and price inertia, Smets and Wouter (2003) and Woodford (2003) have formulated rational expectations models with enough shocks and rigidities to fit macro data well enough to be useful to research departments of leading central banks. These models can be estimated and simulated with Dynare.

The idea of rational expectations was essential for formulating the problem of time inconsistency in macroeconomics. Three ideas underlie the time consistency problem in multi-agent dynamic games and macroeconomic models: (1) the communism of models brought by rational expectations, (2) backward induction by all agents, and (3) the observation that different timing protocols generally imply different outcomes. The time inconsistency 'problem' was recognized in macroeconomics by Kydland and Prescott (1977) and Calvo (1978), who studied macro models in which a competitive economy with a representative agent confronts a benevolent government. These papers compare outcomes under two timing protocols. In one timing protocol, private agents choose sequentially but the government has a commitment technology that allows it once and for all at time zero to choose an entire history contingent sequence of actions (for example, tax rates or money supplies). In the other, the government, or a sequence of government administrations if you prefer, must choose sequentially, that is, anew each period. Outcomes under these two timing protocols typically differ, with outcomes being better under the timing protocol that allows the government to choose once and for all at time 0. The difference in outcomes shows the value of being able to commit at time 0. In the problem under commitment, among the constraints that the government faces at time 0 are a sequence of private agents' Euler equations that involve their (rational) expectations of future government actions. The equilibrium time t values of the Lagrange multipliers on these 'implementability constraints' encode the costs in terms of the government's time t continuation value of confirming the time t expectations that the government's time 0 plan had induced private agents to expect. The presence of those implementability conditions in the government's constraint set gives rise to a conflict between the preference orderings of the government and the representative agent over outcomes. That conflict is the ultimate source of the timing inconsistency problem. Recursive methods for computing the optimal plan under commitment were first suggested by Kydland and Prescott (1980) and are surveyed in Ljungqivst and Sargent (2004). These methods are used extensively in the literature on rational expectations monetary models with ad hoc inertial wages and prices that Woodford (2003) catalogues and extends.

An important literature studies whether reputation can overcome the time inconsistency problem. The finding of this literature is that, by allowing history- dependent strategies, reputation can substitute for the ability to commit if the discount factor is sufficiently close to one. This literature, which is surveyed critically in Ljungqivst and Sargent (2004), exploits the communism of expectations inherent in rational expectations.

See Also

- ▶ Business Cycle Measurement
- ▶ Inflation Expectations
- ▶ New Classical Macroeconomics
- ▶ Self-confirming Equilibria

Bibliography

Blanchard, O.J., and C.M. Kahn. 1980. The solution of linear difference models under rational expectations. *Econometrica* 48: 1305–1311.

Brock, W.A. 1982. Asset prices in a production economy. In *Economics of information and uncertainty*, ed. J.J. McCall. Chicago: University of Chicago Press.

Brock, W.A., and L. Mirman. 1972. Optimal economic growth under uncertainty: The discounted case. *Journal of Economic Theory* 4: 479–513.

Cagan, P. 1956. The monetary dynamics of hyperinflation. In *Studies in the quantity theory of money*, ed. M. Friedman. Chicago: University of Chicago Press.

Calvo, G.A. 1978. On the time consistency of optimal policy in a monetary economy. *Econometrica* 46: 1411–1428.

Evans, G.W., and S. Honkapohja. 2001. *Learning and expectations in macroeconomics*. Princeton: Princeton University Press.

Friedman, M. 1957. *A theory of the consumption function*. Princeton: Princeton University Press.

Fudenberg, D., and D.K. Levine. 1998. *The theory of learning in games*. Cambridge, MA: MIT Press.

Hansen, L.P., and T.J. Sargent. 1980. Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2: 7–46.

Hansen, L.P., and K.J. Singleton. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50: 1269–1286.

Judd, K. 1998. *Numerical methods in economics*. Cambridge, MA: MIT Press.

Kreps, D.M. 1998. Anticipated utility and dynamic choice. In *Frontiers of research in economic theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, ed. D.P. Jacobs, E. Kalai, and M.I. Kamien. Cambridge: Cambridge University Press.

Krusell, P., and A. Smith. 1998. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 106: 867–896.

Kydland, F.E., and E.C. Prescott. 1977. Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy* 85: 473–491.

Kydland, F.E., and E.C. Prescott. 1980. Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control* 2: 79–91.

Kydland, F.E., and E.C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50: 1345–1370.

Ljungqvist, L., and T.J. Sargent. 2004. *Recursive macroeconomic theory*, 2nd ed. Cambridge, MA: MIT Press.

Lucas Jr., R.E. 1976. Econometric policy evaluation: A critique. In *The Phillips curve and the labor market*, Vol. 1 of Carnegie-Rochester conference in public policy, ed. K. Brunner and A. Meltzer. Amsterdam: North-Holland.

Lucas Jr., R.E. 1978. Asset prices in an exchange economy. *Econometrica* 46: 1429–1445.

Lucas Jr., R.E., and E.C. Prescott. 1971. Investment under uncertainty. *Econometrica* 39: 659–681.

Marcet, A., and T.J. Sargent. 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory* 48: 337–368.

Muth, J.F. 1960. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association* 55: 299–306.

Muth, J.F. 1961. Rational expectations and the theory of the price movements. *Econometrica* 29: 315–335.

Nerlove, M. 1967. Distributed lags and unobserved components in economic time series. In *Ten economic studies in the tradition of Irving Fisher*, ed. W. Fellner. New York: Wiley.

Rosen, S., and R.H. Topel. 1988. Housing investment in the United States. *Journal of Political Economy* 96: 718–740.

Rosen, S., K.M. Murphy, and J.A. Scheinkman. 1994. Cattle cycles. *Journal of Political Economy* 102: 468–492.

Ryoo, J., and S. Rosen. 2004. The engineering labor market. *Journal of Political Economy* 112: S110–S139.

Sargent, T.J. 1977. The demand for money during hyperinflations under rational expectations: I. *International Economic Review* 18: 59–82.

Sargent, T.J. 1979. *Macroeconomic theory*. New York: Academic Press.

Sargent, T.J. 1984. Autoregressions, expectations, and advice. *American Economic Review* 74: 408–415.

Sargent, T.J. 1999. *The conquest of American inflation*. Princeton: Princeton University Press.

Sargent, T.J., and N. Wallace. 1976. Rational expectations and the theory of economic policy. *Journal of Monetary Economics* 2: 169–183.

Smets, F., and R. Wouter. 2003. An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association* 1: 1123–1175.

Whiteman, C.H. 1983. *Linear rational expectations models: A users guide*. Minneapolis: University of Minnesota Press.

Whittle, P. 1983. *Prediction and regulation by linear least-square methods*, 2nd ed. Minneapolis: University of Minnesota Press.

Woodford, M. 1990. Learning to believe in sunspots. *Econometrica* 58: 277–307.

Woodford, M. 2003. *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.

Rational Expectations Models, Estimation of

Monika Piazzesi

Abstract

Rational expectations impose cross-equation restrictions that have important implications for the estimation of models. These implications have lead to the development of new estimation and testing techniques. More recently, this development has generated techniques that handle models that cannot be solved analytically. Together with the rapid increase in computing power, these methods offer insights in to the working of these models and thereby enable their refinement.

Keywords

Bayesian methods on macroeconometrics; Cross-equation restrictions; Distributed lags; Estimation; Euler equations; Generalized method of moments; Maximum likelihood; Rational expectations; Simulation-based estimation; Term structure of interest rates; Testing; Vector autoregressions

JEL Classifications

D4; D10

Most dynamic models in economics assume that agents form expectations rationally. An equilibrium of a dynamic model can typically be described by a probability distribution over sequences of data. The rational expectations assumption says that every agent's subjective belief about the data is a conditional of this equilibrium probability distribution, where the conditioning is on the agent's information set. Expectations are thus consistent with outcomes generated by the model. They are also optimal, in the sense that they correctly use all information available to the agent.

The rational expectations assumption was first proposed by John F. Muth in the early 1960s in his analysis of linear macroeconomic models. Prior to Muth's work, expectations in those models had been parametrized distributed lags. In the early 1970s, Robert E. Lucas Jr. studied the rational expectations equilibrium of a model with optimizing agents who have different information sets. It was recognized early on that taking rational expectations models to data required new techniques. Building on the early work on tests of the natural rate hypothesis by Sargent (1971), there has been much progress in rational expectations econometrics since the mid-1970s (for example, see Hansen and Sargent 1980, 1991; Lucas and Sargent 1981). In the meantime, the rational expectations assumption has come to be used in many fields of economics, including finance, labour economics and industrial organization.

Rational expectations impose cross-equation restrictions that have important implications for the estimation of models, which I will describe below. These implications have lead to the development of new estimation and testing techniques. More recently, this development has generated techniques that handle models that cannot be solved analytically. Together with the rapid increase in computing power, these methods offer insights in to the working of these models and thereby enable their refinement.

Cross-Equation Restrictions

The rational expectations assumption implies cross-equation restrictions that constrain parameters and shocks different places of the model. There are (at least) three reasons for why these restrictions have important implications for estimation. First, cross-equation restrictions constrain the parameters associated with agents' expectations to be consistent with the parameters from the equilibrium probability distribution. These restrictions reduce the overall number of parameters that have to be estimated. In particular, they eliminate any free parameters associated with expectations. To see why, consider a dynamic model with an agent who maximizes some objective function subject to constraints. To solve this optimization program, the agent needs to form expectations about future variables such as growth rates. In a model without rational expectations, these expectations might be based on some subjective belief about the future. This belief introduces free parameters that need to be estimated in addition to other model parameters, such as preference parameters.

Take, for example, an endowment economy populated by a representative agent with time separable power utility. The agent may be optimistic and believe in high mean growth rates for the endowment. This optimistic belief will have an affect on equilibrium outcomes. For example, the agent's Euler equation will only hold for a high short real rate, because the high mean growth rate implies a strong consumption smoothing motive. However, the actual mean growth rate in this economy may be lower than what the agent believes (so that the agent will be disappointed by the endowment realizations.)

The estimation of the model with an optimistic agent involves two parameters, the subjective mean of endowment growth and its true mean, which is the mean of the data generating process of endowment growth. The assumption of rational expectations reduces the number of parameters to estimate, because the two mean parameters collapse: the agent's subjective belief is equal to the true data-generating process. In this simple example, the cross-equation restrictions only eliminated one parameter. In more realistic examples, the agent's subjective belief may involve many parameters (for example, because it is described by a vector autoregression in many variables and with many lags), so that the restrictions are important for keeping the estimation tractable.

The second important implication of cross-equation restrictions is that the processes for different endogenous variables often involve the same parameters and shocks. As a consequence, different data series are informative about the same set of parameters. This implication can be used to increase the efficiency of the estimation. Going back to the example of a representative agent endowment economy, the equation describing the equilibrium process of an interest rate on a bond with m-period maturity is intimately related to the equation describing the process of an n-period interest rate for some $m \neq n$. The relationship between different interest-rate equations, or restrictions across equations, consists of parameters that enter both equations (for example, expected growth) and also shock processes that affect both equations (for example, surprises in growth). These restrictions help in the estimation and can be tested empirically with data on interest rates with different maturities.

Some of the earliest tests of cross-equation restrictions were indeed tests of the implications of rational expectations for the term structure of interest rates. Sargent (1979) specifies a vector autoregression (VAR) for short and long rates. Assuming Gaussian disturbances, Sargent estimates this VAR using maximum likelihood and performs likelihood ratio tests to see whether the restrictions imposed by the expectations hypothesis are satisfied. Subsequently, these tests were further refined, and the expectations hypothesis (which is a stronger assumption than rational expectations) was rejected in many empirical studies. The lessons from these statistical rejections have resulted in refined models with rational expectations but time-varying risk premia (for example, Ang and Piazzesi 2003).

The third important implication of rational expectations is that the data-generating process that underlies agent beliefs is equal to the true data-generating process. This enables the estimation of rational expectations models using the generalized method of moments based on moment conditions derived from Euler equations (see Hansen 1982; Hansen and Singleton 1982; generalized method of moments estimation). Using the law of iterated expectations, such a GMM estimation also allows for the case that agents in the model have more information than the econometrician.

Estimation Methods

Estimation methods for rational expectations models can be distinguished by the amount of information they require. Generally speaking, there are full information methods and limited information methods. The goal of full information methods is to estimate the entire model by exploiting all its cross-equation restrictions. This estimation method is efficient and produces estimates for all the parameters in the model. These methods are maximum likelihood and its Bayesian counterparts (see Bayesian methods in macroeconometrics). To apply these methods, the econometrician needs to specify the entire structure of the model, including the distribution of shocks.

Limited information methods require less structure. The goal of these methods is to exploit only some of the restrictions imposed by the model and to obtain estimates for only some of the model parameters. These methods lose some of the efficiency of the full information methods, but they help the researcher to avoid contaminating the estimation results by model misspecification in parts of the model that are not of interest. For example, Hall (1978) and Hansen and Singleton (1982) use the Euler equations from a single agent model as moment conditions for GMM and measure the empirical counterparts of these moments using data on consumption and financial returns. This procedure gives estimates for preference parameters and does not depend on any specific assumption on the distribution of shocks in the model.

Faced with the difficulty that many models do not have analytical solutions and have to be solved numerically, there has been progress regarding simulation-based estimation methods. These methods compare moments of data simulated from the model using some parameter values with their empirical counterparts. For a textbook treatment of these methods, see Gourieroux and Monfort (1996), Gourieroux and Jasiak (2001), and Singleton (2006).

See Also

- ► Bayesian Methods in Macroeconometrics
- ► Generalized Method of Moments Estimation
- ► Rational Expectations

Bibliography

Ang, A., and M. Piazzesi. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* 50: 745–787.

Gourieroux, C., and J. Jasiak. 2001. *Financial econometrics*. Princeton: Princeton University Press.

Gourieroux, C., and A. Monfort. 1996. *Simulation-based econometric methods*. Oxford: Oxford University Press.

Hall, R.E. 1978. Stochastic implications of the life cycle permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86: 971–987.

Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

Hansen, L.P., and T.J. Sargent. 1980. Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2: 7–46.

Hansen, L.P., and T.J. Sargent. 1991. *Rational expectations econometrics*. Boulder: Westview Press.

Hansen, L.P., and K. Singleton. 1982. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91: 249–265.

Lucas, R.E., and T.J. Sargent. 1981. *Rational expectations and econometric practice*. Minneapolis: University of Minnesota Press.

Sargent, T.J. 1971. A note on the accelerationist controversy. *Journal of Money, Credit, and Banking* 3: 721–725.

Sargent, T.J. 1979. A note on maximum likelihood estimation of the rational expectations model of the term structure. *Journal of Monetary Economics* 5: 133–143.

Singleton, K. 2006. *Empirical dynamic asset pricing*. Princeton: Princeton University Press.

R

Rational Expectations: Econometric Implications

N. E. Savin

It has long been recognized that forecasts affect outcomes. Similarly, outcomes affect expectations. Thus, there is a mapping from expectations to outcomes and back to expectations and so from expectations to expectations. A rational expectations equilibrium is a fixed point of this mapping in which expectations generate outcomes which confirm the original expectations. A rational expectations equilibrium is a natural solution concept in a model with expectations. The heuristic reasoning is that outside rational expectations equilibria agents make systematic mistakes; expectations are not confirmed by outcomes in that the expectations are not correct on the average. Consequently, it is very plausible that outside rational expectations equilibria agents will eventually notice that they are making systematic mistakes and attempt to revise the way they forecast in order to eliminate the sources of the systematic errors. This suggests that agents are not in equilibrium until they have learned to form rational expectations.

Econometric analysis typically assumes that the econometrician is an outside observer: nothing which the econometrician does affects the data generation process. In particular, it is assumed that the forecasts based on the econometrician's estimated model do not influence the forecasts of the agents in the economy. By contrast, the agents in the economy are inside econometricians. If the agents' forecasts are derived from an estimated econometric model, then the data generation process changes when the agents update the parameter estimates or change the model specification. A single atomistic agent can act like the outside econometrician, but this is not so for agents as a whole. The collective impact of the forecasting activity of the agents is to change the data generation process; this is the essence of forecast feedback.

This entry concentrates on three topics which involve the econometric implications of rational expectations: solutions, estimation and learning. The issues surrounding the solutions are discussed in section "Solutions" in the context of a second order linear expectational difference equation. Section "Estimation" considers maximum likelihood and general method of moment estimators which can be used by an outside econometrician to estimate the parameters of a rational expectations model. The question of whether agents – inside econometricians – can learn to form rational expectations is addressed in section "Learning". The concluding comments are in section "Concluding Comments".

Solutions

A prototype for many rational expectations models is the second order expectational difference equation

$$E_t y_{t+1} - (\rho_1 + \rho_2) y_t + \rho_1 \rho_2 y_{t-1} = x_t \qquad (1)$$

where t indexes the integers, $\{x_t\}$ and $\{y_t\}$ are scalar stochastic processes and where for expositional purposes ρ_1 and ρ_2 are assumed to be real numbers. The variable x is called the driving process' and '$E_t y_{t+1}$' is the forecast of y based on the information available at time t. The reduced form of the model is the solution of the equation which expresses y as a function of current and past values of x_t, which is the information available at time t. Second order expectational difference equations arise as necessary conditions for optima in linear-quadratic versions of costly adjustment models and in this context are called Euler equations. Examples can be found in Kennan (1979), Sargent (1979), Hansen and Sargent (1980), Eichenbaum (1983) and Hansen and Singleton (1982).

There is a long list of methods for finding solutions to linear expectational difference equations. These include 'state-space' techniques in Lucas (1972), 'methods of undetermined coefficients' in Muth (1961) and Aoki and Canzoneri (1979), 'forward and backward' solutions in Blanchard (1979) and Blanchard and Kahn (1980) and a 'method of undetermined coefficients in the frequency domain' in Saracoglu and Sargent (1978), Futia (1981) and Whiteman (1983).

The solutions presented below are those obtained by the approach of Whiteman (1983). This method is analytically straightforward and has the virtue that it finds all the solutions within a certain set. Whiteman assumes that the driving process is covariance stationary and looks for solutions with the same general structure as the driving process, that is, for solutions in the same 'space' as the driving process. The motivation for this approach is twofold. The first is that without any restrictions on the $\{x_t\}$ and without any side conditions there is a plethora of solutions. The second is that stationarity is assumed in the estimation theory for expectational difference equations.

The Whiteman solution technique employs four assumptions. First, x_t has a known Wold decomposition

$$x_t = \sum_{j=0}^{\infty} A_j \in_{t-j} \tag{2}$$

with
$\in_t = x_t - E(x_t|x_{t-1}, x_{t-2}, \ldots), \sum_{j=0}^{\infty} A_j^2 < \infty$ and the function $A(z) = \Sigma_{j=0}^{\infty} A_j z^j$ must be analytic on the open unit disk. Thus (2) can be written as

$$x_t = \sum_{j=0}^{\infty} A_j L^j \in_t = A(L) \in_t \tag{2$'$}$$

where L is the lag operator: $L^n = x_{t\,-\,n}$ Second, the solutions are in the space of the driving process (2) and are of the form

$$y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j} = C(L) \in_t. \tag{3}$$

Third, the forecasting procedure is rational and the forecasts are computed using the Wiener–Kolmogorov formula

$$\begin{aligned} E_t y_{t+1} &= E_t[C_0 \in_{t+1} + C_1 \in_t + C_2 \in_{t-1} + \ldots] \\ &= C_1 \in_t + C_2 \in_{t-1} + \ldots \\ &= [C(L) - C_0]L^{-1} \in_t \end{aligned} \tag{4}$$

Since $E_t \in_{t+1} = 0$. Note that the forecast is computed using a solution to the model and hence is model consistent. Fourth, the rational expectations restrictions hold for all realizations of the driving process. Using (2), (3) and (4) equation (1) can be written as

$$\begin{aligned} &[C(L) - C_0]L^{-1} \in_t - (\rho_1 + \rho_2)C(L) \in_t \\ &+ \rho_1\rho_2 C(L)L \in_t \\ &= A(L) \in_t \end{aligned} \tag{5}$$

where it is assumed that (5) holds for all realizations of $\{\varepsilon_t\}$The solutions are obtained by exploiting the property that the z transforms of the sequences represented in (5) must be identical as analytic functions on the open unit disk.

The solutions are now presented for the three cases corresponding to three different sets of values for the parameters ρ_1 and ρ_2. First suppose $|\rho_1| < 1, |\rho_2| < 1$.The Wold representation for the solutions $\{y_t\}$ is

$$y_t = \{(1 - \rho_1 L)(1 - \rho_2 L)\}^{-1}\{A(L)L + C_0\} \in_t \tag{6}$$

which can be written as

$$y_t = \{(1 - \rho_1 L)(1 - \rho_2 L)\}^{-1}\left\{L - C_0 A(L)^{-1}\right\} x_t \tag{6$'$}$$

provided that $\{x_t\}$ has an autoregressive representation. In this case the expectational difference equation (1) does not uniquely determine the solution $\{y_t\}$. For any finite value of C_0 (6) gives a process lying in the space of the driving process which satisfies equation (1). Since C_0 is a parameter in the forecasting formula the model does not completely determine the forecasting procedure of the agents.

The second case is $|\rho_1| < 1 < |\rho_2|$.In this case (1) and (2) determine a unique solution for y_t:

$$y_t = \{(1 - \rho_1 L)(1 - \rho_2 L)\}^{-1}\left\{(L - \rho_2^{-1} A(\rho_2^{-1}) A(L)^{-1}\right\} x_t. \tag{7}$$

This case applies when (1) is interpreted as the Euler equation in a linear-quadratic costly adjustment model; see Kennan (1979).

The third case is where $1 < |\rho_1|, 1 < |\rho_2|$. In this case there is no solution lying in the space of the driving process.

There are several econometric implications of the solutions. First, the parameters of (6) and (7) depend on the parameters of both the driving process and the expectational difference equation. Thus, there are cross-equation restrictions between the parameters of the reduced form and the driving process.

Sargent (1981) has called the cross-equation restrictions the ‘hallmark of rational expectations’. If x is a policy variable and if a change in policy is described by a change in the parameters of the x process, then a policy change induces a change in the values of the reduced form parameters. The consequence is that if the reduced form parameters are estimated from data generated by the existing policy regime, the resulting estimates may produce a misleading forecast of what will happen under a different policy regime. This point is spelled out in Lucas's (1976) critique of econometric policy evaluation. The connection between the notion of exogeneity and the Lucas critique is discussed by Engle et al. (1983). See also Sims (1982).

The second point is that when $|\rho_1| < 1$ and $|\rho_2| < 1$ there may be many stationary solutions in the space of the driving process. The nature and implication of the multiple solutions has been discussed by Gourieroux et al. (1982), Broze et al. (1985) and Evans and Honkapohja (1986). A number of criteria have been proposed for eliminating some of the solutions. Examples include Taylor's (1977) 'minimum variance' criterion, which chooses the solution with the smallest variance, McCallum's (1983) 'minimum state variable' criterion, which chooses the solution which depends on the fewest other variables and Evans's (1985) 'expectational stability' criterion, which chooses solutions that are stable given a small deviation of the expectations functions from rational expectations equilibrium.

The search for selection criteria in linear rational expectations models has a resemblance to a parallel activity in game theory. Games of complete as well as incomplete information can have multiple equilibria. Several selection criteria – 'refinements' to the concept of Nash equilibrium – have been developed for the purpose of eliminating some of these equilibria. These criteria, or refinements, include Selten's (1965) 'subgame perfection', Selten's (1975) 'trembling hand perfection' and the Kohlberg and Mertens (1986) 'stability' criterion. It is difficult to find appealing arguments for eliminating solutions for linear rational expectations models when the expectational difference equation is not the first order condition to a well posed optimization problem.

Third, in the case of the unique solution (7) it is the relation between the C_j's and not the absolute size of these coefficients that is determined. This can be seen from the renormalization $C^* = C_0^{-1}C_j$ and $\upsilon_t^* = C_0^{-1} \in_t$. The same rescaling procedure can be applied to the representation for x_t.

Fourth, the second order case is the simplest case where all three possibilities exist: many, one and no solutions in the space of a stationary driving process. The case of no solutions is of special interest since in empirical studies the estimation procedures assume that a stationary solution exists.

Fifth, there are solutions lying outside the space of the driving process, some of which are nonstationary. Nonstationary solutions exist whether or not the driving process is stationary.

Estimation

The problem considered here is the estimation of a rational expectations model by an outside econometrician. Hansen (1982) has shown that under certain assumptions there are strongly consistent estimators for the parameters of linear and nonlinear rational expectations models. A key assumption is that the driving process and the solution are stationary and ergodic. This assumption again highlights the importance of the driving process.

A stationary and ergodic driving process $\{x_t\}$ is illustrated by the first order autoregressive process:

$$x_t = ax_{t-1} + \in_t, \quad |a| < 1, \tag{8}$$

where $\{\varepsilon_t\}$ is independently identically normally distributed. The moving average representation of the $\{x_t\}$ process (8) is

$$x_t = (1 - aL)^{-1} \in_t = A(L) \in_t. \tag{8'}$$

If $\{y_t\}$ is in the space of the driving process, then it is also stationary and ergodic.

For empirical work the assumption of stationarity and ergodicity is a demanding one, especially for times series as opposed to cross-sectional series. Nelson and Plosser (1982) have provided evidence that a number of macroeconomic variables such as GNP and the money supply behave very similar to random walks or integrated processes rather than stationary processes about a trend. In practice the driving variables are often detrended. The detrending of a random walk produces a number of spurious effects. For example, Nelson and Kang (1981, 1983) have shown that regressing a trend-free random walk against a time trend will result in the misleading inference that the trend is significant and that the detrended series is serially correlated. Some examples of the empirical implications of trends versus random walks for rational expectations models are discussed in Deaton (1986). The general asymptotic theory for testing the random walk versus the time trend model has been recently developed by Durlauf and Phillips (1986). They examine analytically the effects of spuriously detrending random walks.

The fact that the driving process is stationary does not imply that the solution is stationary since the solution may not be in the space of the driving process. Hence there is the additional problem of testing whether the solution is stationary.

Turning to estimation, the objective is to estimate the parameters of the structural equation rather than the parameters of the driving process. The parameters of the expectational difference equation (1) are often interpreted as the coefficients of the utility function or production function of a representative agent or firm and hence it is these parameters which are of economic interest.

There are two approaches to the estimation of the structural parameters. One is to estimate the structural parameters via the reduced form. The estimates produced by the reduced form approach depend on the specification of the driving process and on which solution is selected. Suppose the driving process is the first order autoregression (8) and the parameters of equation (1) are estimated from the non-unique solution (6). Substituting $(1-aL)^{-1}$ for $A(L)$ in (6) gives

$$y_t = \{(1-\rho_1 L)(1-\rho_2 L)\}^{-1}\{L + C_0(1-aL)\}x_t, \tag{9}$$

which can be rearranged as

$$\begin{aligned} y_t = {} & (\rho_1+\rho_2)y_{t-1} - \rho_1\rho_2 y_{t-2} + C_0 x_t \\ & + (1-C_0 a)x_{t-1}. \end{aligned} \tag{9$'$}$$

Since (9′) is an exact relation the coefficients of (9′) can be calculated exactly from four sample points of the form $y_{t'}, y_{t'-1}, y_{t'-2}, x_{t'}, x_{t'-1}$ Given the prior information $|\rho_1| < 1$ and $|\rho_2| < 1$ only the sum and the product of ρ_1 and ρ_2 can be identified from the data.

Assume next that $|\rho_1| < 1 < |\rho_2|$ Then substituting $(1-aL)^{-1}$ for $A(L)$ in the unique solution (7) yields

$$y_t - \{(a-\rho_2)(1-\rho_1 L)\}^{-1}x_t, \tag{10}$$

which can be rewritten as

$$y_t - \rho_t y_{t-1} = (a-\rho_2)^{-1}x_t. \tag{10$'$}$$

The parameter a can be consistently estimated by applying least squares to (8). In this case (10) and (10′) are also exact relations so that the coefficients of y_{t-1} and x_t can be determined exactly from the two sample points (y_t, y_{t-1}, x_t) and (y_{t+1}, y_t, x_{t+1}) Given an estimate of a an estimate of ρ_2 is obtained from the coefficient of x_t in (10′) and the ρ_1 is determined exactly since it is the coefficient of y_{t-1} in (10′). In this case the prior information $|\rho_1| < 1 < |\rho_2|$ allows the parameters ρ_1 and ρ_2 to be identified.

In empirical studies the sample data does not satisfy exact relations such as (9′) and (10′). This has led to the construction of models based on stories where the agents have more information than the outside econometrician. For example, in a model in which agents face several driving variables, the econometrician may have observations on only some of the driving variables. This is illustrated by

$$E_t y_{t+1} - (\rho_1+\rho_2)y_t + \rho_1\rho_2 y_{t-1} = x_t = x_t' + \eta_t. \tag{11}$$

where the outside econometrician observes only x_t' Whiteman calls (11) a 'perturbed equation'.

Observe that in the perturbed version of (9′) the parameters C_0 and a are over-identified since given an estimate of a from the driving process two estimates of C_0 can be calculated from the perturbed version of (9′) and also two estimates of a; one obtained from the driving process and one from (9′).

The coefficients of the perturbed equation (11) can be consistently estimated from the reduced form provided certain conditions are satisfied. As an illustration suppose the autoregression (8) is the driving process where the ε_t's and n_t's are serially and mutually independent. Applying least squares to the driving process and to the perturbed version of (10′) produces consistent estimates of the coefficients in these equations. A consistent estimates of the structural parameter ρ_2 is derived from the coefficient of x_t in the perturbed version of (10′) using the least squares estimate of the parameter a in the autoregression (8).

Asymptotically efficient estimators of the parameters of the driving process and the structure can be obtained by using the method of maximum likelihood. Hansen and Sargent (1980) show that the maximum likelihood estimator is asymptotically efficient only if it maximizes the joint likelihood function of the driving process and the reduced form.

Two problems are encountered using maximum likelihood. First, it is difficult to solve explicitly for the reduced form if structural equation is nonlinear and the driving process is complicated. For the case of nonlinear expectational difference equations Fair and Taylor (1983) have proposed an approximate maximum likelihood procedure which circumvents some of the computational difficulties of obtaining a complete characterization of the reduced form. In particular, they develop a method for solving numerically for the reduced form. The second is that the maximum likelihood estimator may not be consistent, or, if consistent, not efficient, when the model is misspecified. Hansen and Singleton (1982) present an example in which the maximum likelihood estimator fails to be consistent due to a misspecification of the stochastic properties of the driving process.

The other approach is to estimate the structural parameters directly. This approach applied to equation (1) can be motivated as follows. The difference between y_{t+1} and the conditional expectation $E_t\, y_{t+1}$ is

$$u_{t+1} = y_{t+1} - E_t y_{t+1} = [C(L) - (C(L) - C_0)]L^{-1}\varepsilon_t = C_0 \in_{t+1},$$

which implies that

$$E_t u_{t+1} = 0 \tag{12′}$$

and hence that y_{t-1} is a conditionally unbiased estimate of the conditional expectation. This condition can be interpreted as the first order condition to a linear-quadratic optimization problem. By assumption the forecast error is orthogonal to the observed forecast and to any other variables in the information set of agents when the forecast is made. Substituting y_{t+1} for $E_t\ y_{t+1}$ in (1) and rearranging gives

$$y_{t+1} = (\rho_1 + \rho_2)y_t - \rho_1\rho_2 y_{t-1} + x_t + u_{t+1}. \tag{13}$$

From (13) it is seen that the 'error' (12′) introduced by the substitution is contemporaneously uncorrelated with the 'regressors' y_1, y_{t-1} and x_t provided that the $\in_t$'s are serially uncorrelated. Thus, consistent estimates of the parameters ρ_1 and ρ_2 can be obtained by applying least squares to (13). If instead of (1) the starting point is the perturbed equation (11), least squares is consistent if the error η_t in (11) is independent of the error u_{t+1} The direct structural approach was used by Kennan (1979) to estimate a perturbed version of an Euler equation.

Note that the value of C_0 and the variance of $\in_{t+1}$ combine to determine the variance of the error (12′) and that the direct structural approach gives a consistent estimate of the error variance. As a consequence, the value of C_0 in the multiple solutions case (6) is (implicitly) consistently estimated.

An alternative motivation for the direct structural approach exploits certain orthogonality conditions. Define

$$\begin{aligned} h(z_{t+1}, b_0) &= y_{t+1} - (\rho_1 + \rho_2)y_t \\ &\quad + \rho_1\rho_2 y_{t-1} - x_t \\ &= 0 \end{aligned} \tag{14}$$

where $z_{t+1} = \left(y_{t+1}y_t, y_{t-1}, x_t\right)$ is the vector of variables and $b_0 = [(\rho_1 + \rho_2), (\rho_1\rho_2)]$ is the vector of parameters in (1). Using these definitions (1) can be written as

$$E_t h(z_{t+1}, b_0) = 0 \tag{14$'$}$$

so that given a set of variables $\{w_q\}$ in the agents' information set which are observed by the econometrician

$$E_t\left[h(z_{t+1}, b_0)w_{q,t}\right] = 0 \tag{15}$$

where the variables w_q can be thought of as instrumental variables. Taking the expectation of (15) over the variables in information set gives the unconditional expectation

$$\begin{aligned} EE_t\left[h(z_{t+1}, b_0)w_{q,t}\right] &= E\left[h(z_{t+1}, b_0)w_{q,t}\right] \\ &= 0. \end{aligned} \tag{15$'$}$$

Hansen (1982) defines the general method of moments estimator of the true parameter b_0 as the estimator which makes the sample versions of the population orthogonality conditions (15′) as close to zero as possible according to some measure of distance. Examples of this method include the least squares procedure of Kennan (1979) and a variety of instrumental variable techniques. For identification there must be at least as many orthogonality conditions as parameters to be estimated.

The general method of moments estimators are in general less asymptotically efficient than maximum likelihood if the model is correctly specified. Heuristically, this is because the method of moments does not use all the stochastic properties of the driving process and all the orthogonality conditions. The chief advantages of the method of moments estimators are robustness to misspecification and computational convenience. The method is robust in the sense that the model does not have to be completely specified; in particular, it is not necessary to make precise assumptions about the stochastic properties of the driving process. The computational advantage is that least squares type procedures can be used and that the model does not have to be solved for the reduced form. Hence, the method is especially suited to the estimation of nonlinear rational expectations models. In many applications of interest the u_{t+1}'s are serially correlated and conditionally heteroskedastic. Hansen (1982) has also stated conditions under which the method is consistent in the presence of serial correlation and conditional heteroskedasticity.

Learning

One possible and appealing justification for the use of rational expectations is that agents learn to form rational expectations. There is a large literature on learning to form rational expectations, much of which is surveyed in Blume et al. (1982). The literature falls into two parts: one is concerned with 'rational learning' in which the model is correctly specified and agents form rational expectations given knowledge of the model and estimates of its parameters. Examples include Townsend (1978, 1983), Brandenburger (1984) and Bray and Kreps (1986). Rational learning is the natural extension of the standard methodology, based on optimization, to learning. Bray and Kreps (1986) show that it also guarantees convergence to rational expectations equilibrium under quite mild assumptions. The case which have been studied suppose a substantial degree of insight and prior knowledge of the part of agents.

The other part of this literature assumes some degree of bounded rationality. Examples of this type include Bray (1982, 1983), Radner (1982) Frydman (1982), Bray and Savin (1986) Fourgeaud et al. (1986) and Marcet and Sargent (1986). In the bounded rationality framework agents are assumed to learn using reasonable model specifications which are often correct in rational expectations equilibrium, but misspecified when there is learning.

R

Following the classic paper by Muth (1961), the cobweb model has been used in the discussion of expectations formation of Townsend (1978), Brandenburger (1984), Frydman (1982), Bray and Savin (1986), Fourgeaud et al. (1986) and others. Townsend (1978) and Bray and Savin (1986) consider a continuum of firms producing a homogeneous good where the set of firms is the unit interval [0,1] indexed by i. The firms make their production decisions at each date t before the realization of an exogenous stochastic demand which depends linearly on p_t, the market clearing price of the good, and on an unobserved exogenous demand shock. Each firm has a quadratic cost function so that the optimal output of firm i at date t is proportional to p_{it}^{e} the mean of firm i's prior on p_t at date t. Setting the average supply to the market equal to the demand gives

$$p_t = x_t'm + a\overline{p}_t^e + u_t \tag{16}$$

where x_t is a vector of exogenous supply shocks observable by firms when the production decision is made, u_t is the difference between the unobservable exogenous shocks in the demand and supply equations and

$$\overline{p}_t^{\mathrm{e}} = \int_0^1 p_{it}^{\mathrm{e}} \mathrm{d}i \tag{17}$$

is the average of the price expectations (prior means) of the firms. In Bray and Savin (1986) the description of the model is completed by assuming that the stochastic processes x_t and u_t are independently identically distributed random variables with bounded forth moments. The equation (16) is a special type of expectational difference equation called a 'withholding equation'. The simplest example of such an equation is

$$E_{t-1}y_t - \rho y_t = x_t. \tag{18}$$

The equation (16) is a perturbed version of (18) due to the addition of the error u_t. Withholding equations are very prevalent in the rational expectations literature. One reason is that a class of models stemming from the absolute versus relative-price confusion paradigm of Lucas (1972, 1975) employs (18). Another reason is that Muth's (1961) cobweb model produces such an equation. The unique solution y_t to (18) lying in the space of a driving process when the ε_t's are independently identically distributed is

$$y_t = (1-\rho)^{-1} x_t \tag{19}$$

provided ρ is not equal to unity. It is important to note that this unique solution exists regardless of the value of ρ. For further details see Whiteman (1983).

From (16) the rational expectations equilibrium price forecast is

$$p_{it}^{\mathrm{e}} = x_t'm(1-a)^{-1} \tag{20}$$

for all i, provided a is not equal to unity. This solution is essentially the same as (19). Substituting this forecast in (16) the price in rational expectations equilibrium is the random variable

$$p_t = x_t'm(1-a)^{-1} + u_t. \tag{21}$$

Hence if agents know the numerical value of $m(1-a)^{-1}$ they can form rational expectations.

The learning procedure followed by agents should depend on how much they know about the model and the way other agents learn. Suppose all agents know the numerical value of a and can observe or infer $\overline{p}_t^{\mathrm{e}}$. Then (16) can be written as $y_t = x_t'm + u_t$ where $y_t = p_t - a\overline{p}_t^{\mathrm{e}}$ is an observable variable. This equation satisfies the assumptions of the standard linear model so that m can be consistently estimated using classical or Bayesian methods.

Townsend (1978) assumes that the agents are Bayesians who know a and have enough common knowledge to infer $\overline{p}_t^{\mathrm{e}}$ and hence can use Bayesian methods to infer m. A similar result under weaker common knowledge assumption has been obtained by Brandenburger (1984). In Townsend and Brandenburger the agents know that they are in a market game where the actual price depends on the collective output decisions of all the firms. Each agent calculates the Bayesian Nash equilibrium price of the game at each date and uses this as the price forecast. These examples assume Bayesian learning based on correctly specified likelihood functions, that is, likelihood functions which take into account the forecast feedback.

By contrast, in Bray and Savin (1986) and Fourgeaud et al. (1986) the agents do not know the value of a and use a misspecified model for forecasting price. The agents assume that

$$p_t = x_t b + u_t \tag{22}$$

and that (22) satisfies the assumptions of the standard linear model and estimate b using classical or Bayesian techniques. For simplicity, suppose that the agents are classical statisticians and that b is estimated after observing $(x_1, p_1), \ldots, (x_{t-1}, p_{t-1})$ by

$$b_{t-1} = \left[\sum_{j=1}^{t-1} x_j x_j'\right]^{-1} \left[\sum_{j=1}^{t-1} x_j p_j\right]. \tag{22}$$

The agent's forecast of p_t is $x_t' b_{t-1}$. Substituting this forecast into (16) gives

$$p_t = x_t'(m + ab_{t-1}) + u_t. \tag{23}$$

Equations (22) and (23) describe the true data generation process. Comparing (23) with (21) it is clear that agents are using a misspecified model since they are assuming that b in (21) is a constant when in fact the learning process induces a time-varying parameter $m + ab_{t-1}$ The specification (21) is not arbitrary since it would be correct in rational expectations equilibrium, that is, if the value $m(1-a)^{-1}$ were known and used in forecasting. On the other hand, the agents are not fully rational because they fail to employ the relevant model to deduce (16) and hence to deduce that the forecasting procedure based on (21) implies the time-varying parameter model (23), which in turn implies that the forecasting procedure based on (21) is inconsistent with the model.

It can be shown that b_t cannot converge to any other value than the rational expectations equilibrium value $m(1-a)^{-1}$ and that if $a < 1$, b_t strongly converges to $m(1-a)^{-1}$. In this case agents eventually learn how to form rational expectations. When $a < 1$ the demand curve crosses the supply curve from above, which is the standard economically plausible case. Fourgeaud et al. (1986) and Marcet and Sargent (1986) present proofs for the case where b is estimated by least squares and Bray and Savin (1986) for the case where agents are Bayesian statisticians.

When $a > 1$ it appears that b_t does not converge. In this case b_t follows one of a variety of divergent processes including a random walk. The nonconvergence is due to the unstable cobweb since the driving process is stable.

The question of the rate of convergence of b_t to the rational expectations equilibrium value is of considerable interest and has been investigated by Bray and Savin (1986). If the rate of convergence is fast, then the learning procedure works in the sense of generating expectations which are very nearly rational in a short time. Rapid convergence justifies the use of the rational expectations equilibrium as a good asymptotic approximation to a learning process and encourages the application of rational expectations models to actual data.

If convergence is slow or does not appear to occur, then the agents will eventually detect that the model (21) is misspecified. As a consequence, the specification of the model may be revised. Whether the sequence of model revisions adopted by agents will eventually lead to rational expectations equilibrium is an open question.

Time-varying models are widely used in empirical studies. Since learning processes can generate data which closely mimics that generated by standard time-varying parameter models, learning is a potentially attractive explanation for the observed phenomenon of time-varying coefficients.

Concluding Comments

The implicit assumption made in the case of an outside econometrician estimating a rational expectations model is that the actions of the outside econometrician do not influence the agents in the economy. This is true if the model estimated by the econometrician is ignored by the agents. If the empirical work of the outside econometrician is in fact ignored, then this raises the question of the motivation for such work. On the other hand, if the outside econometrician's model specification and estimates do have influence, then it is no longer true that the outside econometrician is indeed outside the economy. In this situation the econometrician's model may be misspecified due to forecast feedback.

The stationarity of the driving process and the solution plays an important role in the analysis of the solutions and in the theory of estimation. Nonstationarity appears to be a characteristic of many macroeconomic time series. It is this which accounts for the popularity of time-varying parameter methods in econometrics. This nonstationarity may be in part a product of agents learning how to form rational expectations. Even if agents are fully rational in the sense that they can calculate the Bayesian Nash equilibrium, this does not rule out nonstationarity. The assumption of stationarity may not be consistent with the notion of agents learning to form rational expectations.

In the typical rational expectations model forecasting is assumed to be a costless activity. In practice forecasting is costly, if for no other reason, because it is a time consuming activity; agents may be playing many market games simultaneously or agents may be playing one game which involves substantial amounts of data collection and processing. This suggests that the choice of a forecasting procedure is the outcome of a constrained optimization problem. Thus the assumption of bounded rationality is not necessary to explain why agents do not forecast with all available information. Alternatively, bounded rationality can be interpreted as the result of a budget constraint. Rule of thumb forecasting procedures may closely approximate the procedures selected by constrained optimization. The time constraint also naturally suggests why there is a market for forecasting services. Given the opportunity cost of time it may be optimal for agents to buy forecasts rather than make their own. The econometricians who supply these forecasts are inside the economy which means that econometric modelling is complicated by the presence of forecast feedback. This in turn may explain why econometric models require frequent revision.

See Also

► Econometrics
► Macroeconometric Models

Bibliography

Aoki, A., and M. Canzoneri. 1979. Reduced forms of rational expectations models. *Quarterly Journal of Economics* 93: 59–71.

Blanchard, O.J. 1979. Backward and forward solutions for economies with rational expectations. *American Economic Review* 69: 114–118.

Blanchard, O.J., and C.M. Kahn. 1980. The solution of linear difference models under rational expectations. *Econometrica* 48: 1305–1311.

Blume, L., M.M. Bray, and D. Easley. 1982. Introduction to the stability of rational expectations equilibrium. *Journal of Economic Theory* 26: 313–317.

Brandenburger, A. 1984. *Information and learning in market games*. Churchill College, Cambridge Mimeo, August 1984.

Bray, M.M. 1982. Learning estimation, and the stability of rational expectations. *Journal of Economic Theory* 26: 318–339.

Bray, M.M. 1983. Convergence to rational expectations equilibrium. In *Individual forecasting and aggregate outcomes*, ed. R. Frydman and E.S. Phelps. Cambridge: Cambridge University Press.

Bray, M.M., and D. Kreps. 1986. Rational learning and rational expectations. In *Essays in honour of K.J. Arrow*, ed. W. Heller, R. Starr, and D. Starrett. Cambridge: Cambridge University Press.

Bray, M.M., and N.E. Savin. 1986. Rational expectations equilibria, learning and model specification. *Econometrica* 57: 1129–1160.

Broze, L., C. Gourieroux, and A. Szafarz. 1985. Solutions of linear rational expectations models. *Econometric Theory* 1: 341–368.

Deaton, A. 1986. *Life-cycle models of consumption: Is the evidence consistent with the theory?* Woodrow Wilson School, Princeton University, Mimeo.

Durlauf, S.N., and P.C.B. Phillips. 1986. *Trends versus random walks in time series analysis*, Cowles Foundation discussion paper, no. 788.

Eichenbaum, M.S. 1983. A rational expectations equilibrium model of finished goods and employment. *Journal of Monetary Economics* 12: 259–277.

Engle, R.F., D.F. Hendry, and J.-F. Richard. 1983. Exogeneity. *Econometrica* 50: 227–304.

Evans, G. 1985. Expectational stability and multiple solutions in linear rational expectations models. *Quarterly Journal of Economics* 99: 1217–1233.

Evans, G., and S. Honkapohja. 1986. A complete characterization of ARMA solutions to linear rational expectations models. *Review of Economic Studies* 53: 227–239.

Fair, R.C., and J.B. Taylor. 1983. Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica* 51: 1169–1185.
Fourgeaud, C., C. Gourieroux, and J. Pradel. 1986. Learning procedure and convergence to rationality. *Econometrica* 54: 845–868.
Frydman, R. 1982. Towards an understanding of market processes: Individual expectations, learning and convergence to rational expectations equilibrium. *American Economic Review* 72: 652–668.
Futia, C.A. 1981. Rational expectations in stationary linear models. *Econometrica* 49: 171–92.
Gourieroux, C., J.J. Laffont, and A. Monfort. 1982. Rational expectations in linear models: Analysis of solutions. *Econometrica* 50: 409–25.
Hansen, L.P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.
Hansen, L.P., and T.J. Sargent. 1980. Formulating and estimating dynamic linear rational expectations. *Journal of Economic Dynamics and Control* 2: 7–46.
Hansen, L.P., and K.J. Singleton. 1982. Generalized instrumental variable estimation of nonlinear rational expectations models. *Econometrica* 50: 1269–1286.
Kennan, J. 1979. The estimation of partial adjustment models with rational expectations. *Econometrica* 47: 1441–1446.
Kohlberg, E., and J.-F. Mertens. 1986. On the strategic stability of equilibria. *Econometrica* 57: 1003–1038.
Lucas Jr., R.E. 1972. Econometric testing of the natural rate hypothesis. In *Econometrics of price determination conference*, ed. O. Eckstein. Washington, DC: Board of Governors of the Federal Reserve System.
Lucas Jr., R.E. 1975. An equilibrium model of the business cycle. *Journal of Political Economy* 83: 1113–1144.
Lucas Jr., R.E. 1976. Econometric policy evaluation: A critique. In *The Phillips Curve and labor markets*, Carnegie-Rochester conference on public policy, vol. 1, ed. K. Brunner and A.H. Meltzer. Amsterdam: North-Holland.
Marcet, A., and T.J. Sargent. 1986. *Convergence of least squares learning mechanisms in self referential linear stochastic models*. University of Minnesota, Mimeo.
McCallum, B.T. 1983. On non-uniqueness in rational expectations models: An attempt at perspective. *Journal of Monetary Economics* 11: 139–168.
Muth, J.F. 1961. Rational expectations and the theory of price movements. *Econometrica* 29: 315–335.
Nelson, C.R. and H. Kang. 1983. *Pitfalls in the use of time as an explanatory variable in regression*, NBER technical working paper, no. 30. Cambridge, MA: National Bureau of Economic Research.
Nelson, C.R., and H. Kang. 1981. Spurious periodicity in inappropriately detrended time series. *Econometrica* 49: 741–751.
Nelson, C.R., and C. Plosser. 1982. Trends and random walk in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10: 139–162.
Radner, R. 1982. Equilibrium under uncertainty. In *Handbook of mathematical economics*, vol. II, ed. K.J. Arrow and M.D. Intriligator. Amsterdam: North-Holland.
Saracoglu, R., and T.J. Sargent. 1978. Seasonality and portfolio balance under rational expectations. *Journal of Monetary Economics* 4: 511–521.
Sargent, T.J. 1979. *Macroeconomic theory*. New York: Academic Press.
Sargent, T.J. 1981. Interpreting economic time series. *Journal of Political Economy* 89: 403–410.
Selten, R. 1965. Spieltheoretische Behandlung eines Oligopomodells mit Nachfragetragheit. *Zeitschrift fur die Gesamte Staatswissenschaft* 122: 301–324.
Selten, R. 1975. A re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4: 25–55.
Sims, C.A. 1982. Policy analysis with econometric models. *Brookings Papers on Economic Activity* 1: 107–152.
Taylor, J.B. 1977. Conditions for unique solutions in stochastic macroeconomic models with rational expectations. *Econometrica* 45: 1337–1385.
Townsend, R.M. 1978. Market anticipations, rational expectations and Bayesian analysis. *International Economic Review* 19: 481–494.
Townsend, R.M. 1983. Forecasting the forecasts of others. *Journal of Political Economy* 91: 545–588.
Whiteman, C.H. 1983. *Linear rational expectations models*. Minneapolis: University of Minnesota Press.

Rational Inattention

Mirko Wiederholt

Abstract

Economists have studied for a long time how decision-makers allocate scarce resources. The recent literature on rational inattention studies how decision-makers allocate the scarce resource attention. The idea is that decision-makers have a limited amount of attention and have to decide how to allocate it. The literature on rational inattention argues that the optimal allocation of attention by decision-makers can explain important features of economic data.

Keywords

Attention; Attention allocation; Business cycles; Consumption; Dynamic stochastic general equilibrium (DSGE); Entropy;

R

Expectations; Finance; Fiscal policy; General equilibrium; Imperfect information; Incomplete information; Information; Information choice; Information flow; Information friction; Information processing; Information theory; Interest rate; Learning; Limited; Attention; Macroeconomics; Monetary policy; Noisy signal; Nominal rigidities; Portfolio choice; Price setting; Rational inattention; Savings; Sims, C. A; Sticky prices; Uncertainty

JEL classifications
D8; D9; E3; E5; G1

The idea of rational inattention is that individuals have a limited amount of attention and therefore have to decide how to allocate their attention.

There is a vast amount of information that is, in principle, available to decisionmakers (e.g. information published in books, magazines, newspapers, and scientific articles; information available on the Internet; knowledge available through colleagues, friends and family), but due to limited attention it is simply impossible to attend to all of this information. Therefore decision-makers have to choose which information to attend to carefully, which information to attend to less carefully, and which information to ignore. According to the theory of rational inattention, decision-makers take this decision optimally. The literature on rational inattention argues that the optimal allocation of attention by decision-makers can explain important features of economic data.

Modelling Limited Attention

Christopher A. Sims proposed modelling attention as an information flow and to model limited attention as a bound on information flow; see Sims (1998, 2003). To implement this idea, one has to quantify information flows. Sims (1998, 2003) suggested following the literature on information theory by quantifying information as reduction in uncertainty, where uncertainty is measured by entropy.

Let us illustrate these concepts with a simple example. Entropy is a measure of uncertainty. The entropy of a normally distributed random variable X equals

$$H(X) = \frac{1}{2}\log_2\left(2\pi e \sigma_X^2\right),$$

where $\sigma^2{}_X$ denotes the variance of X. Conditional entropy is a measure of conditional uncertainty. The conditional entropy of X given S, when X and S have a multivariate normal distribution, equals

$$H(X|S) = \frac{1}{2}\log_2\left(2\pi e \sigma_{X|S}^2\right),$$

where $\sigma^2{}_{X|S}$ denotes the conditional variance of X given S. Equipped with measures of uncertainty and conditional uncertainty, one can quantify the information that one random variable contains about another random variable as reduction in uncertainty. For example, the amount of information that S contains about X equals

$$I(X;S) = H(X) - H(X|S).$$

Think of X as a variable that a decision-maker may be interested in. Paying attention to the variable X can be modelled as receiving a signal $S = X + \varepsilon$, where the noise ε is interpreted as coming from the decision-maker's limited attention and is assumed to be independent of X and normally distributed with mean zero and variance σ_ε^2. The idea is that limited attention leads to a noisy perception of the true realisation of X. Limited attention can be modelled as a bound on information flow

$$I(X;S) \le \kappa.$$

The constraint on information flow implies

$$\frac{\sigma_X^2}{\sigma_{X|S}^2} \le 2^{2\kappa},$$

or equivalently

$$\frac{\sigma_X^2}{\sigma_\varepsilon^2} \le 2^{2\kappa} - 1.$$

In this simple example, limited attention simply imposes a bound on variance reduction. Equivalently, limited attention imposes a bound on the signal-to-noise ratio in the signal concerning X. The noise in the signal is interpreted as arising from the decision-maker's own nervous system.

One advantage of measuring uncertainty by entropy is that entropy also summarises in a single number the uncertainty associated with a multivariate distribution. For example, the entropy of an n-dimensional random vector X that has a multivariate normal distribution equals

$$H(X) = \frac{1}{2}\log_2\left[(2\pi e)^n \det\Omega_X\right],$$

where det Ω_X denotes the determinant of the covariance matrix of X. Think of X as a vector of variables that a decision-maker may be interested in. Paying attention to the vector X can be modelled as receiving an m-dimensional signal S with the property that X and S have a multivariate normal distribution. Limited attention can again be formalised as a constraint on information flow

$$H(X) - H(X|S) \leq \kappa.$$

The constraint on information flow now implies

$$\frac{\det\Omega_X}{\det\Omega_{X|S}} \leq 2^{2\kappa},$$

where $\Omega_{X|S}$ denotes the conditional covariance matrix of X given S. One can then ask how the decision-maker would want to use the available information flow. What is the optimal dimension of S? Which elements of X would the decision-maker want to learn about? Would the decision-maker want to learn about linear combinations of elements of X?

Moreover, entropy can be computed for discrete and continuous distributions. Let X be a discrete random variable with support χ and probability mass function $p(x) = \Pr\{X = x\}$. Then the entropy of X equals

$$H(X) = -\sum_{x\in\chi} p(x)\log_2 p(x).$$

Paying attention can again be modelled as receiving a signal S and limited attention can be modelled as a constraint on information flow

$$H(X) - H(X|S) \leq \kappa.$$

One can once more ask how the decision-maker would want to use the available information flow. Is it optimal to find out only whether X is above or below a certain level? What is the optimal form of the joint distribution of X and S?

For a different approach to modelling attention, see Reis (2006). In his work, paying attention is modelled as incurring a fixed cost and then learning everything perfectly.

Rational Inattention

Let us now study a rational inattention problem. In particular, let us study the problem of a decision-maker who is responsible for setting a price and has to decide how to allocate his or her attention. Let p_i denote the price of good i. Setting a price p_i that differs from the profit-maximising price p_i^* causes a profit loss equal to $\frac{\omega}{2}\left(p_i - p_i^*\right)^2$. The profit-maximising price equals $p_i^* = \phi x$ where x is a normally distributed random variable with mean zero and variance σ_x^2. Here ω and ϕ are parameters. Paying attention to the variable x is modelled as receiving a signal $s_i = x + \varepsilon_i$ where the noise ε_i is independent of x and normally distributed with mean zero and variance σ_ε^2. The decision-maker chooses the amount of attention κ devoted to the variable x. The decision-maker faces a marginal cost of attention $\mu > 0$. This cost can be interpreted as the opportunity cost of devoting some of the scarce resource attention to the variable x. Formally, the decision-maker solves

$$\min_{\kappa\geq 0}\left\{\frac{\omega}{2}\left[E\left(p_i - p_i^*\right)\right] + \mu\kappa\right\},$$

subject to $p_i^* = \phi x$, $s_i = x + \varepsilon_i$, $p_i = E\left[p_i^*|s_i\right]$, and

$$\frac{1}{2}\log_2\left(\frac{\sigma_x^2}{\sigma_{x|s}^2}\right) \leq \kappa.$$

The optimal amount of attention devoted to the variable x equals

$$\kappa^* = \begin{cases} \frac{1}{2}\log_2\left(\frac{\omega\phi^2\sigma_x^2\ln(2)}{\mu}\right) & \text{if } \frac{\omega\phi^2\sigma_x^2\ln(2)}{\mu} \geq 1 \\ 0 \ \text{otherwise} \end{cases};$$

and the price set by the decision-maker equals

$$p_i = \left(1 - 2^{-2\kappa^*}\right)\phi(x + \varepsilon_i).$$

The ratio $\frac{\omega\phi^2\sigma_x^2\ln(2)}{\mu}$ is the marginal benefit of paying attention to the variable x at $\kappa = 0$ divided by the marginal cost of paying attention to the variable x. If this ratio exceeds one, the decision-maker pays some attention to the variable x. The larger the cost of a price-setting mistake and the larger the variance of the profit-maximising price due to the variable x (i.e. the larger ω and $\phi^2\sigma_x^2$), the more attention is devoted to the variable x and the stronger is the response of the price p_i to changes in x.

The example given above is static. One can make the example dynamic by introducing time and by specifying the stochastic process $\{x_t\}_{t=0}^{\infty}$. One can then solve for the optimal signal process $\{s_{i,t}\}_{t=0}^{\infty}$, subject to a constraint on information flow. If the variable x_t follows a stationary first-order autoregressive process, it is optimal to receive a signal of the form $s_{i,t} = x_t + \varepsilon_{i,t}$, where the noise $\varepsilon_{i,t}$ is independent across time; see Maćkowiak and Wiederholt (2009), Propositions 3 and 4. The attention devoted to the variable x_t is again increasing in $\phi^2\sigma_x^2$. Furthermore, the more attention devoted to the variable x_t, the smaller the variance of noise in the signal and the faster the response of the price $p_{i,t}$ to changes in x_t. Maćkowiak and Wiederholt (2009) argue that this can explain why prices respond quickly to idiosyncratic shocks and slowly to aggregate shocks. If there is a tradeoff between attending to idiosyncratic conditions and attending to aggregate conditions and idiosyncratic conditions are more volatile or more important than aggregate conditions, then price setters devote more attention to idiosyncratic conditions and prices respond faster to idiosyncratic shocks.

In the example given above, the signal s_i and the external state x are assumed to have a multivariate normal distribution, implying that the action p_i and the external state x have a multivariate normal distribution. Sims (2006) argues that an agent with rational inattention will also choose the optimal form of the joint distribution of the action and the external state, subject to the constraint on information flow. Sims (2006) and Tutino (2009) study this question in the case of consumption saving problems. Maćkowiak and Wiederholt (2009), Woodford (2009) and Matejka (2010) study this question in the case of price-setting problems. When the decision-maker's objective is quadratic and the external state has a normal distribution, a normally distributed signal is optimal.

Sims' idea of rational inattention has been applied to a variety of different decision problems, not just to price-setting problems. Sims (2003, 2006), Luo (2008) and Tutino (2009) study consumption saving problems with a constant interest rate. Maćkowiak and Wiederholt (2010) study a consumption saving problem with a variable interest rate. Van Nieuwerburgh and Veldkamp (2009) and Mondria (2010) study portfolio choice under rational inattention. Maćkowiak and Wiederholt (2010) solve a dynamic stochastic general equilibrium model with rational inattention on the side of decision-makers in firms and households.

Much remains to be done. There are numerous potential applications of the idea of rational inattention, but so far the theory of rational inattention has only been applied to price setting, wage setting, consumption and portfolio choice. Furthermore, it is interesting to test models of rational inattention empirically. See Maćkowiak et al. (2009) and Kacperczyk et al. (2010) for empirical tests of rational inattention theories of price setting and of portfolio choice. Finally, it seems interesting to study policy implications of models with rational inattention. Maćkowiak and

Wiederholt (2010) and Paciello (2010) conduct monetary policy experiments in models with rational inattention, but fiscal policy, for example, has not yet been studied in models with rational inattention.

See Also

- ▶ Monetary Business Cycles (Imperfect Information)
- ▶ Perfect Information
- ▶ Uncertainty and General Equilibrium

Bibliography

Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L. 2010. Attention Allocation over the Business Cycle. Manuscript

Luo, Y. 2008. Consumption dynamics under information processing constraints. *Review of Economic Dynamics* 11(2): 366–385.

Maćkowiak, B., and M. Wiederholt. 2009. Optimal sticky prices under rational inattention. *American Economic Review* 99(3): 769–803.

Maćkowiak, B., and Wiederholt, M. 2010. Business cycle dynamics under rational inattention. CEPR Discussion Paper 7691.

Maćkowiak, B., E. Moench, and M. Wiederholt. 2009. Sectoral price data and models of price setting. *Journal of Monetary Economics* 56(S): 78–99.

Matejka, F. 2010. Rationally Inattentive Seller: Sales and Discrete Pricing. Manuscript

Mondria, J. 2010. Portfolio Choice, Attention Allocation, and Price Comovement. Manuscript

Paciello, L. 2010. Monetary Policy Activism and Price Responsiveness to Aggregate Shocks under Rational Inattention. Manuscript

Reis, R. 2006. Inattentive producers. *Review of Economic Studies* 73(3): 793–821.

Sims, C.A. 1998. Stickiness. *Carnegie-Rochester Conference Series on Public Policy* 49: 317–356.

Sims, C.A. 2003. Implications of rational inattention. *Journal of Monetary Economics* 50(3): 665–690.

Sims, C.A. 2006. Rational inattention: Beyond the linear-quadratic case. *American Economic Review: Papers and Proceedings* 96(2): 158–163.

Tutino, A. 2009. The Rigidity of Choice. Lifecycle Savings with Information Processing Limits. Manuscript

Van Nieuwerburgh, S., and L. Veldkamp. 2009. Information immobility and the home bias puzzle. *Journal of Finance* 64(3): 1187–1215.

Woodford, M. 2009. Information-constrained state-dependent pricing. *Journal of Monetary Economics* 56(S): 100–124.

Rationality

Lawrence E. Blume and David Easley

Abstract

Economic theory takes the individual consumer and firm as a primitive unit of analysis, and so a theory of individual agency is required to derive hypotheses about the behaviour of markets and other systems of economic interest. One such theory is the principle of rationality, whereby agents act in their perceived best interest. This article surveys the implementation of this principle in economic models, and discusses the critiques of the rationality principle and some proposed alternatives from the perspective of the economic modeller.

Keywords

Altruism; Behavioural economics; Bentham, J.; Bounded rationality; Choice under uncertainty; Cognitive models; Consistency; Cooperation; Decision theory; Economic laws; Economic man; Evolutionary models; Expected utility; General choice theory; Hyperplanes; Interpersonal utility comparisons; Marginal utility; Mechanisms; Menu choice models; Methodological individualism; Neuroeconomics; Non-cooperative game theory; Non-expected utility decision theory; Preference formation; Preference relation; Prospect theory; Rational choice; Rational expectations; Rational expectations equilibrium; Rationality; Rationality principle; Reciprocity; Revealed preference theory; Savage's subjective expected utility model; Self-interest; Social norms; Social preferences; Statistical decision theory; Transitivity; Trust; Utilitarianism; Utility

JEL Classifications

D0

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, . . .
> (Justice Potter Stewart, 378 U.S. 184, 197)

Rationality is for economists as pornography was to the US Supreme Court, undefinable but nonetheless easily identified; and yet, like the Justices of the Court, no two economists share a common definition. This article details some of the common meanings of individual (as opposed to social) rationality and discusses their uses. Our point of view is that of working economists rather than that of psychologists. Economics is committed to methodological individualism, the claim that social phenomena must be explained in terms of individual actions which in turn must be explained through individuals' motivations. This commitment requires a theory of human action. The *rationality principle*, that individuals act in their best interest as they perceive it, provides such a theory. In this article we evaluate the rationality hypothesis and its alternatives from the perspective of how they explain social phenomena such as the behaviour of a market. Our interest is in social life rather than in the psychology of an individual.

History and Description

The use of the rationality principle in economics certainly predates the utilitarianism with which it is so often conflated. Adam Smith (1789, p. 19) describes, in his discussion of the division of labour, a tribe of hunters in which one person is particularly deft at making bows and arrows. 'He frequently exchanges them for cattle or for venison with his companions; and he finds at last that he can in this manner get more cattle and venison, than if he himself went to the field to catch them. From a regard to his own interest, therefore, the making of bows and arrows grows to be his chief business . . .' Moving from intuition to analysis, however, requires a sharp understanding of what it means to regard one's own interest, and this has become a source of endless debate for rational-actor social scientists.

The utility-maximization version of rationality springs from the utilitarianism of Bentham and Mill. According to Bentham (1789, p. i),

> Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne.

Although many thinkers toyed with utilitarian approaches to economic analysis, it was not until the 1870s, through the work of Jevons, Menger and Walras, that utility maximization began to assume the important role in economic analysis it has since held. For this trio, utility was a short cut to a theory of value. Perhaps this is why they were not overly concerned with the issues of measurable utility and the possibility of interpersonal utility comparisons which so exercised their successors. Utility for Bentham, on the other hand, was a physical measure of pain and pleasure which could be computed according to his 'felicific calculus'. Although utility as a merely hedonic measure was rejected even by Mill, only in the 1930s, and after a half century's work beginning with Fisher (1892) and Pareto (1895) was it generally recognized that properties of demand derived from the shape of indifference curves, and so utility could admit a purely ordinal interpretation. This 'shift in emphasis away from the physiological and psychological hedonistic, introspective aspects of utility', as Samuelson (1947, p. 90–1) put it, led to the 'purging out of objectionable, and sometimes unnecessary, connotations . . . of the Bentham . . . variety'. The ultimate expression of this non-psychological view is the theory of revealed preference, whose purpose is '. . .to develop the theory of consumer's behavior freed from any vestigial traces of the utility concept' (Samuelson 1938a, p. 71). The result is a mathematical structure that Edgeworth would have understood, interpreted in a manner completely foreign to his way of thinking.

Expected utility in the theory of choice under uncertainty is older than Benthamite utilitarianism. Both an expectation argument and a dominance (admissibility) argument for the existence of God were carefully laid out by Pascal (1672, p. 233). These remarkable few paragraphs touch on many important issues in contemporary decision theory, including the principle of insufficient reason, the problem of infinite utility payoffs, and incomplete preferences: 'Yes; but you must wager. It is not optional. You are embarked. Which will you choose then?' Even the concept of marginal utility predates Bentham, in Gabriel Cramer's and Daniel Bernoulli's famous near-resolutions of the St. Petersburg paradox. But despite this early progress, the formalization of the modern theory of choice under uncertainty begins only with Wald (1939), who at one go describes the key structures of statistical decision theory: loss functions, a priori distributions, and Bayes, admissible, and minimax decision rules. Interest quickly coalesced, however, around the expected utility models described in the two great testaments of decision science, von Neumann and Morgenstern (1947) and Savage (1956). Expected utility (EU) quickly became such a dominant paradigm for choice under uncertainty that research into alternatives was a backwater for 20 years. But criticisms of the expected utility models emerged almost before the ink was dry on the two manuscripts, in Allais' (1953) experiments and Cyert et al. (1956) empirical studies of firm behaviour, and by the late 1970s behavioural economics and non-EU decision theory were active areas of research.

Psychological utilitarianism and decision theory are the two traditions which most inform the modern economist's thinking about 'rationality', and yet, despite the long intellectual history of these ideas, no single vision of what it means to be a 'rational actor' has emerged. In the remainder of this article we single out several sources of confusion and disagreement. We discuss five models of rationality.

General Choice Theory (GCT)

A set A of alternatives is given, along with a collection $\mathcal{B}$ of non-empty subsets of A. The set A is the set of possible alternatives and any member B of $\mathcal{B}$ is a set of feasible alternatives, a set from which the decision-maker must choose. A *choice function* C assigns to each $B \in \mathcal{B}$ a nonempty subset of $\mathcal{B}$, the objects chosen by the decision-maker from the feasible set. In the theory of demand, for instance, A is the consumption set, B is the collection of possible budget sets and the choice function is the demand function. A textbook treatment of the rational decision-maker requires that she have a *preference relation* $\succcurlyeq$ on A, and we understand $a \succcurlyeq b$ to mean that she finds a to be 'at least as good as' b. By 'preference relation' we mean a binary relation which is complete, all alternatives can be compared, and transitive. Transitivity means that if a is at least as good as b and b is at least as good as c, then a is at least as good as c. Chosen objects in a set B of feasible alternatives are those maximal with respect to the preference relation; b is chosen from B, that is, $b \in C(B)$, if and only if $b \succcurlyeq a$ for all $a \in B$. Preference is the primitive expression of rationality. The role of utility is to provide a convenient *representation* of preference. A utility function u is a real-valued function on A, and to say that *u represents* $\succcurlyeq$ means that $u(a) \succcurlyeq u(b)$ if and only if $a \succcurlyeq b$. While the decision theory toolkit of the working economist mostly specializes this model, much contemporary economic theory does not require this much of rationality. In particular, the completeness and transitivity assumptions can be done away with in general equilibrium theory, and numerical representations for incomplete and non-transitive preferences are available. See, for instance, Aumann (1962), Chipman et al. (1971), and Gale and Mas-Colell (1975).

R

Expected Utility Theory (EU)

Expected utility is a specialization of GCT in which the set A and the preference relation have a specific structure. In EU theory, X is a finite set of prizes or outcomes, and the alternative set A is the set of all probability distributions on X. Preferences have the following representation: A *payoff function* v is a real-valued function on X, and any two probability distributions p and q in A are compared according to their expected values of v; that is $p \succcurlyeq q$ iff $E_p v \succcurlyeq E_q v$. The content of this theory is that, geometrically speaking, indifference curves are parallel straight lines (hyperplanes). The first characterization of EU preferences was provided by von Neumann and Morgenstern (1947); today's standard axiomatic characterization of EU preference orders is due to Herstein and Milnor (1953).

Subjective Expected Utility Theory (SEU)

When we choose whether to play roulette or a slot machine, we are choosing among probability distributions. When we bet on the outcome of a horse or political race, we are betting on the realization of uncertain outcomes, but not objects to which probabilities are necessarily attached. Savage's (1956) contribution was to provide a theory of what he called 'personal probability', a specialization of GCT, here interpreted as a decision-maker's degree of belief in the occurrence of some event. He characterized those preference relations which could be represented by the expectation of some payoff function with respect to a personal probability. In Savage's subjective expected utility (SEU) theory, S is a set of states, such as the possible outcomes of the election. There is also a set X of outcomes. A bet on the election is a function which assigns an outcome to every state. Savage called such functions $f: S \to X$ *acts*, and the set of acts is the alternative set A. A preference relation k on the set A has an SEU representation if there is a payoff function v on outcomes X and a probability distribution p on states S such that $f \succcurlyeq g$ if and only if $E_p\{v(f(s))\} \succcurlyeq E_p\{v(g(s))\}$.

Methodological individualism requires the analysis of social phenomena to be 'bottom-up', that is, to begin with individuals. It is a stronger statement to claim, however, that the description of the individual is entirely pre-social; that in economic models, for instance, that individuals come to the market with preferences and beliefs already formed. Most modern economists do not make this claim, and instead work with models in which the description of the individual is an equilibrium outcome. The two most prominent examples of this method are rational expectations equilibrium and non-cooperative game theory.

Rational Expectations Equilibrium (REE)

The rational expectations hypothesis supposes a population of individuals solving decision problems which have a common state space, and furthermore that the state will be chosen according to the 'true distribution' μ, which is determined by the individuals' choices. The payoff $v(c,s)$ to a choice c depends on the state realization s, and preferences over choices are EU: $U_i(c) = E_\mu v\{c, s\}$. The hypothesis asserts that all beliefs will be correct; that is, that all SEU decision-makers have preference representations in which the beliefs are in fact the probability distribution μ, and μ in turn is the distribution of states which is determined by their actions. Rational expectations is a misuse of the adjective. Unfortunately it is probably too late to abandon the term. There is no connection between the rationality principle, which claims that individuals act in their perceived best interest, and the rational expectations hypothesis, which claims that those perceptions meet some *ex ante* standard of correctness. But so labelling a theory is certainly a nice rhetorical move for how it structures subsequent debate.

Non-cooperative Game Theory (NGT)

A population of individuals chooses actions. Individual i's payoff to action c_i, $v(c_i, c_{-i})$ depends upon the choices c_{-i} of the others. He holds probabilistic beliefs about the actions of others, and evaluates a choice according to EU. The social construction of the individual is seen in the determination of beliefs. *Undominated strategies* are those which can be rationalized by *some* choices of beliefs. *Rationalizable strategies* are those which can be justified by some beliefs satisfying a belief restriction, that it be *common knowledge* that all members of the population are EU-rational with some beliefs, and that payoffs be common knowledge (see epistemic game theory: an overview). Nash equilibrium requires, like REE, that everyone's beliefs are correct. Various Nash equilibrium refinements also have belief interpretations (see Nash equilibrium, refinements of).

Rationality and Mind

The merits of the rational choice foundation of economics have been much discussed, both by its practitioners and by its critics. This discussion is often confused, in part because economists are not consistent in how they understand the contents of the rationality hypothesis. Economic theory holds two views of rationality. One is that rationality is consistency of choice, that the tools of choice theory are just an alternative encoding of certain choice functions; the other is that rationality is a theory of intentional behaviour, in which beliefs and desires are meaningful constructs.

Revealed preference theory is the sharpest formulation of the consistency view. It takes demand as primitive and asks if it is *consistent* with the maximization of a preference order. It recovers desires from choice, and only to the extent that choices are different can two desires, preference orders, be distinguished. This view permeates the foundations of decision theory. For Savage (1956, p. 17), 'It is possible that the person **prefers f** to **g**. Loosely speaking, this means that, if he were required to decide between **f** and **g**, no other acts being available, he would decide on **f**'. In this account, preference is defined by choice. This means specifically that if the choice function C on a collection $\mathcal{B}$ of choice sets satisfies certain conditions, then there is a complete and transitive binary relation such that for every $B \in \mathcal{B}$, $C(B)$ contains exactly the elements of B which are maximal in B with respect to the relation. The binary relation is nothing more than an alternative description of C on $\mathcal{B}$. Suppose a new choice set $B' \notin \mathcal{B}$ is considered. What can we guess about the contents of $C(B)$? Knowing that the decision-maker is consistent on B allows the observer to infer nothing at all about $C(B')$.

If revealed preference represents at all a psychology of choice, that psychology is a form of *radical behaviourism*. Radical behaviourism asserts that two mental states are distinguishable only to the extent that some observable behaviour distinguishes them. Behaviours are all that one can theorize about. Samuelson (1938b, p. 344) writes 'of a steady tendency toward the removal of moral, utilitarian, welfare connotations . . .' and of 'the rejection of hedonistic, introspective, psychological elements'. Although the behaviourist position seems extreme, the leading graduate microeconomics textbook writes approvingly of revealed preference: 'Perhaps most importantly, it makes clear that the theory of individual decision making need not be based on a process of introspection but can be given an entirely behavioral foundation' (Mas-Colell et al. 1995, p. 5). Consistency is often justified as discipline. It requires a minimum of assumptions about the beliefs and desires of individuals, and minimizes the possibility of researchers' values and beliefs slipping unbidden into their analyses. It allows the data maximal scope to speak for itself.

Although received economics talks approvingly of rationality as mere consistency, this is not in fact what most economists do. Much of economics involves invisible-hand explanations; aggregate market behaviour emerges from the decisions of many agents. Whether the invisible hand lifts the cup aloft or knocks it over, economic explanation entails explaining how it coordinates for good or ill the motives and interests of diverse individual actors. These kinds of question call for explanations based on the motivations of economic actors, which purely behaviouralist explanations cannot provide. So economists in practice take an intentional view.

The intentional view holds that rational choice theory is a common-sense or 'folk' (as opposed to 'scientific') psychology. Just as in our everyday transactions we use the language of beliefs and desires to interpret and forecast the behaviour of others, so do economists interpret choice behaviour. The investor *believes* that the asset price will be higher tomorrow. She *wants* greater wealth tomorrow. So she *acts* by purchasing the asset. In this view belief and desire are in fact mental states that are connected to action. The folk psychology is a theory of mind which is presumed by economists to be both adequate for a descriptive psychology of decision and accurate enough in its predictions of individual behaviours for the uses to which it is put. Although utility does not exist as a psychophysical quantity, rational choice models provide a representation of the mental states involved in judgment and decision. (The stronger claim that mental process is a more or less efficient utility maximization algorithm is a view held only by the straw man regularly beaten up by rationality's critics.)

The economist's folk psychology goes further than everyday folk psychology by specifying analytic representations of beliefs, desires, and how they interact. No matter what representation is ultimately chosen by the textbook economist, his folk psychology rests on two points. (1) Rationality is instrumental. Its concern is the efficient pursuing of ends by available means, not the sensibility of the ends. (2) Desire is not anchored by any other aspect of the decision problem, whether the feasible set or the context of choice. Formally, desires are captured by a preference ordering on possible objects of choice whose existence is independent of the feasible set or the context of choice. This is the content of GCT.

The tension between the demands for a parsimonious behavioural theory and the need for an intentional theory of choice is often resolved by holding that, of course beliefs and desires exist, but we economists have access to them only as they are revealed in observed choice behaviour. In a recent critique of neuroeconomics, two well-known theorists write, 'In standard economics, the testable implications of a theory are its content; once they are identified, the non-choice evidence that motivated a novel theory becomes irrelevant' (Gul and Pesendorfer 2005, p. 6). This view has a long history, perhaps with origins in the defence of marginal analysis against its early critics. Machlup (1946, p. 537) writes, 'Psychologists will readily confirm that statements by interviewed individuals about the motives and reasons for their actions are unreliable or at least incomplete', and also raises the oft-heard incentive problem of eliciting survey data, namely, that survey respondents may choose answers to meet their own goals, which may not include truth.

One source of confusion in evaluating claims for and against the economist's psychology is that the theory has both positive and normative components. According to Marshak (1950, p. 111), 'The theory of rational behavior is a set of propositions that can be regarded either as idealized approximations to the actual behavior of men or as recommendations to be followed'. Savage's early work with Milton Friedman (1948, 1952) was explicitly descriptive, but Savage (1956) is just as explicitly normative. It is not surprising that a description of decision in terms of beliefs and desires should have a normative component which evaluates how well goals are achieved. Confusion arises, however, when the descriptive and prescriptive positions are inappropriately conflated to justify the rationality assumptions as a statement of fact. Many undergraduate microeconomics texts justify transitivity assumptions by a money pump argument as a prelude to demand theory. The Dutch book is used to defend probabilistic descriptions of belief. But both of these arguments are, at their source, explicitly normative (see Davidson et al. 1955, p. 146; Ramsey 1931).

A descriptive theory of choice which is grounded not in empirical reality but in logical deductions from normative principles, like Dutch books and money pumps, is not science, but metaphysics. Furthermore, normative justifications are implicitly introspective. A money pump argument really says, 'you wouldn't fall into this trap, would you?' Significant empirical work in psychology (Nisbett and Wilson 1977), however, indicates that introspective evidence is simply unreliable. When individuals turn to review and justify their decisions, they may have no access to the mental states which guided their choice. On the other hand, it seems to us quite reasonable to build models of financial asset pricing which assume that traders are probabilistically sophisticated, on the supposition that traders who are not will either not long survive in the market or not, as a group, be large enough to have a significant effect on prices. Financial markets, unlike Dutch books, actually exist, and the claim that individuals with probabilistically incoherent beliefs do not fare well is a claim of fact, to be tested against market data.

The conflation of positive and normative concerns in decision theory is more fundamental than simple carelessness in an argument. In his criticism of the fact/value dichotomy, Putnam (2002) asks us to consider the word 'cruel'. He observes that the word often has both descriptive and normative content, and in most uses they cannot be separated. The same could be said of the adjective 'rational' in economists' usage. Marshak (1950, p. 111) illustrates this perfectly when he writes that the purpose of EU is '. . . to describe the behavior of men who, it is believed, cannot be "all fools all the time". . ..' When the word 'rational' is used to describe a system in which all agents hold accurate probabilistic beliefs, the implication is that someone holding inaccurate beliefs gets it wrong. REE is often informally defended by the assertion that, if an economic actor's beliefs were incorrect, he would observe this and form new ones. The assertion is either a positive assertion, that actors do indeed have such beliefs, or a normative assertion, that they should hold such beliefs. The normative assertion is a metaphysical defence of the validity of the rational expectations hypothesis. The positive assertion is a claim of fact whose validity could in principle be put to test, but testing the claim would in fact require so rich a set of ancillary maintained hypotheses that practically it is infeasible.

Given all the problems of the two views of rationality, one might wonder why economics needs a rational actor. Dennett (1971, p. 92) provides perhaps the best defence of belief/desire explanations. He contrasts what he calls the *design stance*, predicting behaviour from an understanding of how an agent is designed, or built, with the *intentional stance*, attributing to the agent beliefs and desires, and predicting from them. The intentional stance is useful, he writes, 'Whenever we have reason to suppose the assumption of optimal design is warranted, and doubt the practicality of prediction from the design . . .stance'. Warranting the optimal design assumption means for Dennett not that the system actually be designed to achieve a fixed set of goals, but that this assumption is a useful first approximation. 'Not surprisingly', he observes,

> as we discover more and more imperfections . . ., our efforts at intentional prediction become more and more cumbersome and undecidable, for we can no longer count on the beliefs, desires, and actions going together that ought to go together. Eventually we end up, following this process, by predicting from the design stance; we end up, that is, dropping the assumption of rationality. (p. 95)

This movement, from rationality to realism, is the motivation for taking behaviour more seriously.

Rationality and Behaviours

Game theory and general equilibrium theory are 'system frameworks'. They imagine a collection of individual agents interacting in some systematic way, strategically in game theory, as described by the normal or extensive form of the game, and through markets in general equilibrium theory. In each case, the model produces an 'equilibrium' of the system. The first stage in the development of a system framework involves determining its consistency and internal coherence, that is, conditions which guarantee the existence of equilibrium. This analysis will be as abstract and general as possible, to encompass as large a repertory of behaviours as possible. The second stage is the application of the framework to derive useful statements about the world. This requires explicit behavioural assumptions about agent behaviour and describing the resulting equilibrium. These statements – predictions about market or game behaviour – can be examined empirically.

There are two difficulties with the received models of decision theory such as expected utility and dynamic programming in this kind of research program. First, as these models are formulated, behaviours are not accessible. For example, using expected utility to derive home bias in financial asset markets – that is, investors tend not to take positions in foreign assets – requires complicated assumptions about traders' beliefs. Second, these models are insufficiently rich to capture all the behaviours one might want to examine. For instance, the additively separable intertemporal expected utility model conflates time preference and risk aversion because the model is too thinly parametrized.

Behavioural economics is a research program which will, its proponents argue, replace rational actor models with a more psychologically informed view of human decision making. Much of behavioural economics, however, is less ambitious (and thus, perhaps, more useful). This work can be described as reformulating or extending rational actor models so as to make those observable behaviours whose implications we wish to examine more accessible. While much of this work is at the core of behavioural economics, many who do this work eschew the label; not only behavioural economists are interested in behaviour. Here we discuss four categories of research which cover much work on behaviours, both by behavioural economics and by its critics.

Recontextualizing Decision

GCT is a very parsimonious simplification of a decision problem. In modelling there is a trade-off between behavioural accuracy and parsimony in the description of decision problems. In general equilibrium models, for example, behavioural accuracy may improve descriptive and explanatory power, but parsimony is required because individual decisions are only one piece of the analysis, and complicated models of individual behaviour may generate only intractable market models.

One implication of GCT is that preferences are not choice-set dependent. Even in the early days of decision theory, important models such as minimax regret (Savage 1951) violated the requirement of a single preference order on a universal space of potential choices. Furthermore, many choice-set effects appear to be perfectly rational. Consider the behaviour of a well-mannered but very hungry person at a dinner party. A plate is passed to him with three pieces of the main course, ordered in size such that $a \prec b \prec c$. Being both well-mannered and hungry, he chooses the second largest piece, b. Suppose now that the plate had been passed around the table in the other direction, so that when it comes to him there remains only a and b. Now according to his rule he chooses a. Is he called irrational by the GCT theorists at the table?

Kahneman and Tversky's (1979) prospect theory illustrates another way in which decision problems can be recontextualized. Here additional context, a status quo, is added to the description of the decision problems. Gambles are viewed as probability distributions over gains and losses relative to the status quo. Given a status quo, a preference order over all possible final outcomes exists, but that preference order varies with the status quo. There is, however, a stable preference order over the universe of all possible gains and losses; more context is added by redefining the objects of choice. A similar transformation is accomplished in Gul and Pesendorfer's (2004) model of choice with self-control problems. In the conventional infinite-horizon optimal consumption problem, the objects of choice are consumption paths. Gul and Pesendorfer, on the other hand, take the objects of choice to be pair consisting of a current period consumption and a decision problem to be solved tomorrow. Gul and Pesendorfer's model is an example of a *menu choice model*. Although used somewhat earlier, the first formal development of such models was by Kreps (1979) to describe preferences for flexibility.

Constructing Rationality

The economist's conventional view of market interaction posits a collection of individuals with well-formed preferences meeting in a marketplace. The preferences, along with endowments and technologies, are exogenous to the system. On the other hand, some attendees at a large outdoor concert are there because they like the music, while others are there because of the crowd. Teenagers' evaluation of clothing style has perhaps as much to do with who wears such clothes as with their cut and pattern. These are all examples of socially constructed preferences.

Socially constructed preferences are a part of conventional economic theory. Both NGT and REE are models of socially constructed preferences. In each case desires are fixed, but beliefs adjust. However, neither of these equilibrium concepts is particularly well-supported by belief adjustment (learning) processes. The literature on learning Nash equilibrium is huge, and the state of the art is that, while one can construct learning dynamics that will find a Nash equilibrium, many intuitive learning processes will often fail. Blume and Easley (1982) show that rational equilibrium can easily fail to be reached by any reasonable learning process.

Restricting the socially determined component of preferences only to beliefs is an artificial constraint, and to limit social influence on preference formation to learning is to miss most of the interplay between the individual and the group. Manski (2000) observes that the implications of social interactions through learning and through tastes are distinct, and the difference is significant for policy analysis. Any theory of the interaction of desires requires a new set of primitives which describe the preference formation mechanism. One popular approach has been to model the evolution and workings of pro-social norms of cooperation and trust. Bowles (1998) is an engaging survey of this work. Much less has been done on the evolution and workings of anti-social norms, such as discrimination and stigmatization. Others have turned to biological metaphors. Here one might look at the population dynamics of rules or preferences on a game form or market where game or market outcomes (not utilities) determine the composition in the next round of the population's decision rules or preference orders (Güth and Kliemt 1998; Blume and Easley 1992, 2006). Pro-social behaviour such as reciprocity and altruism has also been investigated from the biological standpoint (Bergstrom 2002; Sethi and Somanathan 2003). One lesson of this literature is that the nature of the interaction between agents is at least as important as the model of choice in determining system outcomes. About the embeddedness of economic action in social life, Granovetter (1985, p. 506) writes,

> The notion that rational choice is derailed by social influences has long discouraged detailed sociological analysis of economic life and led revisionist economists to reform economic theory by focusing on its naive psychology. My claim here is that however naive that psychology may be, this is not where the main difficulty lies – it is rather in the neglect of social structure.

The Content of Preferences and Beliefs

It has been conventional in economic analysis to construe self-interest very narrowly. No 'other-regarding' values are expressed in preferences, and conventionally to do otherwise is frowned upon. For instance, it is hard to explain why an individual votes in an election by her effect on the outcome, without referring to the psychic rewards of the act of voting. Yet the claim that people vote because of norms of citizenship and the like is often regarded as 'nearly tautological' (Ordeshook 1986, p. 50.). On the other hand, critics of economic man often incorrectly assert that rational actors are excessively self-interested; incorrectly, because the existence of preferences and the content of preferences are distinct issues. The rationality hypothesis does not preclude other-regarding desires. Interest in those externalities that arise from ethical concerns, social norms, and other social constructions has increased enormously since the mid-1990s. Much of the literature on social interactions is a study of the consequences of other-regarding preferences. Not surprisingly, other-regarding preferences usefully model both altruism and racism. This is not a fix for those critics who see the selfishness of traditional neoclassical models as a moral failing rather than a behavioural one.

A distinct problem which, unfortunately, has not been much addressed by behavioural economics is the use of individual preferences in the economist's version of moral philosophy. The same preferences which are revealed through shopping behaviour at the grocery store are supposed to be informative for the ethical questions posed by welfare economics. One could, in fact, distinguish 'ethical preferences' from 'subjective preferences' as Harsanyi (1955) has done, and it would be interesting to know if social psychology has anything to say about the relationship between the two types of decision problems, individual and social, which economists address.

Different Psychologies

Some economists look to replace the folk psychology of beliefs and wants with something altogether different. Neuroeconomics is one such attempt, although the neuroeconomics literature seems to eschew drawing economic conclusions from imaging data. Unfortunately, the link between brain and mind is elusive. Eliminative materialism is a position taken by some cognitive scientists, which claims that beliefs and desires do not exist as mental states, and will have no place in an accurate account of the mind. Theoretical and methodological arguments in its support can be found, for instance, in Churchland (1981). An economics which takes its microfoundations entirely from cognitive science could look extremely different than the economics of today. But even if one is more hopeful than Churchland for the utility of the economist's folk psychology, the goal is far off. As one leading neuroimaging specialist puts it, 'Despite fantastic technical developments, lingering methodological and conceptual limitations hinder progress in understanding how mental processes (wrapped up in folk psychology) reduce to or emerge from neural processes' (Schall 2004, p. 44). Savoy (2001, p. 36) has a bleaker view:

> Do the new discoveries about human brain function based on neuroimaging experiments really teach us things that are relevant for the study and understanding of behaviour? That is a question which you must answer. My own impression is that, at present, the overwhelming thrust of these data are toward understanding brain organisation, rather than human behaviour. Of course, we assume that when brain organisation is sufficiently well understood, it will lead to increases in our understanding of behaviour. But I do not think, as yet, there is a great deal of progress in that direction.

Neuroeconomics hopes to replace the belief/desire folk psychology that informs most of modern analytical economics with a more accurate scientific psychology. Alternatively, one could construct a different folk psychology which, like utility maximization, has no scientific pretensions, but is more descriptively accurate. Models of *intrapersonal* conflict are the most familiar example of this kind of framework. Strotz (1955) demonstrated the possibility of time-inconsistent planning in intertemporal utility maximization problems, and Pollak (1968) subsequently displayed the essentially strategic nature of the planning problem as a problem of competition between the selves choosing at different dates. Schelling (1984) described a variety of decision problems with aspects of intrapersonal conflicts, and discussed them from a game-theoretic perspective. He, for instance, wrote about the competition between that part of him which desires nicotine and that part which wants to give it up. This is a contest for self-control. The literature today contains a number of intertemporal models which, following Pollak (1968) distinguish two kinds of behaviour. Sophisticated behaviour chooses today with full knowledge that her future selves may try to undo her decision. A choice is a subgame perfect equilibrium of a game played by all her selves. Naive behaviour chooses today assuming, perhaps incorrectly, that her future selves will stick with her decisions. These two models are intrinsically no more realistic than GCT, just different.

For the working economist, the ultimate test of a psychologically more accurate theory of individual choice is how it performs in explaining market and other social outcomes rather than how well it predicts the behaviour of an individual. Could theory A, more informed with insights from psychology and cognitive science, possibly be less useful for economists? Here are three possibilities: (1) Theory A might be extremely complex. Its application to a heterogeneous-agent financial market model, for instance, is simply impossible to work with, and no conclusions can be derived. (2) Theory A might require for its application data that we can observe in a controlled and heavily instrumented setting but simply cannot collect in the field. (3) Theory A may not be posed with concepts which are useful for the economist's interpretation of social outcomes. For instance, theory A might be couched in terms of chemical states of the brain, and not speak at all about agents' intentions, beliefs or desires. While it may be possible to construct a biochemical model of the invisible hand, it would not be useful for welfare economics.

Evidence on the question of whether these models lead to better market analyses is sparse, and mixed, and there is no evidence on how these models perform relative to menu choice models, which address the same questions from a rational choice perspective. More generally, more work needs to be done in evaluating behavioural models with respect to their economic performance. How useful are they for deriving implications about the performance of aggregate economic variables such as prices? This kind of research is already under way. Two examples are Kocherlakota (2001) and Laibson (1997).

An instance of point (3) can be seen in the time-inconsistency literature. Pollak (1968) and his followers (O'Donoghue and Rabin 1999) see choice not as the expression of a single desire, but as the outcome of conflict, perhaps inefficient and destructive, between competing desires. Now the Pareto ranking of alternatives in a social interaction either becomes dependent on which of the many competing preference orders we modellers choose for each individual or it becomes empty if we try to respect them all. The advantage of menu choice models, the rational-choice approach to modelling problems of self-control, is that there is a well-defined notion of preference for each agent, from which a Pareto ranking can be constructed. To be fair, rational choice modelling also poses problems for welfare economics. If individuals make consistent errors in a class of choice problems, what can revealed preference say about intentions? In the presence of systematic error, a welfare economics built from revealed choice is at best misleading.

Conclusion

The purpose of decision models in economics is to explain the behaviour not of a single individual but of aggregates of individuals. Sometimes economists explain by appeal to ‘Laws’, such as ‘the Law of Supply and Demand’. But this mode of explanation is mostly an intermediate product; useful, perhaps, for generating back-of-the-envelope predictions about the effects of a tax on market price, but not a source of understanding. There are few natural laws in the social sciences, and the domains of the few we can identify are very limited.

More often, economists appeal to ‘mechanism’. We try to understand economic phenomena, such as the determination of prices in different kinds of markets, in terms of the mechanisms which generate them. Given our commitment to methodological individualism, this requires an explanation of how individual economic actors interact with one another. This is where rational actor theories are employed, and it is with respect to how these models do in this discussion rather than how they do in other domains, such as explaining individual behaviour, that the rationality principle should be evaluated.

Unfortunately, perhaps, at this point there are no serious alternatives to the rationality principle. For all of its buzz, proponents of *bounded rationality*, by which we mean models of behaviour that consider beliefs and desires but that do not optimize, have so far failed to deliver decision models which are robust and not tightly tied to a small class of decision problems.

It is perhaps too early in its intellectual history to ask for as much from *cognitive models*. We are sceptical about the value for social and economic systems analysis of unpacking the black box of consumer behaviour by deploying a rich and sophisticated model of cognitive process within a general equilibrium or game theoretic model. There is a point to reductionism. On the other hand, we are enthusiastic about the possibility that cognitive science will contribute to sharpening the rationality principle. The focus of much modern decision theory, such as Kahneman and Tversky (1979), Gilboa and Schmeidler (1989) and Gul and Pesendorfer (2004), has been to make the black-box model better by looking for formulations of rational choice models that better conform to the data. A better understanding of decision mechanism will doubtless suggest constraints on black-box behaviour which can be captured in reduced-form decision models, and perhaps it will uncover constraints that cannot be observed from behaviour alone.

Evolutionary models have also been proposed as an alternative framework to rational choice decision-making. Market forces, or a combination of markets and biology, favour some decision rules over others. In the long run, the market will be populated mostly by those decision rules that are ‘most fit’, rational or not. One can indeed ask if the forces of market selection favour rational decision rules (Blume and Easley 1992; Sandroni 2000), but the study of market population dynamics is complementary to rather than a substitute for rational choice models. Blume and Easley (2006), for instance, demonstrate how market forces select within the class of rational decision rules, favouring some kinds of preferences and beliefs over others.

Although there appear no be no serious alternatives to the rational choice paradigm on the near horizon, there is much to regret in how the rationality principle is discussed. The following statements should be self-evident, but clearly are not, judging by our reading of the literature: (1) Rationality does not mean complete or symmetric information. In fact, much of rational actor social science attempts to understand social outcomes when these conditions do not obtain. (2) Rationality does not require individuals to be entirely selfish. While much effort has been made to understand social norms from the point of view of entirely individualistic preferences, the insistence on relying on self-regarding rather than pro-social preferences is a matter of the content of preferences, rather than an axiom of rationality per se. (3) Rationality does not mean expected utility. Expected utility is one small class of decision models for choice under uncertainty. Its dominance in application was understandable in the 1970s, when few alternatives were on the table. Since then decision theorists have been creative in developing better-behaved alternatives, and equilibrium and game theorists have been clever in applying them. (4) Rationality does not mean 'rational expectations'. For a belief restriction to be a requirement of rationality, it must be clear that all those who are not 'all fools all the time' must have correct beliefs. No research into learning in economics suggests this is the case in any kind of complex environment.

There is also much to regret in how the rationality principle has been deployed in economic analysis. Given the explosion of decision-theoretic research since the 1970s, it is surprising how little this research has affected market and game theoretic analysis. The norm still seems to be self-interested preferences, expected utility and rational expectations (or Nash equilibrium). At this point the question of whether contemporary decision models such as Choquet expected utility and cumulative prospect theory have anything new to say about, say, asset pricing, is open. The value to economists of new decision theories, rational choice or not, is not in how they perform in a laboratory but how they perform in the analysis of markets and other social systems. Too rarely have modern decision theories been exposed to this test.

Rational actor social science is a broader tent than both its supporters and its critics make it out to be. We expect the rational choice framework to be as dominant when the next edition of the *New Palgrave* goes to press as it is today. But we also expect the set of decision-theoretic models deployed in the analysis of social systems will be quite different, and probably more diverse, than it is now.

See Also

▶ Expected Utility Hypothesis
▶ Methodological Individualism
▶ Rationality, History of the Concept
▶ Savage's Subjective Expected Utility Model
▶ Uncertainty
▶ Utilitarianism and Economic Theory
▶ Utility

Bibliography

Allais, M. 1953. Le comportement de l'homme rationel devant le risque: Critique des postulats et axiomes de l'Ecole Américaine. *Econometrica* 21: 503–546.

Aumann, R.J. 1962. Utility theory without the completeness axiom. *Econometrica* 32: 445–462.

Bentham, J. 1789. *An introduction to the principle of morals and legislations. Based on information from English Short Title Catalogue. Eighteenth Century Collections Online*. Farmington Hills: Gale Group.

Bergstrom, T.C. 2002. Evolution of social behavior: individual and group selection. *Journal of Economic Perspectives* 16(2): 67–88.

Blume, L., and D. Easley. 1992. Evolution and market behavior. *Journal of Economic Theory* 58: 9–40.

Blume, L.E., and D. Easley. 1982. Learning to be rational. *Journal of Economic Theory* 26: 340–351.

Blume, L.E., and D. Easley. 2006. If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets. *Econometrica* 74: 929–966.

Bowles, S. 1998. Endogenous preferences: the cultural consequences of markets and other economic institutions. *Journal of Economic Literature* 36: 75–111.

Chipman, J.S., L. Hurwicz, M.K. Richter, and H.F. Sonnenschein. 1971. *Preferences, utility, and demand*. New York: Harcourt Brace Jovanovich.

Churchland, P.M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78(2): 67–90.

Cyert, R.M., H.A. Simon, and D.B. Trow. 1956. Observation of a business decision. *Journal of Business* 29: 237–248.

Davidson, D., J.C.C. McKinsey, and P. Suppes. 1955. Outlines of a formal theory of value, I. *Philosophy of Science* 22: 140–160.

Dennett, D.C. 1971. Intentional systems. *Journal of Philosophy* 68(4): 87–106.

Fishburn, P.C. 1991. Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty* 4: 113–134.

Fisher, I. 1892. *Mathematical investigations in the theory of value and prices*. New Haven: Connecticut Academy of Arts and Sciences.

Friedman, M., and L.J. Savage. 1948. The utility analysis of choices involving risk. *Journal of Political Economy* 56: 279–304.

Friedman, M., and L.J. Savage. 1952. The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy* 60: 463–474.

Gale, D., and A. Mas-Colell. 1975. An equilibrium existence theorem without ordered preference. *Journal of Mathematical Economics* 2: 9–15.

Gilboa, I., and D. Schmeidler. 1989. Maximin expected utility with a non-unique prior. *Journal of Mathematical Economics* 18: 141–153.

Granovetter, M. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology* 91: 481–510.

Gul, F., and W. Pesendorfer. 2004. Self-control and the theory of consumption. *Econometrica* 72: 119–158.

Gul, F., and W. Pesendorfer. 2005. *The case for mindless economics*. Princeton: Princeton University.

Güth, W., and H. Kliemt. 1998. The indirect evolutionary approach: Bridging the gap between rationality and adaptation. *Rationality and Society* 10: 377–399.

Harsanyi, J.C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63: 309–321.

Herstein, I., and J. Milnor. 1953. An axiomatic approach to measurable utility. *Econometrica* 21: 291–297.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.

Kocherlakota, N. 2001. Looking for evidence of time-inconsistent preferences in asset market data. *Federal Reserve Bank of Minneapolis Quarterly Review* 25(3): 13–34.

Kreps, D.M. 1979. A representation theorem for 'preference for flexibility'. *Econometrica* 47: 565–578.

Laibson, D. 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112: 443–477.

Machlup, F. 1946. Marginal analysis and empirical research. *American Economic Review* 36: 519–554.

Manski, C.F. 2000. Economic analysis of social interactions. *Journal of Economic Perspectives* 14(3): 115–136.

Marshak, J. 1950. Rational behavior, uncertain prospects, and measurable utility. *Econometrica* 18: 111–141.

Mas-Colell, A., M.D. Whinston, and J.R. Green. 1995. *Microeconomic theory*. Oxford: Oxford University Press.

Nisbett, R.E., and T.D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 83: 231–259.

O'Donoghue, T., and M. Rabin. 1999. Doing it now or later. *American Economic Review* 89: 103–124.

Ordeshook, P.C. 1986. *Game theory and political theory: An introduction*. Cambridge: Cambridge University Press.

Pareto, V. 1895. Considerazioni sui Principi Fondamentali dell'Economia Politica Pura, part V. *Giornale degli Economisti, ser.* 2(5): 119–157.

Pascal, B. 1672. *Pascal's pensées*, 1958. New York: E.P. Dutton and Co.

Pollak, R.A. 1968. Consistent planning. *Review of Economic Studies* 35: 201–208.

Putnam, H. 2002. The entanglement of fact and value. In *The collapse of the fact/value dichotomy and other essays*. Cambridge, MA: Harvard University Press.

Ramsey, F.P. 1931. Truth and probability. In *The foundations of mathematics and other logical essays*, ed. R.B. Braithwaite. London: K. Paul, Trench, Trubner and Co..

Samuelson, P.A. 1938a. A note on the pure theory of consumer's behavior. *Economica* 5(17): 61–71.

Samuelson, P.A. 1938b. The empirical implications of utility analysis. *Econometrica* 6: 344–356.

Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Sandroni, A. 2000. Do markets favor agents able to make accurate predictions? *Econometrica* 68: 1303–1342.

Savage, L.J. 1951. The theory of statistical decision. *Journal of the American Statistical Association* 46(253): 55–67.

Savage, L.J. 1956. *The foundations of statistics*. 2nd ed, 1972. New York: Dover.

Savoy, R.L. 2001. History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica* 107(1–3): 9–42.

Schall, J.D. 2004. On building a bridge between brain and behavior. *Annual Review of Psychology* 55: 23–50.

Schelling, T. 1984. The intimate contest for self-command. In *Choice and consequence: Perspectives of an errant economist*. Cambridge, MA: Harvard University Press.

Sethi, R., and E. Somanathan. 2003. Understanding reciprocity. *Journal of Economic Behavior and Organization* 50: 1–27.

Smith, A. 1789. *An inquiry into the nature and causes of the wealth of nations*. 5th ed, 1976. Chicago: University of Chicago Press.

Strotz, R.H. 1955. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* 23(3): 165–180.

von Neumann, J., and O. Morgenstern. 1947. *Theory of games and economic behavior*. 2nd ed. Princeton: Princeton University Press.

Wald, A. 1939. Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics* 10: 299–326.

Rationality, Bounded

Herbert A. Simon

Abstract

'Bounded rationality' refers to rational choice that takes into account the cognitive limitations of the decision-maker – limitations of both knowledge and computational capacity. It is a central theme in the behavioural approach to economics. Theories of bounded rationality can be generated by relaxing one or more of the assumptions of subjective utility theory underlying neoclassical economics. They insist that the model of human rationality must be derived from detailed and systematic empirical study of human decision-making behaviour in laboratory and real-world situations. For example, a satisficing strategy may be postulated instead of the maximization of a utility function.

Keywords

Behavioural approach to economics; Bounded rationality; Choice; Cognitive limits; Expected utility; Money illusion; Neoclassical economics; Optimal search; Probability; Rational expectations; Satisficing; Simon, H. A.; Subjective utility theory; Substantive vs. procedural rationality; Uncertainty

JEL Classifications

B4

The term 'bounded rationality' is used to designate rational choice that takes into account the cognitive limitations of the decision-maker – limitations of both knowledge and computational capacity. Bounded rationality is a central theme in the behavioural approach to economics, which is deeply concerned with the ways in which the actual decision-making process influences the decisions that are reached.

The theory of subjective utility (SEU theory) underlying neo-classical economics postulates that choices are made: (1) among a given, fixed set of alternatives; (2) with (subjectively) known probability distributions of outcomes for each; and (3) in such a way as to maximize the expected value of a given utility function (Savage 1954). These are convenient assumptions, providing the basis for a very rich and elegant body of theory, but they are assumptions that may not fit empirically the situations of economic choice in which we are interested.

Theories of bounded rationality can be generated by relaxing one or more of the assumptions of SEU theory. Instead of assuming a fixed set of alternatives among which the decision-maker chooses, we may postulate a process for generating alternatives. Instead of assuming known probability distributions of outcomes, we may introduce estimating procedures for them, or we may look for strategies for dealing with uncertainty that do not assume knowledge of probabilities. Instead of assuming the maximization of a utility function, we may postulate a satisficing strategy. The particular deviations from the SEU assumptions of global maximization introduced by behaviourally oriented economists are derived from what is known, empirically, about human thought and choice processes, and especially what is known about the limits of human cognitive capacity for discovering alternatives, computing their consequences under certainty or uncertainty, and making comparisons among them.

Generation of Alternatives

Modern cognitive psychology has studied in considerable depth not only the processes that human subjects use to choose among given alternatives, but also the processes (problem-solving processes) they use to find possible courses of action (i.e., actions that will solve a problem) (Newell and Simon 1972). If we look at the time allocations of economic actors, say business executives, we find that perhaps the largest fraction of decision-making time is spent in searching for possible courses of action and evaluating them (i.e., estimating their consequences). Much less time and effort is spent in making final choices, once the alternatives have been generated and their consequences examined. The lengthy and crucial processes of generating alternatives, which include all the processes that we ordinarily designate by the word ‘design’, are left out of the SEU account of economic choice.

Study of the processes for generating alternatives quickly reveals that under most circumstances it is not reasonable to talk about finding ‘all the alternatives’. The generation of alternatives is a lengthy and costly process, and one where, in real-world situations, even minimal completeness can seldom be guaranteed. Theories of optimal search can cast some light on such processes, but, because of limits on complexity, human alternative-generating behaviour observed in the laboratory is usually best described as heuristic search aimed at finding satisfactory alternatives, or alternatives that represent an improvement over those previously available (Hogarth 1980).

Evaluation of Consequences

Cognitive limits, in this case lack of knowledge and limits of ability to forecast the future, also play a central role in the evaluation of alternatives. These cognitive difficulties are seen clearly in decisions that are taken on a national scale: whether to go ahead with the construction of a supersonic transport; the measures to be taken to deal with acid rain; Federal Reserve policies on interest rates; and, of course, the supremely fateful decisions of war and peace.

The cognitive limits are not simply limits on specific information. They are almost always also limits on the adequacy of the scientific theories that can be used to predict the relevant phenomena. For example, available theories of atmospheric chemistry and meteorology leave very wide bands of uncertainty in estimating the environmental or health consequences of given quantities and distributions of air pollutants. Similarly, the accuracy of predictions of the economy by computer models is severely limited by lack of knowledge about fundamental economic mechanisms represented in the models’ equations.

Criteria of Choice

The assumption of a utility function postulates a consistency of human choice that is not always evidenced in reality. The assumption of maximization may also place a heavy (often unbearable) computational burden on the decision maker. A theory of bounded rationality seeks to identify, in theory and in actual behaviour, procedures for choosing that are computationally simpler, and that can account for observed inconsistencies in human choice patterns.

Substantive and Procedural Rationality

Theories of bounded rationality, then, are theories of decision making and choice that assume that the decision maker wishes to attain goals, and uses his or her mind as well as possible to that end; but theories that take into account in describing the decision process the actual capacities of the human mind.

The standard SEU theory is presumably not intended as an account of the process that human beings use to make a decision. Rather, it is an apparatus for predicting choice, assuming it to be an objectively optimal response to the situation presented. Its claim is that people choose as if they were maximizing subjective expected utility. And a strong *a priori* case can be made for the SEU theory when the decision making takes place in situations so transparent that the optimum can be reasonably approximated by an ordinary human mind.

Theories of bounded rationality are more ambitious, in trying to capture the actual process of decision as well as the substance of the final decision itself. A veridical theory of this kind can only be erected on the basis of empirical knowledge of the capabilities and limitations of the human mind; that is to say, on the basis of psychological research.

The distinction between substantive theories of rationality (like the SEU theory) and behavioural theories is closely analogous to a distinction that has been made in linguistics between theories of linguistic competence and theories of linguistic performance. A theory of competence would characterize the grammar of a language in terms of a system of rules without claiming that persons who speak the language grammatically do so by applying these rules. Performance theories seek to capture the actual processes of speech production and understanding.

The question of the desirability and usefulness of a procedural theory of decision involves at least two separate issues. First, which kind of theory, substantive or procedural, can better predict and explain what decisions are actually reached. Does SEU theory predict, to the desired degree of accuracy, the market decisions of consumers and businessmen, or does such prediction require us to take into account the cognitive limits of the economic actors?

Second, are we interested only in the decisions that are reached, or is the human decision making process itself one of the objects of our scientific curiosity? In the latter case, a substantive theory of decision cannot meet our needs; only a veridical theory of a procedural kind can satisfy our curiosity.

Bounded Rationality in Neoclassical Economics

It should not be supposed that mainstream economic theory has been completely oblivious to human cognitive limits. In fact, some of the most important disputes in macroeconomic theory can be traced to disagreements as to just where the bounds of human rationality are located. For example, one of the two basic mechanisms that accounts for under employment and business cycles in Keynesian theory is the money illusion suffered by the labour force – a clear case of bounded rationality. In Lucas's rational expectationist theory of the cycle, the corresponding cognitive limitation is the inability of businessmen to discriminate between movements of industry prices and movements of the general price level – another variant of the money illusion. Thus the fundamental differences between these theories do not derive from different inferences drawn from the assumptions of rationality, but from different views as to where and when these assumptions cease to hold – that is, upon differences in their theories of bounded rationality.

What distinguishes contemporary theories of bounded rationality from these ad hoc and casual departures from the SEU model is that the former insist that the model of human rationality must be derived from detailed and systematic empirical study of human decision making behaviour in laboratory and real-world situations.

See Also

▶ Rational Behaviour
▶ Satisficing

Bibliography

Cyert, R.M., and J.G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs: Prentice-Hall. 2nd ed., 1975.
Hogarth, R.M. 1980. *Judgment and choice: The psychology of decision*. New York: Wiley.
Nelson, R.R., and S.G. Winter. 1982. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
Newell, A., and H.A. Simon. 1972. *Human problem solving*. Englewood Cliffs: Prentice-Hall.
Savage, L.J. 1954. *The foundations of statistics*. New York: Wiley.
Simon, H.A. 1982. *Models of bounded rationality.* 2 vols. Cambridge, MA: MIT Press (especially Sections VII and VIII).
Williamson, O. 1975. *Markets and hierarchies*. New York: Free Press.

Rationality, History of the Concept

Esther-Mirjam Sent

Abstract
This article offers a historical and methodological perspective on the concept of rationality. It gives an overview of the various interpretations of the notion, from self-

interest to rational choice and expected utility to strategic rationality and rational expectations. It pays special attention to the ethical dimensions of the concept. The article further places rationality within a long-ranging discussion concerning the status of assumptions within economics. It explicitly considers efforts to test rationality directly. The article concludes with an evaluation of recent efforts to replace rationality with the notion of bounded rationality.

Keywords

Adaptive control learning; Adaptive expectations; Arrow's theorem; Assumptions; Bayes' rule; Bayesian learning; Behavioural economics; Bounded rationality; Cairnes, J. E.; Centipede game; Chain store paradox; Commitment; Completeness of preferences; Computation of equilibria; Confirmation; Error-term justification; Ethics and economics; Expected utility; Falsificationism; Folk theorem; Friedman, M.; Fundamental theorems of welfare economics; Game theory; Hutchison, T.; Instrumental rationality; Introspection; Keynes, J. N.; Learning; Machlup, F.; Microfoundations; Multiple equilibria; Muth, J. F.; Nash equilibrium; Nash program; Neoclassical economics; Neural network learning; New behavioural economics; Notrade theorems; Perfect equilibrium; Pluralism in economics; Power; Preference satisfaction; Preferences; Prisoner's Dilemma; Procedural and substantive rationality; Rational behaviour; Rational expectations; Rationality; Regret; Repeated games; Robbins, L. C.; Satisficing; Self-interest; Senior, N. W.; Simon, H.; Social norms; Strategic behaviour; Subjective probability; Sure-thing axioms; Sympathy; Testing; Transitivity; Verification; Well-being

JEL Classification

B1

Economics has always relied on some notion of rationality. Unlike philosophers, economists are not concerned with the rationality of beliefs, which are taken as data (Tisdell 1975). In the 18th century, economics was integrated into the great scheme of the natural law and a rationalistic world view (Daston 1983; Weber 1904). During this time, the moral sciences aimed to reveal the rational grounds for action and belief. Overall, their focus was individualistic, psychological, and prescriptive. During the 19th century, a transition took place from a psychological framework to a sociological one. At the same time, the search for inexorable social laws replaced the computation of rational self-interest. However, economics continued to cling to rationality. Throughout much of the 20th century, many economists would separate economics from sociology upon the basis of rational or irrational behaviour (Samuelson 1947).

Rationality is usually combined with a variety of other concepts (Arrow 1987; Sen 1987). Indeed, the force of the hypothesis comes from the addition of supplementary hypotheses. What has changed over time, then, is the interpretation of rationality. While it was initially associated with self-interest, in later readings, such as rational choice and expected utility, it became linked with ideas such as consistency and indifference. Recent appeals to it include strategic aspects of behaviour. Within macroeconomics, rational expectations economists have taken rationality to its extreme. Interpretations of rationality cover a wide range that includes it having the status of axiom, a priori truth, self-evident proposition, useful fiction, utopia, ideal type, analytical construct, heuristic construct, indisputable fact of experience, and typical behavioural pattern under capitalism.

Rationality is ubiquitous in modern economics, with the result that economists frequently make the assumption that it has the same meaning in all the contexts in which it is used. However, this is not the case. The assumption of rationality may be motivated by an appeal to the notion of self-interest, with due allowance made for the fact that preferences may extend to the welfare of others, but its use in expected utility theory, the analysis of strategic behaviour, rational expectations, and so on raise issues that are sufficiently profound that the meaning of the concept of rationality is fundamentally changed.

Rationality may further be interpreted as either a positive or a normative notion. Efforts to test rationality interpret the notion in a descriptive manner. That is, rationality is presumed to characterize how people actually go about the business of reasoning. In response, one may investigate the psychological mechanisms and processes that underlie the patterns of reasoning that are observed. By contrast, a normative interpretation of rationality is concerned not so much with how people actually reason as with how they should reason (Suppes 1961). The goal is to discover rules or principles that specify what it is to reason rationally – to specify standards against which the quality of human reasoning can be measured.

In the remainder of this article we first take a closer look at historical debates concerning the overall status of the rationality assumption. We then consider methodological concerns associated with the various historical interpretations of rationality, and subsequently address efforts to test rationality. After this, we take the historical debates up to the present, where we find that more and more economists are moving away from rationality towards the notion of bounded rationality.

The Rationality Assumption

Since rationality is such a central notion within economics, many of the debates about the status of assumptions within economics are related to it. In the early 19th century, John Stuart Mill (1836) argued that economics is an abstract science because it reasons from assumed premises, such as rationality. As a result, its conclusions are true in the abstract. Moreover, Mill continued, that which is true in the abstract is always true in the concrete with proper allowances. This view found support among Nassau Senior, John Elliott Cairnes, and John Neville Keynes. In a similar vein, in the early 20th century, Lionel Robbins (1935) argued that the basic postulates of economics, such as rationality, are simple and indisputable facts of experience. In the opinion of Robbins, the propositions of economic theory are deductions from a series of such self-evident postulates. These insights came under serious attack by Terence Hutchison (1956). First, Hutchison claimed, the propositions of pure theory are empty. Second, maximization and equilibrium require perfect expectations. Third, economics needs more extensive use of empirical techniques. Finally, economists can use the psychological method of a priori facts, the method of Verstehen, and the method of introspection only for suggesting hypotheses, in Hutchison's opinion, and not for establishing them. He concluded that the rationality postulate was treated as analytic by economists, meaning that it is a priori true yet with empirical content. Instead, he claimed that it must be synthetic, meaning that it must be stated in testable form.

Hutchison's arguments about the status of the rationality assumption found a serious critic in Fritz Machlup (1956). First, Machlup argued, rationality is a theoretical construct. Second, empirical studies judge applicability. That is, they confirm rather than offer complete verification. Third, economists need to focus on 'realistic' assumptions embedded in a system of interrelated hypotheses. Finally, they need a suitable replacement in case of 'rejection'. Machlup agreed with Hutchison that testing is important. However, whereas Hutchison wanted to test all statements, Machlup restricted this to specific assumptions and low-level hypotheses. And whereas Hutchison was after verification, Machlup sought confirmation.

A contrasting perspective on the status of the rationality assumption in economics came from Milton Friedman (1953), who argued that assumptions are largely irrelevant to the validation of theories. Instead, the latter should be judged, in Friedman's opinion, almost solely in terms of their instrumental value in generating accurate predictions. He did consider other criteria besides valid and meaningful predictions, but these were subsidiary. They included logical consistency, categories with meaningful empirical counterparts, advancing a substantive hypothesis capable of being tested, and simplicity and fruitfulness. Friedman argued that the standard theory in economics is successful due to its countless applications. He further claimed that the positive record follows from the dynamics of competition over time. In his opinion, the role of assumptions such as rationality is limited. They specify the conditions of validity, but do not determine these. They offer an economical mode of describing or presenting a theory. And indirect evidence may follow if assumptions are the implications of related hypotheses. In a similar vein, Armen Alchian (1950) had argued that individuals who act in a rational fashion will be successful and 'selected' for survival by the economic system.

Herbert Simon (1963) endeavored to rescue interest in the rationality assumption in economics by criticizing Friedman's so-called principle of unreality. According to Simon, one cannot use the validity of the market level to support the actor level. Instead, economists need to explain the market level through the use of the actor level. In Simon's opinion, valid theories about the market level follow from empirically valid assumptions about actors together with empirically valid composition laws. He therefore suggested the so-called principle of continuity of approximation instead. This holds that, if the conditions of the real world approximate sufficiently well the assumptions of an ideal type, the derivations from these assumptions will be approximately correct. Simon argued that the unreality of premises is not a virtue but a necessary evil – a concession to the finite computing agency of the scientist that is made tolerable by the principle of continuity of approximation.

We return to Simon's alternative when we follow the historical narrative to the present towards the end of this article. We now look at methodological concerns associated with the various historical interpretations of rationality.

Rationality as Self-Interest

If rationality is interpreted in terms of self-interest, one of the questions that arises concerns the status of norms (Elster 1989). Are norms rationalizations of selfinterest? No, because some norms override self-interest. And norms need to have some kind of grip to be manipulated. Are norms followed out of self-interest? No, because norms do not need external sanctions. Moreover, some sanctions are performed for other motives. Do norms exist to promote self-interest? No, because followers of norms abide by them even when it is not in their interest to do so. Do norms exist to promote common interests? No, because not all norms are Paretoimprovements. In addition, some norms that would be Pareto-improvements are not observed. Do norms exist to promote genetic fitness? No, because self-interest and fear of sanctions do not provide the full explanation for adherence to norms. And we need to study emotions, envy, honour, and conformism. Additional questions arise with the later interpretations of rationality.

Rational Choice and Expected Utility

Within expected utility theory, rationality was associated with (a) subjective probability, (b) Bayesian learning, and (c) maximization of expected utility (Sugden 1991). In this interpretation, preferences are revealed by choice, and choices are supported by reasons. Efforts have been made to develop philosophical foundations of expected utility theory by appealing to Hume's instrumental rationality. According to the latter, actions can be motivated only by desires, and no desire can be brought into existence by reason alone. That is, reason is an instrument for achieving ends that are not themselves given by reason. However, there are two problems when it comes to linking expected utility theory with Hume's instrumental rationality. First, determinacy is not implied by Hume's theory of motivation. Second, consistency also does not follow from it. That is, the axioms associated with expected utility are much stronger than instrumental rationality. First, there is no justification for the completeness of preferences presumed within expected utility theory. Evidence for this can be found in framing effects, according to which the alternative framing of information in positive or negative terms affects judgments and decisions.

Second, there is no justification for transitivity and sure-thing axioms due to the restricted interpretation of consequences. Evidence for this can be found in regret theory, according to which people take anticipated regret into account when they decide, which probably makes them loss averse.

Some have criticized the rational choice interpretation of rationality for presuming economic agents to be rational fools due to the severe constraints on the nature of the models that can be admitted into analysis (Sen 1977). On the one hand, rational choice presumes too little. This is because choice may reflect a compromise among a variety of considerations of which personal welfare may be just one. Rational choice has further come under attack for circularity because behaviour is explained in terms of preferences defined by behaviour. And it has been criticized for having too little structure; that is, it does not consider sympathy and commitment. With sympathy, which is egoistic, concern for others directly affects one's own welfare, which can be seen as an externality that would upset some standard results. With commitment, which is non-egoistic, concern for others does not affect one's own welfare but does cause action. Such action can be seen as involving counter-preferential choice requiring reformulation of the economic models, since personal choice can not longer be equated with personal welfare. In response, it has been argued that commitment needs to be accommodated as a part of behaviour by considering meta-rankings of preference rankings to express our moral judgments.

Similar concerns arise as a result of a wide range of impossibility results, such as Arrow's impossibility theorem, according to which supra-individual entities such as societies and nations cannot be said to have well-behaved preferences of the sort attributed to individual agents in rational choice approaches, under fairly general circumstances (Arrow 1987). This devastated the hope that statements about collectivities could have solid microfoundations in individual rationality. We will return to ethical and justice matters after taking a closer look at the appeals to rationality within game theory.

Strategic Rationality

Nash equilibrium, the basis for much game theory, goes further in assuming not only rationality but also common knowledge of rationality (Sugden 1991). That is, it presumes that there is common knowledge of the mathematical description of the game, of the rationality of the players, and of the logical or mathematical theorems. However, common knowledge of rationality is not sufficient for Nash equilibrium. In addition, it is incoherent since it requires subjective probabilities to be formed, which may not be possible. Moreover, it is circular, though it establishes internal consistency of the infinite chain of reasoning, since it cannot explain choice because the outcome is not determinate. As a result, common knowledge of rationality should be seen as an equilibrium concept, where equilibrium may not exist.

Arguments have been made that the rationality associated with Nash equilibrium is self-defeating (Sent 2004a). That is, all kind of frictions have been encountered within the Nash program. First, the folk theorem illustrates the (very real) possibility of encountering multiple equilibria in repeated games. The folk theorem states that in infinitely repeated games, for a range of discount factors that are high enough – though less than 1 – any payoff vector that is feasible in the set of payoffs between two players who are simultaneously individually rational is a Nash equilibrium payoff. Second, intuitively unreasonable equilibria may be selected in the finitely repeated Prisoner's Dilemma game, the chain store paradox, and the centipede game. As a result, the standard game-theoretic solutions yield results that are considered quite unintuitive. Finally, under certain conditions, theorems concerning the non-existence of trade and the impossibility of 'agreeing to disagree' about an event have been proved for Nash equilibria. Moreover, speculative trade cannot be explained as an outcome of different information structures. One possible resolution is to disconnect Nash equilibrium and common knowledge of rationality. It has been shown that common knowledge could generate many non-Nash equilibria. Likewise, it has been shown that even with common knowledge of rationality there may be no Nash equilibrium. Overall, however, the foundations of theories associated with rationality are not secure.

Rational Expectations

In recent economics, there has been much concern with how expectations are formed. The dominant approach argues that the rules of thumb, such as adaptive expectations, which fail to use available information optimally are hard to reconcile with the idea of rationality that was the foundation of most economic analysis (Muth 1961). Instead, it is argued that, since agents are claimed to be optimizers, it is only natural to presume that they will also form their expectations rationally. Hence, some argue that the rational expectations hypothesis is nothing but a direct application of the rationality principle to the problem of expectations of future events. In particular, optimizing over perceptions implies that agents do the best they can and form their views of the future by taking account of all available information, including their understanding of how the economy works. If perceptions were not optimally chosen, there would exist unexploited utility or profit-generating possibilities within the system. The implication is that all such unexploited possibilities must disappear. When applied to macroeconomics, this appeared to be in sharp contrast with Keynesian theories, which modelled firms and consumers in ways that were seen as being ad hoc and inconsistent with the idea of rational behaviour. The typical Keynesian assumptions that markets did not clear and that economic agents did not always pursue optimizing strategies could be criticized on similar grounds, as implying ad hoc departures from the axiom of rational behaviour. From this perspective, to adopt rational expectations is thus to replace earlier ad hoc treatments with an approach squarely based on the microfoundations of incentives, information, and optimization.

A variety of problems have arisen within rational expectations economics as a result of its rationality assumption (Sent 1997). First, how can there be trade among economic agents who are all rational? One suggestion, following a line of research started by Robert Lucas, is that equilibrium probability beliefs differ and that agents actually trade on the basis of different information. However, a whole series of notrade theorems overrule this common-sense intuition (Varian 1987). The second obstacle encountered by rational expectations economists involved error-term justification. In particular, close scrutiny of the justification of error terms revealed that the econometrician needed to be outwitted by the agents (Sargent 1981). Finally, how can policy recommendations be made when agents, economists, and governments are put on an equal footing based on rational expectations? When policy recommendations are possible, symmetry is impossible. For making recommendations for improving policy amounts to assuming that in the historical period the system was not really in a rational equilibrium. When symmetry is possible, policy recommendations are impossible. For making the assumption that in the historical period the system was in a rational equilibrium raises the question of why we study a system that we cannot influence (Sargent 1984).

Having considered methodological concerns associated with the various historical interpretations of rationality, we now take up ethical concerns explicitly, since they bear upon rationality in general.

Ethics

Some have criticized economists for focusing narrowly on rationality while ignoring ethics (Hausman and McPherson 1984). They claim that ethics is relevant to economists for a variety of reasons. First, economists need to know some morality to know what questions to ask. Second, economists evaluate moral commitments while describing them. Third, economists affect what they see by how they describe it. Fourth, economists are influenced by their moral values and their attitudes towards the values of the agents they study. Hence, ethics is part and parcel of economics, even when economists fail to acknowledge this in their focus on rationality.

R

An additional concern is that they oftentimes mistakenly identify well-being with preference satisfaction. This is problematic for a variety of reasons. First, what people prefer may not be good for them. Second, people make mistakes. Third, people may prefer to sacrifice their own well-being in pursuit of some other end. As a result, appraisals of economic institutions and outcomes must consider moral concerns, such as freedom, rights, and justice, besides rationality.

Rationality, which is central to the notion of a competitive equilibrium, is a key element in the two fundamental theorems of welfare economics. The first states that any competitive equilibrium leads to an efficient allocation of resources. The second asserts the converse, that any efficient allocation can be sustainable by a competitive equilibrium. The first theorem appears to make a case for non-intervention: let the markets do the work and the outcome will be desirable. The second theorem states that out of all possible efficient outcomes (of which there may be many) one can achieve any particular efficient outcome by enacting a lump-sum wealth redistribution and then letting the market take over. It has been argued that perfectly competitive markets with individual factor endowments and private goods, free market activity and mutual concern, and the absence of rationality are morally free zones (Gauthier 1991). Because of the free activity, there is liberty. Due to the absence of externalities, there is impartiality. And as a result of the first welfare theorem, there is optimality. This morally free zone, the argument goes, arises in a deeper moral framework, according to which moral constraint is compatible with mutual unconcern and rationally required. In addition, morality as a system of rationally required constraints makes possible the realization of one's interests and the fulfilment of one's preferences. This perspective has been criticized on four accounts. First, there are market failures. Second, the initial distribution is relevant. Third, there may be multiple equilibria. Fourth, no account of social policy is given. Instead, it has been claimed that markets are political, cultural and economic. First, they support a well-defined structure of power. Second, they shape our culture. Third, they foster or thwart desirable forms of human development. Fourth, they allocate resources and distribute income. Since rationality is only one element in these arguments, we shall not dwell on these concerns further.

Having considered ethical concerns explicitly, we now turn to efforts to test rationality, since they also bear upon rationality in general.

Testing Rationality

As cautioned in the introduction of this article, efforts to test rationality interpret the notion in a descriptive manner. With this in mind, much energy has also been put into trying to test rationality directly (Blaug 1992). According to some, this involves stating the problem situation, testing the predictions, assessing the evidence, considering the nature of explanation, considering alternative theories, and stating the hard core and heuristics. However, these steps are more difficult than appears at first sight. As philosophers of science have argued, immunizing stratagems are sometimes defensible, verisimilitude is difficult to implement, and there is no metric of corroboration. Economics poses additional problems, because narrow falsificationism is too restrictive and broad falsificationism has no prescriptive force. Moreover, economics is characterized by many initial conditions and no general laws. Testing its models is not the same as testing theories. Its data do not correspond to the concepts. Finally, falsificationism is hardly ever practiced in economics. Indeed, in response to these problems with the Popperian position, Popper himself accorded a special status to the rationality principle within his situational logic as a 'zero principle'. This situational logic is used to explain actions and events in social science. According to Popper, the rationality principle is an integral part of every, or nearly every, testable social theory. At the same time, he believed there to be good reasons for the rationality principle to be false, while a good approximation to truth. Still, he felt that social scientists should retain it despite the fact that it is false. This is because he believed that we learn more if we blame our situational model. Indeed, he saw the policy of upholding the rationality principle as part of our methodology.

After this methodological 'detour', we now return to the historical narrative by evaluating recent efforts to replace rationality with the notion of bounded rationality.

Bounded Rationality

Recently, especially game theorists and rational expectations economists have embraced the notion of Bounded rationality. Game theorists have looked towards Bounded rationality in their efforts to save the rationality of the Nash equilibrium. This was needed because of frictions within the Nash program (Aumann 1997; Rubinstein 1998). First, the folk theorem illustrates the (very real) possibility of encountering multiple equilibria in repeated games. Second, intuitively unreasonable equilibria may be selected in the finitely repeated Prisoner's Dilemma game, the chain store paradox, and the centipede game. Finally, under certain conditions, theorems concerning the non-existence of trade and the impossibility of 'agreeing to disagree' about an event have been proved for Nash equilibria. It could be argued that one response of game theorists to these problems has been to incorporate Bounded rationality (Sent 2004a). First, Bounded rationality functioned as a dynamic for selection among multiple equilibria by promising to 'refine' equilibria. Moreover, the evolutionary stable strategy concept of evolutionary game theory may be viewed as a further refinement of perfect equilibrium, one of the most common notions used to refine the Nash equilibrium. Second, Bounded rationality has been used to rule out unintuitive equilibria in the Prisoner's Dilemma game, the chain store paradox, and the centipede game. Third, absence of a fully rational treatment of knowledge may circumvent the no-trade theorems by allowing speculative trade. These attempts to strengthen Nash, then, lead to the paradoxical observation that rationality in games depends critically on irrationality.

Likewise, rational expectations economists have sought to reinforce the rational expectations hypothesis by focusing on convergence to this equilibrium through boundedly rational 'learning'. They have also used Bounded rationality to deal with some of the problems associated with rational expectations such as multiple equilibria and the computation of equilibria (Sargent 1993).

Economists have tried to capture Bounded rationality by replacing rational players with computing devices such as Turing machines, finite automata, or neural network algorithms. Players' rationality is bounded in the sense that they cannot consider strategies other than those that can be played by these computing devices. Rationality is bounded in Turing machines because these machines will sometimes compute for ever to give correct answers. In order to come up with a solution, the output follows an arbitrary guessing rule after the machine has been stopped. Bounded rationality in finite automata is captured by imposing constraints on the number of states of the automata or assuming that states are costly. Neural networks, finally, increase the computational capability of a finite automaton by increasing the states of the machine.

Both game theorists and macroeconomists have developed models of boundedly rational learning (for example, Bray and Kreps 1987). The basic idea in these models is that boundedly rational agents utilize one of three procedures for making and changing their choices on the basis of past outcomes. First, Bayesian learning assumes that players update their subjective probabilities in the face of inconsistencies through the use of Bayes' rule until consistency is achieved. This technique has been used in the problem of equilibrium selection. It has been criticized for not accurately representing 'true' learning. Second, least squares or adaptive control learning assumes that players use standard statistical or econometric procedures for estimation. These procedures are boundedly rational in that economic agents use models that are misspecified and forecasting procedures that are not part of the optimal decision making of these individuals. This approach has been used in the context of rational expectations models. It has been criticized for requiring the agents to still be quite smart. Third, neural network learning assumes that individuals construct explicitly approximate models of their environment which are updated as their information improves. In contrast to least squares or adaptive control learning, agents here know they hold misspecified models of reality. This technique has been used to explore how boundedly rational players can achieve consistent beliefs and, possibly, a Nash equilibrium.

Behavioural Economics

Developments within game theory and rational expectations economics combined with the rise of behavioural economics suggest that economics as a whole is moving away from rationality and towards Bounded rationality (Sent 2004b). In the 1960s, appeals to Bounded rationality on the part of behavioural economists were designed to develop an alternative to the mainstream model (for example, Simon 1955). We could label these endeavours as old behavioural economics. At that time, few economists exhibited any interest in these efforts. In the 1970s, cognitive psychologists suggested ways to incorporate behavioural insights in ways that provided less of a threat to the standard model (for example, Kahneman and Tversky 1974). We could label these efforts as new behavioural economics. At the same time, the mathematical foundations of the mainstream started showing some flaws. In the 1980s, disagreements emerged between old and new behavioural economists with the latter emerging as the victors in the 1990s, partly because Herbert Simon abandoned his efforts and partly because new behavioural economists suggested ways in which their insights may help rebuild the mainstream stronghold.

Bounded rationality is not a field in itself, but rather an approach to doing economic research (Simon 1976). There are many interpretations of Bounded rationality and these are not always consistent. Herbert Simon, the father of Bounded rationality, used the term 'Bounded rationality' to highlight limitations of both knowledge and computational capacity. These bounds affect human cognitive capacity for discovering alternatives, computing their consequences under certainty and uncertainty, and making comparisons among them. Theories of Bounded rationality, then, are generated by analysing processes for generating alternatives, procedures such as heuristics for evaluating consequences, and strategies such as satisficing for making choices. Decision making is characterized as a selective search in which heuristics are used to determine what paths should be taken, and the search halts when a satisfactory solution has been found. In this process, aspiration levels are adapted in response to success or failure.

To distinguish behavioural economics from neoclassical economics, Simon introduced a distinction between procedural and substantive rationality (Simon 1976). He argued that psychologists have considered the former concept while economists have focused on the latter. Whereas the former involves decision makers following specific rules or procedures, the latter has decision makers consider their own total preference ordering. That is, procedural rationality is about the rationality of the procedure used to reach a decision, while substantive rationality is about the rationality of the decision itself. While procedural rationality is interested in how individuals make decisions, substantive rationality focuses attention on why they do so. That is, the former focuses on the methods individuals employ, whereas the latter considers the outcomes that follow. On the one hand, procedural rationality helps decision makers decide how to get there, but not where to go. On the other hand, substantive rationality tells them where to go, but not how to get there. Simon used the concept of Bounded rationality to explain why substantive rationality is often inappropriate as well as impossible.

While models of Bounded rationality have not always appeared as attractive as the axiomatized certainties of neoclassical economics, more and more economists are embracing one form or another of Bounded rationality. Models of Bounded rationality owe their revival partly to attempts to develop a viable alternative to neoclassical economics and partly to attempts to strengthen neoclassical economics. Whereas the first reason is in the spirit of Simon's contributions, Simon certainly would have opposed the second.

Conclusion

In adopting Bounded rationality, rational expectations economists and game theorists found themselves in the paradoxical position of using Bounded rationality to define rationality. In addition, they were confronted with the question how much Bounded rationality is admitted and how it is clarified. In fact, the dependence of the definition of rationality on irrationality is reminiscent of debates in philosophy concerning the definition of concepts in terms of their opposites, which has led to efforts to destabilize dichotomies. In addition, there is the practical challenge of overcoming the Bounded rationality of economists in modelling the Bounded rationality of economic agents. Ironically, when economists made the agents in their models more bounded in their rationality, they had to be smarter because these models became larger and more demanding econometrically. Bounded rationality researchers face innumerable decisions about how to represent decision-making processes and the ways that they are updated, which requires a large amount of rationality on their part.

Overall, the efforts of rational expectations economists and game theorists to embrace one form or another of Bounded rationality have more to do with intellectual puzzles than empirical anomalies (Aumann 1997; Rubinstein 1998; Sargent 1993). Hence, in effect, economists have replaced one set of puzzles, concerning, for instance, the non-existence of trade, with another paradox, concerning the dependence of rationality on irrationality (Aumann and Sorin 1989).

The move from rationality to Bounded rationality is part of the observation that the present situation in economics could be characterized as one of moderate pluralism. That is, recent years have witnessed not only efforts to incorporate Bounded rationality approaches and behavioural insights, but also chaos theory, complexity approaches, evolutionary insights, experimental methods, and neuroeconomics. Therefore, the benchmark from which new behavioural economics considers deviations may itself be evolving – none of which alone or in combination appears yet to have established a new orthodoxy, leading to debate over the direction and future content of economics. Elaborating these, however, interesting, would take us beyond the scope and word limit of this article.

See Also

- Behavioural Economics and Game Theory
- Rationality, Bounded
- Ethics and Economics
- Rational Expectations
- Simon, Herbert A. (1916–2001)

Bibliography

Alchian, A.A. 1950. Uncertainty, evolution and economic theory. *Journal of Political Economy* 58: 211–222.

Arrow, K.J. 1987. Economic theory and the hypothesis of rationality. In *The New Palgrave: A dictionary of economics*, vol. 1, ed. J. Eatwell, M. Milgate, and P. Newman. New York: W.W. Norton.

Aumann, R.J. 1997. Rationality and bounded rationality. *Games and Economic Behavior* 21: 2–14.

Aumann, R.J., and S. Sorin. 1989. Cooperation and bounded recall. *Games and Economic Behavior* 1: 5–39.

Becker, G.S. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70: 1–13.

Blaug, M. 1992. *The methodology of economics: Or, how economists explain*. Cambridge: Cambridge University Press.

Bowles, S. 1991. What markets can – And cannot – Do. *Challenge* 34(4): 11–16.

Bray, M., and D. Kreps. 1987. Rational learning and rational expectations. In *Arrow and the ascent of modern economic theory*, ed. G. Feiwel. New York: New York University Press.

Brunswik, E. 1955. In defense of probabilistic functionalism: A reply. *Psychological Review* 62: 236–242.

Caldwell, B.J. 1994. *Beyond positivism: Economic methodology in the twentieth century*. London: Routledge.

Daston, L. 1983. Rational individuals versus laws of society. In *Probability since 1800*, ed. M. Heidelberger, L. Krüger, and R. Rheinwald. Bielefeld: Universität Bielefeld.

Elster, J. 1989. Social norms and economic theory. *Journal of Economic Perspectives* 3(4): 99–117.

Friedman, M. 1953. The methodology of positive economics. Reprinted in Hausman (1984).

Gauthier, D. 1991. *Morals by agreement*. New York: Cambridge University Press.

Gigerenzer, G., and P.M. Todd. 1999. Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics*

R

That make us smart, ed. G. Gigerenzer, P.M. Todd, and A.R. Group. New York: Oxford University Press.
Hausman, D. (ed.). 1984. *The philosophy of economics*. Cambridge: Cambridge University Press.
Hausman, D., and M. McPherson. 1984. Economics, rationality, and ethics. In Hausman (1984).
Hutchison, T.W. 1956. On verification in economics. Reprinted in Hausman (1984).
Kahneman, D., and A. Tversky. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.
Kirzner, I.M. 1962. Rational action and economic theory. *Journal of Political Economy* 70: 380–385.
Knight, F.H. 1925. Fact and metaphysics in economic psychology. *American Economic Review* 15: 247–266.
Machlup, F. 1956. On indirect verification. Reprinted in Hausman (1984).
Mill, J.S. 1836. On the definition and method of political economy. Reprinted in Hausman (1984).
Muth, J.F. 1961. Rational expectations and the theory of price movements. *Econometrica* 29: 315–355.
Neumann, J. von., and O. Morgenstern. 1944/1955. *Theory of games and economic behavior*. Princeton: Princeton University Press.
Peterson, C.R., and L.R. Beach. 1967. Man as an intuitive statistician. *Psychological Bulletin* 68: 29–46.
Robbins, L. 1935. The nature and significance of economic science. Reprinted in Hausman (1984).
Rubinstein, A. 1998. *Modeling bounded rationality*. Cambridge, MA: MIT Press.
Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.
Sargent, T.J. 1981. Interpreting economic time series. *Journal of Political Economy* 89: 213–248.
Sargent, T.J. 1984. Autoregressions, expectations, and advice (with discussion). *American Economic Review* 74: 408–421.
Sargent, T.J. 1993. *Bounded rationality in macroeconomics*. Oxford: Oxford University Press.
Savage, L.J. 1954. *The foundations of statistics*. New York: John Wiley & Sons.
Sen, A. 1977. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* 6: 317–344.
Sen, A. 1987. Rational behavior. In *The New Palgrave: A dictionary of economics*, vol. 4, ed. J. Eatwell, M. Milgate, and P. Newman. New York: W.W. Norton.
Sent, E.M. 1997. *The evolving rationality of rational expectations: An assessment of Thomas Sargent's achievements*. Cambridge: Cambridge University Press.
Sent, E.M. 2004a. The legacy of Herbert Simon in game theory. *Journal of Economic Behavior and Organization* 53: 303–317.
Sent, E.M. 2004b. Behavioral economics: How psychology made its (limited) way back into economics. *History of Political Economy* 36: 735–760.
Simon, H.A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69: 99–118.
Simon, H.A. 1976. From substantive to procedural rationality. In *Method and appraisal in economics*, ed. S. Latsis. Cambridge: Cambridge University Press.
Simon, H.A. 1963. Testability and approximation. Reprinted in Hausman (1984).
Smith, V. 1991. Rational choice. *Journal of Political Economy* 99: 877–897.
Smith, V. 1992. Game theory and experimental economics. In *Toward a history of game theory*, ed. E.R. Weintraub. Durham: Duke University Press.
Sugden, R. 1991. Rational choice: A survey. *Economic Journal* 101: 751–783.
Suppes, P. 1961. The philosophical relevance of decision theory. *Journal of Philosophy* 58: 605–614.
Tisdell, C. 1975. Concepts of rationality in economics. *Philosophy of Social Science* 5: 259–272.
Tversky, A., and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76: 105–110.
Varian, H.R. 1987. Differences of opinion in financial markets. In *Financial risk: Theory, evidence, and implications, Proceedings of the 11th annual economic policy conference of the Federal Reserve Bank of St. Louis*.
Weber, M. 1904/1999. The area of economics, economic theory, and the ideal type. In *Essays in economic sociology*, ed. M. Weber. Princeton: Princeton University Press, 1999.
Winter, S.G. 1964. Economic 'natural selection' and the theory of the firm. *Yale Economic Essays* 4: 225–272.

Rationalization of Industry

K. D. George

Rationalization of industry is a term used to describe the changes, usually quite drastic, that are needed to correct a position of fundamental disequilibrium in an industry. The case most commonly associated with the problem is that of adjustments needed to correct overcapacity, but also involved are adjustments to correct substantial cost differences between firms arising from differences in technology or from variations in the efficiency with which a given technology is used.

We start with the case of overcapacity and ask why this state of affairs arose in the first place. There are several possible reasons which will be dealt with briefly because the main concern of this essay is with corrective mechanisms. First, overcapacity may have resulted from a fall in domestic demand brought about by a fall in world demand, or by a fall in the price of imports, or by a switch in demand to a close substitute. Second, it may be due to government policies aimed at stimulating production. Third, it may have arisen because the competitive process resulted in an overexpansion of capacity during a period of increasing demand. There is no convincing theoretical or empirical evidence to suggest that this is more likely to result in one type of industry structure than another. In perfectly competitive markets the problem hinges around the fact that, following an increase in demand, each firm bases its expansion plans on current market prices which are assumed to stay constant. In relation to long-run equilibrium these prices are too high in the case of output and too low in respect of inputs. If all firms are equally assiduous in spotting profitable expansion opportunities the result will be an overexpansion of capacity. This result may be avoided or at least ameliorated if there are speculators who have more accurate information than producers about the course of future prices so that speculative selling of the product reduces the extent of short-run price increases. The same result may obtain if producers hold large stocks which can be used to meet part of the increased demand. More important however is the likelihood that firms' capacity extensions will be staggered so that some bring additional output on to the market, thus depressing the price, before others have put their investment plans into effect. In oligopolistic industries a key factor is the importance that firms attach to customer goodwill and in particular the ability to meet orders promptly. As a result an important aspect of oligopolistic competition during an upswing is competition in capacity extensions. This tendency may be accentuated if capacity extensions are also used as a barrier to entry.

Whatever the reasons for the emergence of over-capacity can it not be left to market forces to correct the imbalance? Market forces will indeed tend to act as a corrective mechanism but intervention may yield a better outcome. This is because of two interrelated factors: the speed with which market forces work, and divergences between private and social costs and benefits.

From the point of view of maximizing the level of output in the economy market forces may, depending on circumstances, operate either too slowly or too quickly – too slowly, because in some industries the long run may be considerably longer for downward adjustments to the capital stock than for additions to it. In industries with high fixed costs prices may have to fall to a very low level before they fail to cover the operating costs of even the most inefficient firm. Typically such aggressive price competition will not occur in oligopolistic industries except perhaps in a severe and prolonged recession. The effectiveness of price competition in eliminating inefficient firms is further reduced if each firm has built up its own clientele. A fall in demand is therefore likely to affect all firms in an industry rather than be concentrated on the most inefficient. Even if aggressive price cutting does occur this does not necessarily mean the elimination of the tail of inefficient firms only; the outcome will also be affected by such factors as the financial resources that each firm has when the slump in demand occurs, and their degree of vertical integration and diversification. Finally, even if price competition does succeed in eliminating inefficient firms this does not necessarily mean any reduction in productive capacity since the plant of a firm that leaves an industry may be acquired by one of the survivors. If for these and other reasons it is indeed the case that resources are released very slowly from declining industries, and if these resources could be more profitably used elsewhere, there is a strong case for a planned rationalization scheme possibly backed by government financial inducements to speed up the process of adjustment.

In other circumstances, however, the optimum policy will be to slow down the process of adjustment. The arguments for contracting an industry are based on existing and expected profitability. Profit is the difference between the value of output produced and the cost of inputs used in production. Its use as an indicator of efficiency is valid if the price of a commodity reflects society's valuation of it (relative to other commodities) and if the cost of inputs reflects the social value of those inputs. When these conditions hold an enterprise making a profit is creating goods whose value is greater than the social value of the inputs used. If, therefore, the government intervenes to support an unprofitable industry it is destroying social value because the labour that is being kept in that industry could be more efficiently used elsewhere in the economy. The argument can be extended to cover competing imports. At an equilibrium exchange rate the price of imports is the value of domestically produced goods that have to be exported to pay for these imports. If the cost of an imported good is less than that of the domestically produced equivalent then the most efficient way of obtaining that good is to import it in exchange for other domestically produced goods.

The above analysis is wholly dependent on the assumptions made. If, for instance, wage rates are rigid in the industry requiring rationalization, contraction of the industry will result in unemployment. The labour market mechanism by which workers, especially those with specific skills, are re-employed elsewhere may work very slowly. There is thus a trade-off between the immediate loss of output resulting from the unemployment and the longer run benefits of transferring workers to more profitable activities. The optimal solution will involve some form of intervention that slows down the rate of decline. More generally, when rationalization leads to unemployment wages are no longer an accurate measure of the social value of labour; i.e the opportunity cost of labour, or the 'shadow wage', is below the market wage.

The case for slowing down the process of contraction is even stronger where the redundancies resulting from a rationalization programme are added to an already high level of unemployment. In these conditions the opportunity cost of certain types of labour may be zero, i.e. the employment of this labour is not preventing more efficient production taking place elsewhere in the economy. Furthermore, additional redundancies have a multiplier effect arising partly from the disruption of backward and forward linkages and partly from the reduced purchasing power available to the newly redundant.

Failure to appreciate the relevance of these considerations is all too common. For instance, in spite of a high level of unemployment and the low shadow wage of coal miners, a substantial programme of rationalization involving the closure of high cost pits was pursued in Britain in 1985 and 1986 on the basis of accounting rates of return. This was said to be necessary to adjust the industry to a lower level of demand for domestically produced coal and to make way for new investment. In the circumstances of this particular case, however, there are strong arguments which suggest that government intervention would have been economically justified to slow down the rate of contraction.

A theoretical issue of general importance in the economics of the rationalization of industry is the legitimacy of using a shadow wage in decisions relating to one particular industry. The use of a shadow wage has been challenged on a number of counts: (i) the shadow wage argument can only be applied to a handful of special cases, otherwise nobody would be paying taxes for common services from which they derive benefit; (ii) if the shadow wage argument were generalized the effect would be to slow down the rate of structural change; (iii) why should any one industry be singled out for special treatment?

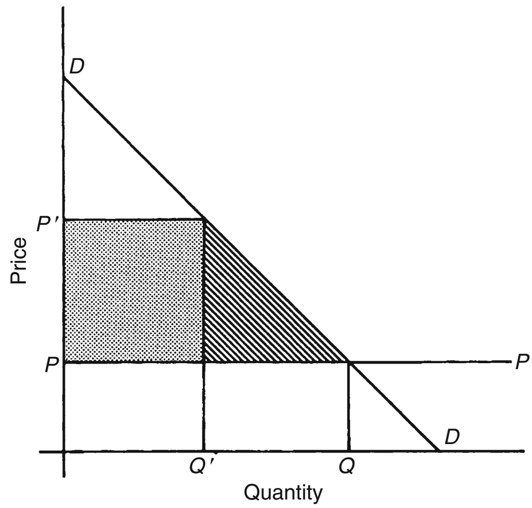
Objections (i) and (ii) are based on a misunderstanding of the rationale of shadow pricing. The purpose of using a shadow wage in policy decisions is to correct for the existence of unemployment and to minimize the loss of output associated with it. Where there is large-scale unemployment there is no reason whatever why the shadow wage argument should not be applied across the board. In practical terms this would involve the use of a general labour subsidy the level of which would fall as the level of unemployment came down. The suggestion that structural change is impeded by shortage of labour at a time of high unemployment is implausible. Objection (iii) is a legitimate query, and indeed on grounds of efficiency a general labour subsidy is preferable to subsidizing high-cost units in a particular industry. However, in the absence of effective macroeconomic policies to bring down the level of unemployment recourse to shadow wages in closure decisions in a particular industry is justifiable on secondbest grounds, particularly where that industry is concentrated geographically.

Government intervention in the rationalization of industry can thus be justified either to speed up the process or to slow it down. But the form of intervention is also important, a point that is well illustrated by reference to European attempts to rationalize agriculture and steel.

In agriculture, massive overproduction of products such as milk and cereals has been due to high price guarantees combined with technological advances. After half-hearted attempts to stem the flow of milk the European Commission in 1983 spelt out the options for a serious attack on the problem: either a price reduction or production quotas with a levy on excess production. In the event and largely for political reasons, quotas were introduced. The preference for quotas over price reductions has been justified on grounds of fairness: whereas price reductions would affect all producers a levy on excess production would affect only those producing the excess and in proportion to their contribution to overproduction. The most likely reason for choosing quotas however was that farmers, who traditionally dislike price cuts, may not readily have appreciated that quotas imply a hidden price reduction; the prices being a weighted average of the guaranteed price on the quota and the disposal price on excess deliveries.

Quotas, however, are an inefficient way of reducing overcapacity. The uneven treatment of producers is actually perverse, preventing efficient farms from expanding while encouraging inefficient farms to maintain production. A straightforward price cut on the other hand would be more easily absorbed by efficient farmers and more discouraging to the inefficient. The anomaly would also be avoided of farmers who had invested heavily in modern dairying facilities not having a quota because they were not in production on day one of the new regime. In addition quotas are administratively cumbersome as compared to a straightforward price reduction and they operate solely on the supply side of the market whereas a price cut would also stimulate demand.

Some of the drawbacks of quotas would be overcome if they were freely marketable, but there would still be a problem of equity. Those who are awarded a large quota are in effect given a valuable capital asset so that the system implies an arbitrary redistribution of wealth within the farm sector. The situation is most unfair on the tenant farmer who once he ceases milk production has no right over the quota, the entire capital being vested in the landowner.



Rationalization of Industry, Fig. 1

The problems of overproduction in European agriculture would be more efficiently tackled by a gradual relaxation of price controls, with the farm sector supported by means of production subsidies financed out of taxation.

This example of the attempt to rationalize milk production in the European Economic Community draws attention to a general problem in rationalization schemes; that of allocating output between firms with different cost functions. Standard economic theory shows that where there are cost differences it is generally possible for a rationalization cartel to reallocate output so as to increase total industry profits. In the diagram, (Fig. 1) *D* is the industry demand curve and *MR* industry marginal revenue. There are two firms, *A* and *B*, each with demand curve *d* and marginal revenue curve *mr*. The profit maximizing price for firm *A* is P_A and for *B*, P_B with outputs of *A* and *B* respectively. Industry profits are maximized where $\sigma MC = MR$, with market shares of A^* and B^* which equalizes the marginal costs of the

R

two firms. This however requires that *A*, the high cost firm, accepts a big reduction in market share and in the profits on its own sales. There is clearly no incentive for *A* to go along with the rationalization scheme unless there is a profit-pooling arrangement and side payments which make it better off than it was before the rationalization scheme was introduced.

However even if a rationalization cartel can increase total industry profits and side payments are feasible it is by no means certain that high-cost firms will participate voluntarily. These firms will fear that by agreeing to a smaller market share they will suffer a loss of bargaining power and perhaps be even worse off than they would have been without cartelization. And once quotas have been agreed it can prove extraordinarily difficult for a firm to improve its market share through the more efficient use of existing assets or through modernization of plant and equipment. Indeed one of the problems often encountered by cartels is that of matching overall capacity to demand without at the same time unduly inhibiting costreducing investments.

These problems have been encountered for instance within the European Coal and Steel Corporation (ECSC). An ECSC Plan to establish a voluntary European cartel in 1977 broke down in a price war. To restore order in the market the ECSC declared a 'manifest crisis', using its powers to compel adherence to quotas, and setting minimum prices for a range of steel products. This was very similar to the method of intervention in agriculture and is subject to the same criticism: firms which have made efforts to improve efficiency are penalized if they are unable to translate this into a larger market share. In its 1984/85 Annual Report the British Steel Corporation complained that, 'with demand for strip mill products markedly higher in the UK than elsewhere in the EEC, it was difficult to secure adequate quota to meet normal UK market share without incurring unacceptably high fines . . . the quota position remains very unsatisfactory'.

Whatever the theoretical possibilities the general experience with cartels is that they tend to cement existing industry structure rather than help in achieving efficiency-enhancing structural changes.

Where cartelization is not feasible the obvious alternative route to greater industry discipline and the benefits of rationalization is through merger. Mergers may lead to benefits of increased specialization within plants and also to larger plants with associated economies of size. In this case, as with cartels, it has to be asked why, if substantial benefits exist, are they not realized through the mechanism of competition between rival producers. if efficiency can be increased by greater specialization within plants it should pay a firm to build up sales of a smaller range of products by offering lower prices on selected brands. However, if there is a great deal of overlap in the product-range of firms this may lead to competitive price-cutting across the board leaving all firms in the end with the same product-range but with lower prices and profits. By concentrating plants into a smaller number of firms mergers may result in economies of rationalization that could not be achieved, or at least could only achieved with less certainty and over a longer period, by internal growth.

A similar argument applies to the size of plants. Existing industry capacity may match demand but many individual plants may be of sub-optimal size. Again if there are costreductions to be secured by operating an optimum-sized plant it should pay a firm to build one and drive the less efficient units out of the market. This, however, may lead to a prolonged period of severe price competition with no certainty that the investment in the optimal-sized plants will yield a normal return.

The advantages which mergers have over internal growth are speed and safety. In the face of strong competition speed may be of the essence if rationalization plans are to succeed. And where an industry has excess capacity the safety factor becomes paramount. Internal growth by creating new capacity threatens to make the situation worse, a risk that can be avoided by merger.

However, for mergers actually to result in these benefits rationalization plans have to be implemented. Increased product specialization within plants and the replacement of sub-optimal plants by optimal sized ones are not brought about without conscious management decisions. Although the opportunity for executing rationalization plans may be enhanced by mergers empirical evidence suggests that the end result is often disappointing.

The rationalization of industry has always been a factor associated with structural change and shifting comparative advange. However the scale of the problem has tended to increase with the increasing importance of international competition. In an oligopolistic grouping consisting entirely of domestic firms a degree of cohesiveness can be expected which is unlikely to be matched in a group that transcends national boundaries. If this is so the pressures to adjust will be greater. Furthermore the problem may be accentuated by direct or indirect state aid to industry. Given the existence of state aid in one country others may feel forced to follow suit, or alternatively to impose some form of control on imports. Rationalization schemes may then turn out to be a cloak for 'beggar-my-neighbour' policies with one country endeavouring to throw a larger share of the burden of adjustment on to another.

This brings us back to the interrelationship between rationalization schemes and other policies – in particular macroeconomic policies and policies such as re-training schemes aimed at increasing the mobility of labour. Government intervention to slow down the rate of contraction of a declining industry is justified if it results in a higher overall level of output. The case for such intervention is strongest when alternative employment opportunities are few, i.e. at times of recession, and when labour is immobile because of specific skills and poor retraining facilities. However, actual experience of schemes, including government-supported rationalization cartels, designed to soften the effect of market forces suggests that they often result in serious allocative inefficiency. On the other hand, the more successful government policy is in maintaining a high level of employment and an adaptable labour force the weaker is the case for policies that slow down the structural change that is the outcome of market forces, and the greater the scope for government intervention that works with market forces rather than against them.

See Also

► Cartels
► Manufacturing and De-industrialization

Bibliography

Allen, G.C. 1981. *The Japanese economy*. London: Weidenfeld & Nicolson.

Caves, R.E., and M. Uekusa. 1976. *Industrial organization in Japan*. Washington, DC: Brookings Institution.

Cowling, K., et al. 1980. *Mergers and economic performance*. Cambridge: Cambridge University Press.

Dixit, A.K. 1980. The role of investment in entry-deterrence. *Economic Journal* 90: 75–106.

Duesenberry, J.S. 1958. *Business cycles and economic growth*. New York: McGraw-Hill.

Foster, B.A., and R. Rees. 1983. The optimal rate of decline of an inefficient industry. *Journal of Public Economics* 22(2): 227–242.

George, K.D., and L. Mainwaring (eds.). 1988. *The Welsh economy*. Cardiff: University of Wales Press.

George, K.D., and C. Joll. 1981. *Industrial organisation; Competition growth and structural change*, 3rd ed. London: George Allen & Unwin.

Jacquemin, A. (ed.). 1984. *European industry: Public policy and corporate strategy*. Oxford: Clarendon Press.

Osborne, D.K. 1976. Cartel problems. *American Economic Review* 66(5): 835–844.

Pinder, J. (ed.). 1982. *National industrial strategies and the world economy*. Beckenham: Croom Helm.

Richardson, G.B. 1960. *Information and investment*. Oxford: Oxford University Press.

Scherer, F.M. 1980. *Industrial market structure and economic performance*, 2nd ed. Boston: Houghton Mifflin.

Shaw, R.W., and S.A. Shaw. 1983. Excess capacity and rationalization in the West European synthetic fibers industry. *Journal of Industrial Economics* 32(2): 149–166.

R

Rationed Equilibria

Jean-Pascal Benassy

Definition and Scope

Equilibria with rationing, also called non-Walrasian equilibria, are a wide class of equilibrium concepts which generalize the traditional notion of Walrasian equilibrium by allowing markets not to clear (in the traditional sense) and therefore quantity rationing to be experienced. Their scope is best described by examining first Walrasian equilibrium as a reference.

In a Walrasian equilibrium by definition all markets clear, that is, demand equals supply for each good. This consistency of the actions of all agents is achieved by price movements solely. No rationing is experienced as each agent is able to exchange as much as he wants at the Walrasian equilibrium price system. As noted by Arrow (1959), there is a 'missing element' in the concept, in that, whereas quantity actions by the agents result from rational behaviour, market clearing is assumed axiomatically. Non-Walrasian theory thus simply abandons this last assumption, allowing prices to be determined by other mechanisms than market clearing. An almost immediate corollary is that in order to obtain equilibrium concepts, quantity signals and quantity adjustments will have to be introduced together with price adjustments. To summarize briefly, the non-Walrasian equilibrium concept is generalized in the following directions:

(1) More general price mechanisms are considered, ranging from full rigidity to full flexibility, with intermediate forms of imperfect competition. Moreover each market may have its own price determination scheme.
(2) Quantity signals are introduced in addition to price signals. They intervene in both demand–supply and price making behaviour.
(3) Equilibrium in the short run is achieved by quantity adjustments as well as by price adjustments.
(4) Expectations about the future concern not only price signals, but quantity signals as well.

History

Equilibria with rationing have a double ancestry: Keynes (1936) because he developed (at the macroeconomic level) a concept of equilibrium where adjustment was made by quantities (the level of income) as well as by prices, and Walras (1874) because he developed a model of general equilibrium with interdependent markets. The Walrasian model has been beautifully developed into a highly elaborate and rigorous concept, notably in Hicks (1939), Arrow and Debreu (1954), Debreu (1959), Arrow and Hahn (1971).

The gap between these two lines of thought was unfortunately total until the stimulating contributions of Clower (1965) and Leijonhufvud (1968), who reinterpreted Keynesian analysis in terms of labour market rationing and quantity adjustments. These insights were included in the first fixprice macroeconomic model by Barro and Grossman (1971) (1976).

Rigorous microeconomic concepts of equilibrium with quantity rationing were then developed: Drèze (1975) built an equilibrium concept with prices variable between preset limits; Benassy (1975a, 1977b, 1982) constructed an alternative concept of fixprice equilibrium, and introduced expectations into that framework. A third concept of fixprice equilibrium was built by Younès (1975). Benassy (1976, 1977a, 1982) also developed a non-Walrasian equilibrium concept with price makers which bridged the gap with another important line of work, that dealing with general equilibrium under imperfect competition, notably associated with the name of Negishi (1961).

Still other concepts of equilibria with rationing were proposed by Glustoff (1968), Hahn (1978), Böhm and Levine (1979), Heller and Starr (1979).

We shall now describe the main concepts of the theory. In order to set the stage and introduce notation, let us first describe the economy considered and the corresponding Walrasian equilibrium concept.

The Economy and Walrasian Equilibrium

We shall describe the various concepts in the framework of an exchange economy. One good, which we shall call money, serves as numeraire, medium of exchange and reserve of value (nonmonetary exchange has been considered in Benassy 1975b). There are l markets where nonmonetary goods, indexed by $h = 1, \ldots, l$, are exchanged against money at the price p_h. Call

p the l-dimensional vector of these prices. Agents are indexed by $i = 1, \ldots, n$. Agent i has an initial endowment of good h ωih, and of money $\overline{m}_i$. Call d_{ih} his purchase of good h, s_{ih} his sale of good h. Define $z_{ih} = d_{ih} - s_{ih}$ his net purchase of good h and z_i the vector of these net purchases. His final holdings of non-monetary goods and money are respectively

$$x_i = \omega_i + z_i \quad m_i = \overline{m}_i - pz_i$$

and we shall assume that the agent has a utility function $U_i(x_i, m_i) = U_i(\omega_i + z_i, m_i)$ which we shall assume throughout strictly quasi-concave in its arguments.

Having described the economy, let us now turn to the notion of Walrasian equilibrium. As indicated above, each agent is assumed to be able to exchange as much as he wants on each market. He thus transmits demands and supplies which maximize his utility subject to the budget constraint, i.e. the Walrasian net demand function is the solution in z_i of the following programme:

$$\begin{aligned}&\text{Maximize } U_i(\omega_i + z_i, m_i) \text{ s.t.}\\ &pz_i + m_i = \overline{m}_i\end{aligned}$$

This yields a vector of Walrasian net demands $z_i(p)$ A Walrasian equilibrium price vector p^* is defined by the condition that all markets clear, i.e.:

$$\sum_{i=1}^{n} z_i(p*) = 0.$$

Transactions realized by agent i are simply equal to $z_i(p^*)$.

Equilibrium with Bounded Prices

This concept, due to Drèze (1975), develops a notion of equilibrium valid when prices are subject to inequality constraints. We shall describe here, for simplicity of exposition, the case where absolute prices are subject to limits of the form:

$$\overline{p}_h \leqslant p_k \leqslant \overline{p}_h$$

Price limits linked to a price index were considered as well in Dreèze (1975) (see also Van der Laan 1980; Dehez and Drèze 1984).

The basic idea behind this concept of equilibrium is that rationing becomes operative when prices hit one of the limits. The rationing considered will take the form of an upper bound on trades. (We shall see in the next section a possible justification for this type of rationing.) More specifically, as in Drèze (1975), consider a uniform rationing on each market, and call $\overline{d}_h$ the upper bound on purchases on market h, $\overline{s}_h$ the upper bound on sales. Net purchases of agent i on market h, z_{ih} are thus limited to the interval:

$$-\overline{s}_h \leqslant z_{ij} \leqslant \overline{d}_h.$$

An equilibrium with price rigidities 'à la Drèze' can be now defined as a set of prices p_h^* transactions z_{ih}^* and quantity constraints $\overline{d}_h$ and $\overline{s}_h$ such that:

(i) $\overline{p}_h \leqslant p_j^* \leqslant \overline{p}_h \ \forall h$

(ii) $\sum_{i=1}^{h} z_{ij}^* = 0 \ \forall h$

(iii) The vector z_i^* is solution in z_i of Maximize

$$U_i(\omega_i + z_i, m_i), \text{s.t.} \begin{cases} \{pz_i + m_i = \overline{m}_i \\ -\overline{s}_{ij} \ \leq \ z_{ij} \ \leq \ \overline{d}_{ij} \quad \forall h \end{cases}$$

(iv) $\forall h \ z_{ih}^* = \overline{d}_j$ for some i implies $z_{jh}^* > -\overline{s}_h \ \forall j$
$z_{ih}^* = \overline{d}_h$ for some i implies $z_{jh}^* < \overline{d}_h \ \forall j$

(v) $\forall h \ p_h < \overline{p}_h$ implies $z_{ih}^* < \overline{d}_h \ \forall i$
$p_h > \overline{p}_h$ implies $z_{ih}^* > -\overline{d}_h \ \forall i$

Condition (i) simply reminds us that prices are bounded upward and downward. Condition (ii) says that transactions should be consistent on every market. Condition (iii) says that transactions must be individually rational, i.e. they must maximize utility subject to the budget constraint and quantity constraints on all markets. Condition (iv) says that rationing may affect either supply or demand, but not both simultaneously. This condition is usually presented as a condition of market by market efficiency. Note also that money is never rationed. This condition is aimed at suppressing trivial equilibria where all agents

R

would be constrained to trade nothing. Condition (v) says that upward (downward) price rigidity must be binding if there is quantity rationing of demand (supply). It thus expresses in an intuitive way that quantity rationing is a substitute for price variations.

We should note at this stage that this concept contains as particular cases both a fixprice equilibrium concept (when both price limits are equal) and Walrasian equilibrium (when the lower bound is zero and the upper bound infinite).

Existence of such an equilibrium with uniform bounds on net trades was proved in Drèze (1975). The concept is easily extended to some non-uniform bounds (Grandmont and Laroque 1976; Greenberg and Müller 1979), but in this last case it is not specified in the concept how shortages are allocated among rationed demanders or rationed suppliers. We shall now study alternative concepts based on different premises which, in particular, make this more explicit. We shall now therefore study in more detail how transactions and quantity signals may be formed in a nonclearing market.

The Functioning of a Nonclearing Market

In this and the two subsequent sections we shall study other non-Walrasian concepts due to Benassy (1975a, 1976, 1977b, 1982). A basic characteristic of these models is that a clear-cut difference is made between demands and supplies on the one hand, and the resulting transactions on the other. Agents express effective demands $\tilde{d}_{ih}$ or supplies $\tilde{s}_{ih}$ which are somehow signals to the market and the other agents, and which do not necessarily match on a specific market. However, the trading process will generate transactions, i.e. purchases d^*_{ih} and sales s^*_{ih} which identically balance on each market:

$$\sum_{i=1}^{n} d^*_{ih} = \sum_{i-1}^{n} s^*_{ih} \ \forall h.$$

A rationing process is thus necessary, which may take various forms, such as uniform rationing, queueing, priority systems, proportional rationing, etc. … To be more explicit, define:

$$\tilde{z}_{ih} = \tilde{d}_{ih} - \tilde{s}_{ih},\ z^*_{ih} = d^*_{ih} - s^*_{ih},$$

A rationing scheme on a market h is described by a set of n functions:

$$z^*_j = F_{ih}(\tilde{z}_{lh},\cdots,\tilde{z}_{nh}),\quad i = 1,\cdots,n \qquad (1)$$

such that:

$$\sum_{i=1}^{n} F_{ih}(\tilde{z}_{1h},\ldots,\tilde{z}_{nh}) \equiv 0.$$

We shall generally assume that F_{ih} is continuous, non-decreasing in $\tilde{z}_{ih}$ and non-increasing in the other arguments. Let us now examine a few possible properties. The first one is that of voluntary exchange, according to which no one can be forced to trade more than he wants, which is expressed by:

$$d^*_{ih} \leqslant \tilde{d}_{ih} \qquad\qquad s^*_{ih} \leqslant \tilde{s}_{ih}$$

or

$$z^*_{ih} \cdot \tilde{z}_{ih} \geqslant 0 \quad \text{and} \quad |\, z^*_{ih}| \leqslant |\, \tilde{z}_{ij}|$$

We shall now assume this property throughout. It allows to classify the agents in two categories: unrationed agents for which $z^*_{ih} = \tilde{z}_{ij}$ and rationed ones who trade less than they wanted. A second property we want to discuss is that of manipulability. A scheme is nonmanipulable if an agent, when rationed, cannot increase the level of his transaction by increasing his demand or supply. Priority or uniform rationing schemes are non-manipulable, a proportional rationing scheme is manipulable. Rationing schemes which satisfy both voluntary exchange and non-manipulability can be expressed under the form:

$$\begin{aligned} d^*_{ih} &= \min\left(\tilde{d}_{ih}, \overline{d}_{ih}\right) \\ s^*_{ih} &= \min\left(\tilde{s}_{ih}, \overline{s}_{ih}\right) \end{aligned} \qquad (2)$$

with:

$$\begin{aligned} \overline{d}_{ih} &= G_{ih}^{d}(\tilde{z}_{1h}, \ldots, \tilde{z}_{nh}) \\ \overline{s}_{ih} &= G_{ih}^{s}(\tilde{z}_{1h}, \ldots, \tilde{z}_{h}) \end{aligned} \tag{3}$$

where $\overline{d}_{ih}$ and $\overline{s}_{ih}$ the quantity constraints faced by agent i, are actually functions only of demands other than $\tilde{z}_{ih}$ (hence the property of non-manipulability). We thus see that a rationing which takes the form of upper bounds on net trades results from both properties of voluntary exchange and non-manipulability, and we shall assume these in what follows (a more general theory covering other cases has been developed in Benassy 1977b, 1982).

A third property which is often used, though it is not necessary for what follows, is that of market efficiency, according to which one should not find rationed demanders and rationed suppliers at the same time on a market. The intuitive idea behind it is that in an efficiently organized market a rationed buyer and a rationed seller should be able to meet, and would exchange until one of the two is not rationed. Of course this condition will be more often met in a small micro-market than on a large aggregated macro-market. Together with voluntary exchange it implies the 'short-side' rule according to which agents on the short side of the market may realize their desired transactions:

$$\tilde{z}_{ih}\left(\sum_{j}\tilde{z}_{jh}\right)\leqslant 0 \Rightarrow z_{ih}^{*} = \tilde{z}_{ih}.$$

Fixprice Equilibrium

The concept we shall describe here was developed in Benassy (1975a, 1977b, 1982). We have already seen in the previous section how transactions and quantity signals are formed in a market where effective demands and supplies have been expressed (equations (1) and (3)). All we need, in order, to obtain a fixprice equilibrium concept, is to show how optimal effective demands are expressed as a function of price and quantity signals.

For that each agent maximizes the utility of his transactions U_i $(\omega_I + z_i, m_i)$ knowing that the transactions he will obtain are related to his demands and supplies by formulas (2). A convenient solution (Benassy 1977b, 1982) is to take the effective demand $\tilde{z}_{ih}$ as the solution (unique because of strict quasiconcavity) of the following programme:

$$\text{Maximize } U_i(\omega_i + z_i, m_i) s.t. \\ \begin{cases} pz_i + m_i = \overline{m}_i \\ -\overline{s}_{ik} \leq z_{ik} \leq \overline{d}_{ik} \quad K \neq h \end{cases}$$

which yields an effective demand function denoted as $\tilde{\zeta}_{ih}(p, \overline{d}_i, \overline{s}_i)$ where $\overline{d}_i$ and $\overline{s}_i$ are the vectors of quantity constraints.

A fixprice equilibrium is now naturally defined as a set of effective demands, transactions and quantity constraints such that:

(a)

$$\tilde{z}_{ih} = \tilde{\zeta}(p, \overline{d}, \overline{s}_i) \quad \forall i, \forall h$$

(b)

$$z_{ih}^{*} = F_{ih}(\tilde{z}_{1h}, \ldots \tilde{z}_{nh}) \quad \forall i, \forall h$$

(c)

$$\overline{d}_{ih} = G_{ih}^{d}(\tilde{z}_{1h}, \ldots \tilde{z}_{nh}) \quad \forall i, \forall h$$

$$\overline{s}_{ih} = G_{ih}^{s}(\tilde{z}_{1h}, \ldots \tilde{z}_{nh}) \forall i, \forall h$$

Equilibria defined by these equations exist for all positive prices and rationing schemes satisfying voluntary exchange and non-manipulability (Benassy 1975a, 1982). Because the concept includes an explicit description of the rationing procedure, the equilibrium is unique for a given price system and rationing scheme under fairly natural assumptions (Schulz 1983).

Equilibria as defined above also possess the optimality properties one would naturally expect: they are consistent at the market level because of (b), and individually rational since effective demands have been constructed to yield optimal transactions, given price and quantity constraints. If moreover the rationing scheme on a market h is efficient, then no demanders and suppliers are rationed at the same time on that market. This last remark suggests that, even though their respective logics of construction are quite different, under the added assumption of market efficiency the Benassy and Drèze concepts should yield similar allocations at given prices. This was shown indeed by Silvestre (1982, 1983) for both exchange and productive economies. Some efficiency (and inefficiency) properties of the corresponding allocations are studied in the entry ► fixprice models.

Non-Walrasian Equilibria with Price Makers

At this stage, the theory is still in need of a description of price making by agents internal to the system. We shall now describe a concept dealing with that problem (Benassy 1976, 1977a, 1982), which synthesizes the previous developments and the theory of general equilibrium with monopolistic competition, as developed notably by Negishi (1961, 1972).

As indicated in the entry ► disequilibrium analysis, the idea behind the modelling of price making in such models is that each price maker uses the prices he controls to 'manipulate' the quantity constraints he faces. To make things more precise, assume that agent i controls the price of a subset H_i of the goods, with H_i and H_j disjoint so that the price of each good is determined by one agent at most. Agent i thus sets a vector of prices p_i,

$$p_i = \{p_h | h \in H_i\}$$

He perceives that his sales constraint in a market h he controls, $\overline{s}_{ih}$ (this constraint is actually equal to the total demand of the other agents, since he is the only seller on that market) depends on the vector p_i through a function, the perceived demand curve, denoted as

$$\overline{S}_{ih}(p_i, \theta_i)$$

where θ_i is a vector of parameters. Symmetrically a demander who sets a price p_h has a perceived supply curve

$$\overline{D}_{ih}(p_i, \theta_i)$$

We assume that the parameters θ_i are estimated as a function of current signals $p, \overline{d}_i, \overline{s}_i$ (and of course any other signal available, including data of past periods. This formulation thus allows some learning about the demand curve). Because we are dealing with a general equilibrium concept, at equilibrium the perceived demand or supply curve must go through the observed point (Bushaw and Clower 1957), i.e.

$$\begin{aligned} \overline{D}_{ih}(p_i, \theta_i) &= \overline{d}_{ih} \\ \overline{S}_{ih}(p_i, \theta_i) &= \overline{s}_{ih} \end{aligned} \tag{4}$$

We can now make explicit the procedure of price formation. Agent i, facing a price p_h and constraints $\overline{d}_{ih}$ and $\overline{s}_{ih}$ on markets $h \notin H_i$ will choose his price so as to maximize his utility, i.e. the solution in p_i to the programme

$$\text{Maximize } U_i(\omega_i + z_i, m_i) \text{s.t.}$$
$$\begin{cases} pz_i + m_i = \overline{m}_i \\ -\overline{s}_{ih} \leqslant z_{ih} \leqslant \overline{d}_{ih}, h \notin H_i \\ -\overline{S}_{ih}(p_i, \theta_i) \leqslant z_{ih} \leqslant \overline{D}_{ih}(p_i, \theta_i) \quad h \in H_i \end{cases}$$

which yields a function $\mathrm{P}_i^*(p, \overline{d}_i, \overline{s}_i)$ since the parameters θ_i are function of $p, \overline{d}_i, \overline{s}_i$. A non-Walrasian equilibrium with price makers is then simply defined as an equilibrium where quantities are optimal given prices and no price maker has interest in changing his price i.e.:

(a) The quantities $\tilde{z}_{ih}, z_{ih}^*, \overline{d}_{ih}, \overline{s}_{ih}$ form a fixprice equilibrium for p^*
(b) $p_i^* = \mathrm{P}_i^*(p*, \overline{d}, \overline{s}_i) \forall i$.

We may note that under reasonable assumptions (though not necessarily always) a price maker will satisfy the demand addressed to him. Sufficient conditions are found in Benassy (1982).

We may note as a final remark that the consistency conditions (4) imposed on the parameters of the perceived demand and supply curves are minimal ones, which thus allows to cover a maximum number of cases, depending on the structure of information available to price makers. More demanding consistency conditions have been searched for (see notably the ideas of an objective demand curve in Nikaido (1975) or of rational conjectures in Hahn 1978) but the problem has not yet received a general satisfactory solution, for lack of a well defined concept of a 'true' demand curve in a general equilibrium situation with price makers.

Expectations and Non-Walrasian Equilibria

Up to now we have dealt with an equilibrium structure in the period considered, implicitly a short-run one, but of course the economy extends further in the future, as we are reminded at least by the presence of money as a store of value. More generally the presence of stocks (inventories, capital goods, financial assets) makes it necessary to form expectations, and these will influence current equilibrium. How this occurs has been studied in Benassy (1975a, 1982), and we shall only briefly outline the method for dealing with that problem.

Each agent actually plans for the current and future periods. Expectations for future periods take the form of prices and quantity constraints (for price takers) or expected demand curves (for price makers). These may be deterministic or stochastic. These expectations are formed via expectations schemes, which link future price quantity expectations to all price quantity signals received in past and current periods. This formulation is thus quite general and covers any expectations scheme, 'rational' or not, based on actually available information.

By a standard dynamic programming technique, one can reduce the multi-period problem to a single period one, where the valuation of all stocks (and notably money) depends upon future expectations, and thus, via the expectations schemes, upon the current and past pricequantity signals, We are thus back to the one period formulation used in the previous section, with the only difference that current and past quantity signals must be added in the valuation functions. We should note that the inclusion of these expectations does not create any problem for the existence of an equilibrium when the prices are fully rigid, but may jeopardize existence when endogenous price setting is considered (see Benassy 1982). We should also note that rational expectations are fully consistent with this type of model, as was pointed out in Neary and Stiglitz (1983).

The most important feature of the introduction of expectations in such models is that, whereas traditional market clearing models deal with price expectations only, these models deal with a richer menu of price and quantity expectations.

Concluding Remarks

The concepts of equilibria with rationing, or non-Walrasian equilibria, described in this entry represent a useful generalization of the traditional Walrasian equilibrium concept in several directions: whereas the Walrasian model covers by definition only the case where all markets clear, these concepts consider more general price mechanisms including full or partial price rigidities or imperfect competition. They introduce quantity signals in addition to price signals in demand–supply theory and mixed price-quantity adjustments in the short run. They integrate quantity expectations as well as price expectations. All this is done with the same rigorous methods which have proved successful in Walrasian theory. Besides their evident microeconomic interest, non-Walrasian equilibria have been widely used in macroeconomics (see, for example, Benassy 1986), where they allow, for example, to study more rigorously states of the economy with involuntary unemployment.

A great strength of the theory is that it gives a rigorous framework within which one can predict which allocations will occur when prices are not the Walrasian ones. It also provides the first steps of a theory of endogenous price making, in line with the traditional theories of imperfect competition in general equilibrium. This theory has certainly called for new interesting developments and applications as new modes of price making without an auctioneer will be integrated within that framework.

See Also

- ▶ Disequilibrium Analysis
- ▶ Fixprice Models
- ▶ General Equilibrium

Bibliography

Arrow, K.J. 1959. Towards a theory of price adjustment. In *The allocation of economic resources*, ed. M. Abramowitz. Stanford: Standford University Press.

Arrow, K.J., and G. Debreu. 1954. Existence of an equilibrium for a competitive economy. *Econometrica* 22: 265–290.

Arrow, K.J., and F. Hahn. 1971. *General competitive analysis*. San Francisco: Holden-Day.

Barro, R.J., and H.I. Grossman. 1971. A general disequilibrium model of income and employment. *American Economic Review* 61: 82–93.

Barro, R.J., and H.I. Grossman. 1976. *Money, employment and inflation*. Cambridge: Cambridge University Press.

Bellman, R. 1957. *Dynamic programming*. Princeton: Princeton University Press.

Benassy, J.P. 1975a. Neo-Keynesian disequilibrium theory in a monetary economy. *Review of Economic Studies* 42: 502–523.

Benassy, J.P. 1975b. Disequilibrium exchange in barter and monetary economies. *Economic Inquiry* 13: 131–156.

Benassy, J.P. 1976. The disequilibrium approach to monopolistic price setting and general monopolistic equilibrium. *Review of Economic Studies* 43: 69–81.

Benassy, J.P. 1977a. A neokeynesian model of price and quantity determination in disequilibrium. In *Equilibrium and disequilibrium in economic theory*, ed. G. Schwödiauer. Boston: D. Reidel.

Benassy, J.P. 1977b. On quantity signals and the foundations of effective demand theory. *Scandinavian Journal of Economics* 79: 147–168.

Benassy, J.P. 1982. *The economics of market disequilibrium*. New York: Academic Press.

Benassy, J.P. 1986. *Macroeconomics: An introduction to the non-Walrasian approach*. New York: Academic Press.

Böhm, V., and P. Levine. 1979. Temporary equilibria with quantity rationing. *Review of Economic Studies* 46: 361–377.

Bushaw, D.W., and R. Clower. 1957. *Introduction to mathematical economics*. Homewood: Richard D. Irwin.

Clower, R.W. 1965. The Keynesian counterrevolution: A theoretical appraisal. In *The theory of interest rates*, ed. F.H. Hahn and F.P.R. Brechling. London: Macmillan.

Debreu, G. 1959. *Theory of value*. New York: Wiley.

Dehez, P., and J.H. Drèze. 1984. On supply constrained equilibria. *Journal of Economic Theory* 33: 172–182.

Drèze, J.H. 1975. Existence of an exchange equilibrium under price rigidities. *International Economic Review* 16: 301–320.

Glustoff, E. 1968. On the existence of a Keynesian equilibrium. *Review of Economic Studies* 35: 327–334.

Grandmont, J.M., and G. Laroque. 1976. On Keynesian temporary equilibria. *Review of Economic Studies* 43: 53–67.

Greenberg, J., and H. Müller. 1979. Equilibria under price rigidities and externalities. In *Game theory and related topics*, ed. O. Moeschlin and D. Pallaschke. Amsterdam: North-Holland.

Hahn, F.H. 1978. On non-Walrasian equilibria. *Review of Economic Studies* 45: 1–17.

Heller, W.P., and R.M. Starr. 1979. Unemployment equilibrium with myopic complete information. *Review of Economic Studies* 46: 339–359.

Hicks, J.R. 1939. *Value and capital*. Oxford: Clarendon Press. 2nd ed, 1946.

Keynes, J.M. 1936. *The general theory of money, interest and employment*. New York: Harcourt, Brace.

Leijonhufvud, A. 1968. *On Keynesian economics and the economics of Keynes*. Oxford: Oxford University Press.

Neary, J.P., and J.E. Stiglitz. 1983. Toward a reconstruction of Keynesian economics: Expectations and constrained equilibria. *Quarterly Journal of Economics* 98-(Supplement): 199–228.

Negishi, T. 1961. Monopolistic competition and general equilibrium. *Review of Economic Studies* 28: 196–201.

Negishi, T. 1972. *General equilibrium theory and international trade*. Amsterdam: North-Holland.

Nikaido, H. 1975. *Monopolistic competition and effective demand*. Princeton: Princeton University Press.

Schulz, N. 1983. On the global uniqueness of fixprice equilibria. *Econometrica* 51: 47–68.

Silvestre, J. 1982. Fixprice analysis of exchange economies. *Journal of Economic Theory* 26: 28–58.

Silvestre, J. 1983. Fixprice analysis in productive economies. *Journal of Economic Theory* 30: 401–409.

Triffin, R. 1940. *Monopolistic competition and general equilibrium theory*. Cambridge, MA: Harvard University Press.

Van Der Laan, G. 1980. Equilibrium under rigid prices with compensation for the consumers. *International Economic Review* 21: 63–74.

Walras, L. 1874. *Eléments d'économie politique pure*. Lausanne: Corbaz. Definitive edition translated by W. Jaffé: *Elements of pure economics*. London: Allen & Unwin, 1954.

Younès, Y. 1975. On the role of money in the process of exchange and the existence of a non-Walrasian equilibrium. *Review of Economic Studies* 42: 489–501.

Rationing

J. Peter Neary

Abstract

Rationing occurs whenever economic agents face quantity constraints on their demand for or supply of particular commodities. This article reviews the main results of rationing theory: a tightening of a ration constraint raises the demand for unrationed substitutes and reduces the price responsiveness of all unrationed goods (the Le Chatelier effect). It shows how the technique of virtual prices can be used to generalize these results to the case of strictly binding rations, and briefly reviews some applications, empirical and theoretical, of rationing theory to public and environmental economics, fix-price macroeconomics, and the effects of quotas on international trade.

Keywords

Compensated demand; Demand price; Fixprice macroeconomics; Income effects; International trade theory; Labour supply; Le Chatelier principle; Nonlinear budget constraints; Nonlinear commodity taxation; Public goods; Quotas; Rationing; Reservation price; Separability; Shadow price; Substitution effect; Uncompensated demand; Unemployment; Virtual price

JEL Classifications

D11

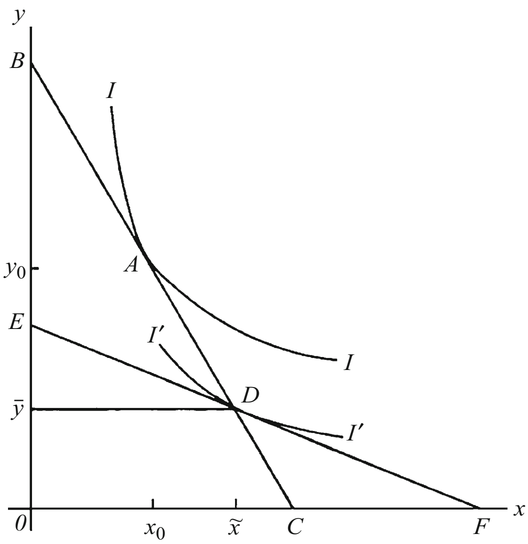
Rationing refers to any situation in which economic agents face quantity constraints on their demand for or supply of particular commodities, unlike the standard situation in which they are free to purchase unlimited quantities subject only to fixed prices and a linear budget constraint.

Quantity constraints may impose consumption levels either below or above those that would be freely chosen: goods rationing in wartime illustrates the former, while examples of the latter include precommitted expenditures and unemployment (which may be viewed as 'forced consumption' of leisure). From an analytic point of view, the two cases are identical and may be described by the general term 'rationing'. This article outlines the principal results of the microeconomic theory of consumer rationing and then notes some of its applications. Similar results apply to producer rationing, with the added simplification that income effects do not arise for a profit-maximizing firm. I concentrate throughout on 'simple' rationing (that is, exogenous restrictions on the consumption of particular commodities); some work has also been done on 'points' rationing (where the consumer has a number of ration 'points' or 'coupons' to be allocated between a group of commodities), and there is an extensive literature on the general case of nonlinear budget constraints, from both theoretical and empirical perspectives (see Hausman 1985).

Consider first the case where only two commodities are consumed. This misses many important aspects of rationing, but allows most of the basic ideas to be introduced using a simple diagram. In Fig. 1, the unconstrained optimal consumption bundle (x_0, y_0) is represented by point A, the point of tangency between the budget constraint BC and the highest attainable indifference curve, II. Suppose now that the consumer is faced with an additional constraint which stipulates that consumption of commodity y cannot exceed the level $\bar{y}$. The consumer is therefore forced to adjust consumption to the point D. Here, the budget constraint is still satisfied, consumption of y is constrained to equal $\bar{y}$, and expenditure has spilled over onto the unrationed commodity x, leading to a new higher consumption level $\tilde{x}$ (where a tilde '~' denotes a demand schedule for unrationed commodities in the presence of the ration constraint $\bar{y}$). The consumer is also at a lower indifference curve $I'I'$, so rationing reduces real income, or increases the true cost of living, even though prices and nominal income are unchanged.

The case illustrated in Fig. 1 is of course extremely special because the number of commodities is the same as the number of independent binding constraints, so the optimal consumption

R



Rationing, Fig. 1

bundle is uniquely determined. In the more usual case, where there are fewer constraints than commodities, the consumer is free to allocate her uncommitted expenditure between a number of unrationed commodities, and a major focus of rationing theory has been on how this allocation, and its responsiveness to changes in exogenous variables, is affected by the presence of rationing. One important special case where rationing has very simple effects is when preferences are weakly separable between the rationed and unrationed commodities. This implies that the direct utility function $v(x, y)$, can be written in the form $U[f(x), y]$ where f is a scalar sub-utility function defined over a vector of unrationed commodities x. Weak separability implies that the demand for each good depends only on the prices of goods within the separable group and on the expenditure allocated to it. Hence the ration constraint has an income effect only and the constrained demand functions for the unrationed goods take the special form $\tilde{x}(p, I - q\bar{y})$. This specification is plausible in the case of some public goods (for example, increased spending on national defence is unlikely to affect the pattern of demand for private goods). Unfortunately, it is less satisfactory in other applications. For example, if leisure is the rationed commodity, weak separability implies that all other goods must be substitutes for it, irrespective of the extent of unemployment.

When preferences are unrestricted, a useful starting point to understanding the effects of rationing is to note that an unrationed consumer would choose to consume at point D under certain circumstances. Specifically, this would occur if the consumer were faced with a relative price ratio equal to the tangent to the indifference curve $I'I'$ at D, and were given an adjusted level of income such that that point represented the unconstrained utility-maximizing consumption bundle. The hypothetical relative price ratio required is given by the slope of the line EF. Following Rothbarth (1941) and Neary and Roberts (1980), the price of the rationed commodity underlying EF is called its *virtual price:* the price which would induce the consumer to purchase the ration level voluntarily. The advantage of this approach is that the effect of any exogenous shock on a rationed consumer may be decomposed into the sum of two effects on an orthodox unrationed consumer: the direct effect of the shock itself and the indirect effect arising from the induced change in the virtual price of the rationed good. (Note in passing that the terms 'virtual price', 'demand price' or, if the ration is set at a zero level, 'reservation price' are preferable to 'shadow price', since the latter risks confusion with the shadow price of the ration constraint which emerges from the consumer's maximization problem.)

It is clear that, for non-zero virtual prices to be unique and well defined, the indifference curve at D must be convex and differentiable. (Further technical details may be found in Neary and Roberts 1980.) Consider then the general case where unrationed commodities are represented by a vector x and their prices by a vector p, while the commodity subject to a binding ration constraint is represented by a scalar y and its market price by q. (The algebra that follows applies equally to the general case with more than one rationed commodity, but it is easier to give intuition for the case of a single rationed good.) It is then straightforward to relate the Hicksian or compensated demand schedules for x in the presence of rationing to the corresponding unrationed schedules, since both are evaluated at the same utility level, that corresponding to the indifference curve $I'I'$:

$$\tilde{x}^c(p,\bar{y},u) = x^c(p,\bar{q},u). \tag{1}$$

Here, a superscript 'c' denotes compensated demands. Crucially, the virtual price $\bar{q}$ is not a parameter but is defined implicitly by the condition that it equate the unconstrained demand for y to the ration level $\bar{y}$:

$$\bar{y} = y^c(p,\bar{q},u). \tag{2}$$

Differentiating (1) and (2) now yields two important comparative statics results. Consider first the effect of a change in the ration level $\bar{y}$ on the demand for unrationed commodities x:

$$\tilde{x}_y^c = x_q^c\left(y_q^c\right)^{-1} \tag{3}$$

where subscripts indicate partial derivatives (for example, $\tilde{x}_y^c$ is the vector whose i'th element gives the partial derivative of the rationed compensated demand function for x_i with respect to the level of the ration constraint). Since the compensated own-price derivative y_q^c is negative, the sign of (3) depends on the sign of x_q^c. Thus, a tightening of the ration constraint (a reduction in $\bar{y}$) raises the compensated demand for unrationed commodities which are net substitutes for y and reduces it for commodities which are net complements for y.

Next, consider the effects on the demand for x of changes in their own prices. Differentiating (1) and (2) and rearranging yields:

$$\tilde{x}_p^c - x_p^c = -x_q^c\left(y_q^c\right)^{-1}x_q^{c\prime}. \tag{4}$$

Since the substitution effect y_q^c is negative, this equation shows that the difference between the matrices of own-price responses of the unrationed commodities with and without rationing is a positive definite matrix. For any particular unrationed commodity, this implies that rationing reduces its responsiveness to its own price. This result is often referred to as the Le Chatelier principle, and was first introduced into economics by Samuelson (1947, pp. 36–9). Strictly speaking, the principle relates only to a comparison of compensated demands (compare (10) below). Moreover, it is a local result only, since it requires that the derivatives of both rationed and unrationed demand schedules be evaluated at the same consumption bundle, point D in Fig. 1 (though Roberts 1999, shows that it applies globally in an average sense). Despite these qualifications, the principle is often interpreted as implying in general that the imposition of restrictions on some aspects of behaviour makes individuals less responsive to exogenous changes in their environment.

Equations (3) and (4) are two of the most important results in rationing theory. However, their simplicity depends crucially on the fact that they refer to the properties of compensated demand schedules. There is one special case where equation (3) holds exactly for uncompensated (Marshallian) as well as compensated (Hicksian) demands, namely, where the ration 'just' binds, in the sense that the ration constraint coincides exactly with the amount of y that would be demanded by an unrationed consumer (so that points A and D in Fig. 1 coincide). This was the case for which Tobin and Houthakker (1950–1) derived their results in a classic paper. For strictly binding ration constraints, any exogenous change has additional income effects, whose implications were first derived by Neary and Roberts (1980).

To illustrate the additional income effects which strictly binding ration constraints introduce, refer again to Fig. 1. The distance OC measures the consumer's actual income in terms of x, I/p or $\tilde{x} + (q/p)\bar{y}$. However, this income would not be sufficient to induce an unrationed consumer faced with prices p and $\bar{q}$ to consume voluntarily at D; to do this they would need an income equal (in terms of good x) to the distance OF. Simple geometry shows that this distance equals $\tilde{x} + (q/p)\bar{y}$ or $[I + (\bar{q} - q)\bar{y}]/p$. Hence, the uncompensated demands of the rationed consumer may be equated to the uncompensated demands of an unrationed consumer, provided the

R

latter are evaluated at the virtual price $\overline{q}$ and at a 'virtual income' $I + (\overline{q} - q)\overline{y}$:

$$\tilde{x}(p, q, \overline{y}, I) = x[p, \overline{q}, I + (\overline{q} - q)\overline{y}]. \tag{5}$$

In addition, the virtual price and income must be such that they induce an uncompensated demand for the rationed good equal to the ration constraint, so that (2) must be replaced by:

$$\overline{y} = y[p, \overline{q}, I + (\overline{q} - q)\overline{y}]. \tag{6}$$

Differentiating (5) and (6) now yields the full effects of exogenous changes on the demand for unrationed commodities. Consider first the effect of a change in income:

$$\tilde{x}_I = x_I - x^c_y y_I. \tag{7}$$

Thus, an increase in income affects demands for unrationed goods in two ways: first, it has a direct effect identical to the effect of an income increase in the absence of rationing (though evaluated at the virtual prices and income, of course); and second, by raising demand for the rationed good (on the assumption that y is normal so that y_I is positive), it is equivalent to a tightening of the ration constraint, and so has an indirect effect given by eq. (3).

Differentiating (5) and (6) also gives the effect of a change in the ration constraint:

$$\tilde{x}_y = \tilde{x}^c_y + \tilde{x}_I(\overline{q} - q). \tag{8}$$

This shows that a tightening of the ration constraint has a compensated or substitution effect given by (3) and an additional income effect, given by the last term in (8). This term vanishes if the virtual and actual prices of the rationed good coincide, which corresponds to the case where the ration constraint 'just' binds. In the case illustrated in Fig. 1, where the consumer would like to consume more of the rationed good, $\overline{q}$ exceeds q, and so a tightening of the ration constraint, by lowering real income, tends to reduce the demand for all normal unrationed goods. (Of course, as already noted in discussing the diagram, the total effect must be an increase in spending on the unrationed goods as a group.)

Finally, the effect of changes in prices may be obtained in a similar manner. First, for an increase in the prices of the unrationed goods themselves:

$$\tilde{x}_p = x_p - \tilde{x}^c_y y_p. \tag{9}$$

This shows that an increase in the price of an unrationed good has a direct effect, equal to its effect in the absence of rationing, and an indirect effect: by changing the demand for the rationed good it is equivalent to a tightening or relaxation of the ration constraint and so has the usual effect given by (3). Equation (9) may be rewritten in a form which, by comparison with (4), shows how income effects may counteract the Le Chatelier principle:

$$\tilde{x}_p - x_p = \left(\tilde{x}^c_p - x^c_p\right) + \tilde{x}^c_y y_I x. \tag{10}$$

By contrast, the effect of a change in the price of the rationed good is much simpler:

$$\tilde{x}_q = -\tilde{x}_I \overline{y}. \tag{11}$$

This price change has no substitution effect, which explains why q is not an argument in the compensated rationed demand schedules (1). Its only effect is to lower real income by requiring the consumer to pay more for the rationed good, and so it reduces the demand for normal unrationed goods.

Before we leave the basic comparative statics of rationing, a problem which is peculiar to this area should be mentioned. All the results which have been derived assume that the values of the exogenous variables are such that the ration constraint is a binding one. However, it is quite possible for a finite change in an exogenous variable to render the constraint non-binding. For example, in Fig. 1 this would occur if the ration constraint $\overline{y}$ rose above the unconstrained demand $y(p, q, I)$. If this happens, the ration constraint ceases to be binding and the ordinary unconstrained demand functions become applicable. Shifts of 'regime' such as this dictate great care in applying rationing theory in cases where large finite changes in exogenous variables occur; and in applications such as fixprice macroeconomics, where interest focuses on the interaction between constraints which impinge on different agents.

In empirical applications of rationing theory, attention has focused on deriving explicit forms for rationed demand functions which are tractable but not too restrictive. As in the case of consumer theory in the absence of rationing, progress in this direction comes most easily not by specifying functional forms for the direct utility function but by adopting a dual approach, which takes the expenditure function as its starting point. In the presence of rationing the constrained expenditure function gives the minimum cost of attaining a given utility level when consumption of the rationed commodity is predetermined:

$$\begin{aligned}\tilde{E}(\bar{y},p,q,u) &= Min[p'x+q\bar{y}\ :\ v(x,y)\geq u]\\ &= p'\tilde{x}^c(p,\bar{y},u)+q\bar{y}.\end{aligned} \tag{12}$$

Substituting from (1) and (2) yields, after some manipulation, the fundamental relationship between constrained and unconstrained expenditure functions:

$$\tilde{E}(\bar{y},p,q,u) = E(p,\bar{q},u) + (\bar{q}-q)\bar{y}. \tag{13}$$

In principle, this identity permits the derivation of a matched pair of rationed and unrationed demand functions, characterizing the behaviour of the same consumer in both environments. Two interesting specifications of the expenditure function which permit this are investigated in the labour supply context by Deaton and Muellbauer (1981). Unfortunately, the derivation of such matched pairs of demand functions is not possible in general. An alternative approach is to specify a general functional form for the rationed expenditure function which imposes fewer restrictions on demand responses though at the cost of an inability to write the unrationed demand functions in closed form. This approach has been pursued by Deaton (1981), who derives a system of rationed demand functions which express budget shares as a linear function of the ration level and the logarithms of prices and real expenditure on unrationed goods. He shows that treating expenditure on housing as predetermined in this framework leads to more plausible results than when it is assumed to be unconstrained. (Specifically, the rationed system goes much of the way towards avoiding the implausible rejection of homogeneity in nominal variables, which has been found in many empirical studies of demand.)

Insights derived from rationing have proved useful in many branches of economic theory other than consumer economics. In public economics and environmental economics, the study of optimal public policy has been extended to public goods and bads (the consumption of which is predetermined from an individual consumer's point of view) and nonlinear commodity taxation (of which government-imposed consumption quotas are a special but empirically important case). In macroeconomics, rationing theory has been used to model both current and expected future quantity constraints on households and firms in formalizations of Keynes's contribution as the economics of 'general disequilibrium', in which the failure of prices to adjust rapidly in the short run faces agents with quantity constraints which 'spill over' to influence their behaviour in other markets (see Neary and Stiglitz 1983, and further references given there). In international trade theory, it has been shown that the behaviour of an economy subject to quotas (quantitative restraints on imports) can be characterized in the same way as that of a household subject to rationing, with the added benefit that the virtual prices can be interpreted as the domestic market-clearing prices (see Anderson and Neary 1992).

Finally, within a utility-maximizing framework, it may be noted that rationing necessarily imposes a welfare loss. This consideration underlies the instinctive preference by most economists for the use of the price system as an allocation mechanism rather than direct controls, a preference which is supported by the two fundamental theorems of welfare economics. Nevertheless, in situations where the conditions for these theorems do not obtain, it may be possible to give a second-best justification for rationing. While work along these lines pertains more to public economics than to rationing theory per se, mention may be made of two especially interesting contributions. One is a paper by Weitzman (1977), who develops a model where the just distribution of a particular commodity on the basis of need alone is considered a socially desirable end in itself. He shows that rationing the commodity is preferable to allocating it via the price system if tastes are homogeneous but income is unevenly distributed. The other is a paper by Guesnerie and Roberts (1984), who show that, in a second-best world with given commodity taxes (so that consumer prices diverge from marginal social valuations), rationing is likely to be welfare improving.

See Also

- ▶ Demand Theory
- ▶ Labour Supply
- ▶ Le Chatelier Principle

Bibliography

Anderson, J.E., and J.P. Neary. 1992. Trade reform with quotas, partial rent retention and tariffs. *Econometrica* 60: 57–76.

Deaton, A.S. 1981. Theoretical and empirical approaches to consumer demand under rationing. In *Essays in the theory and measurement of consumer behaviour in honour of Sir Richard Stone*, ed. A. Deaton. Cambridge: Cambridge University Press.

Deaton, A., and J. Muellbauer. 1981. Functional forms for labour supply and commodity demands with and without quantity constraints. *Econometrica* 49: 1521–1532.

Guesnerie, R., and K.W.S. Roberts. 1984. Effective policy tools and quantity controls. *Econometrica* 52: 59–86.

Hausman, J.A. 1985. The econometrics of nonlinear budget sets. *Econometrica* 53: 1255–1282.

Neary, J.P., and K.W.S. Roberts. 1980. The theory of household behaviour under rationing. *European Economic Review* 13: 25–42.

Neary, J.P., and J.E. Stiglitz. 1983. Towards a reconstruction of Keynesian economics: Expectations and constrained equilibria. *Quarterly Journal of Economics* 98(Supplement): 199–228.

Roberts, K. 1999. Rationality and the Le Chatelier Principle. *Journal of Economic Theory* 87: 416–428.

Rothbarth, E. 1941. The measurement of changes in real income under conditions of rationing. *Review of Economic Studies* 8: 100–107.

Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Tobin, J., and H.S. Houthakker. 1950–51. The effects of rationing on demand elasticities. *Review of Economic Studies* 18: 140–153.

Weitzman, M. 1977. Is the price system or rationing more effective in getting a commodity to those who need it most? *Bell Journal of Economics* 8: 517–524.

Rau, Karl Heinrich (1792–1870)

H. C. Recktenwald

Keywords

Cameralism; Freiburg School; Historical School; German; Public finance; Rau, K. H.; Smith, A.

JEL Classifications

B31

Rau was born in Erlangen and was a lecturer (*Privatdozent*) and professor (1816). In 1822 Rau was appointed to a chair of economics at the University of Freiburg. Involved in political affairs, as were many German professors in the 19th century, he was appointed a member of the upper Chamber of Baden and in 1848 was elected to the Frankfurt Assembly.

At first influenced by Cameralist ideas, Rau was one of the main mediators and defenders of Smith's 'system of natural liberty', whose central principles, abstractly exposed, he embodied in a rich supply of illustrative facts in his famous *Lehrbuch* (1826–1837) yet without attempting to test his hypotheses empirically, that is, to use factual materials as confirmation instead of pure description. To that extent he was not an original thinker. Yet he was a great teacher. Similar to Samuelson's *Economics* in our time, his best-selling textbook, published in eight editions

(1862–1869), was an authoritative work for the majority of economists teaching at German universities for several generations. Based on classical ideas, it thus shaped the economic and political *Weltbild* of future civil servants and lawyers.

Rau's tripartite division of economics, which was obviously influenced by Smith, was divided into three volumes, theory (economic laws), policy (*Polizeiwissenschaft*) and public finance; this division became the established tradition in the teaching of political economy at German universities and is divisive up to the present day. With the rise and the establishment of the German Historical School and its stress on both the ethical aspect of economic issues, that is, of the distribution of income and property, and on the historical character of economic principles, Rau's star faded, although his work on public finance became the foundation of Wagner's famous treatise.

Viewed in a historical continuum, the Freiburg School (Eucken, Röpke, von Hayek), Erhard's liberal economic policy and, more recently, a group of German economists who are attempting to revive Smith's tripartite theory of order (ethics, economics and politics as an entity) all indirectly resume that thread of Rau's concept, although on a different analytical level (Recktenwald 1973, 1985). In the light of a worldwide Smith renaissance in our epoch, Rau's editing function seems to merit secular attention.

Selected Works

For a complete list, see C. Meitzel, *Handwörterbuch der Staatswissenschaft,* vol. 6, 4th ed. Jena: G. Fischer, 1925.

1825. Über die Kameralwissenschaft. Heidelberg: Groos.

1826–1837. *Lehrbuch der politischen Ökonomie*, 3 vols, 8th ed. Heidelberg: C.F. Winter.

Bibliography

Recktenwald, H.C., ed. 1973. *Political economy: A historical perspective*. London: Collier-Macmillan.

Recktenwald, H.C., ed. 1985. *Ethik, Wirtschaft und Staat: Adam Smiths politische Ökonomie heute*. Darmstadt: Wissenschaftliche Buchgesellschaft. 110–20 and 345–90 and G.J. Stigler's and P.A. Samuelson's contributions.

Ravenstone, Piercy

D. P. O'Brien

Attention was drawn to a book bearing this author's name because it is mentioned in the Ricardo–Malthus–Mill correspondence. The true identity of the author is not known for certain, although Sraffa (1973), on the basis of two separate identifications in libraries, suggests it was Richard Puller (*fl.* 1789–1831); on the basis of internal evidence, Dorfman (1966) had previously suggested that the author was an Anglican clergyman, Edward Edwards.

The author has been classified (by Seligman and by Halévy) as a Socialist; as a Tory Democrat (by Beer); and as an Institutionalist (by W.C. Mitchell). The content of the work discussed by Ricardo is, however, essentially Physiocratic; there is no understanding of division of labour, productivity, or exchange; tradesmen add no value but simply pass wealth through their hands; and a large sterile class enjoys revenue generated by a productive class employed in agriculture. Property is perfectly justifiable where it originates in labour, but it has developed into an institution enabling the sterile class to live on the productivity of the productive class. This abstraction takes place through both rent and profits – capital is an imaginary concept designed to justify the revenue enjoyed as profit. Poverty is not caused by Malthusian population pressures – with remarkable self–confidence in the employment of arithmetical ratios, the author satisfied himself that the rate of population increase was everywhere constant – but to the abstraction of revenue for the sterile classes. Since debt service taxes were also used to this end, public debt was undesirable.

Selected Works

1821. A few doubts as to the correctness of some opinions generally entertained on the subjects of population and political economy. London: J. Andrews. Reprinted with an introduction by J. Dorfman. New York: A.M. Kelley, 1966.

1824. Thoughts on the funding system and its effects. London: J. Andrews.

Bibliography

Sraffa, P. (ed.) 1952, 1973. *The works and correspondence of David Ricardo*, vols. IX, XI. Cambridge: Cambridge University Press for the Royal Economic Society.

Raymond, Daniel (1786–1849)

Henry W. Spiegel

Keywords

Lauderdale, Eighth Earl of; List, F.; National system; Protectionism; Raymond, D.

JEL Classifications

B31

The first American to publish a treatise on economic topics, Raymond was born in Connecticut but made his home in Baltimore, where he practised law. *Thoughts on Political Economy* (1820) was written to while away the time as the young attorney waited for clients. The book constituted a challenge to classical orthodoxy and as such was warmly received by the protectionists. To make his voice more resounding they tried (without success) to secure Raymond a professorship at the University of Maryland that they were willing to underwrite. Raymond was an original thinker, whose ideas reverberated in the later writings of Frederick List, the historical economists and the 20th-century literature on economic development.

Raymond's principal concern was national economic development and, unlike the classics, he placed the nation rather than the individual in the centre of his analysis. Following Lauderdale, he distinguished between national and individual wealth, but unlike Lauderdale, to whom usefulness was the characteristic feature of public wealth and scarcity that of private wealth, Raymond interpreted national wealth in terms of its 'capacity' to produce goods. This view opens up to government a central position in promoting economic development by means of tariff protection. Raymond also underlines the conflicts of interest among different groups in the economy and again calls on government for their resolution.

While Raymond's basic ideas reflect the influence of Alexander Hamilton, his distrust of paper money and bank credit echoes the related views of Jefferson. Raymond also was highly critical of corporations. These incongruities were bound to affect the impact of his work.

See Also

▶ National System

Selected Works

1819. *The missouri question.* Baltimore: Schaeffer & Maund.

1820. *Thoughts on political economy.* Baltimore: F. Lucas, Jr.

1828. *The American system.* Baltimore: Lucas & Deaver.

n.d. *The elements of constitutional law.* 2nd ed. Baltimore: F. Lucas, Jr. and E.J. Cooder, 1832.

Bibliography

Conkin, P.K. 1980. *Prophets of prosperity: America's first political economists*. Bloomington: Indiana University Press.

Dorfman, J. 1946. *The economic mind in American civilization 1606–1865*. New York: Viking.

Read, Samuel (fl. 1816–29)

R. M. Rauner

There is no biography of Samuel Read. Nothing is known of his life apart from his writings, which were all published between 1816 and 1829 in Edinburgh. The Preface of his principal work was annotated from Roslin, which is about eight miles south of Edinburgh.

Read's earlier tracts were on currency and government debt, plus a small pamphlet opposing Malthus on population. His major work was *An Inquiry into the Natural Grounds of Right to Vendible Property or Wealth,* published in 1829, in which he tried to put certain moral laws against the importance that economists attributed to material wealth. He used the utilitarian calculus as the basis for human action and the determination of economic justice, which included the natural right of the poor to public support. Labour he deemed the foundation and only certain measure of value. But wealth was more than just the product of labour, since accumulated capital and land also contributed to its production.

Read admitted Samuel Bailey's (1825) influence, but this extended only to showing the absurdity of reducing capital and the time needed for its production into mere labour expended; he did not follow Bailey on the relativity of value. With Bailey and some other Ricardian critics, Read objected to the theory that wages and profits varied inversely and he hinted at an abstinence factor in the supply of capital, thus predating Nassau Senior (1836).

Read was also somewhat ahead of the economists of his time in urging nationalization of local poor rates (i.e., the taxes necessary to support the unemployed, indigent, and other needy), but Samuel Whitbread had already proposed regularization of the poor relief rates in 1807.

Selected Works

1816. *On money and the Bank restriction laws*. Edinburgh.

1818. *The problem solved; an explication of a plan of a safe, steady, and secure government paper currency, and legal tender*. Edinburgh.

1819. *Exposure of certain plagiarisms of J.R. McCulloch Esq., author of two essays on reduction of the interest of the National Debt, committed in the last published of those essays, the Scotsman Newspaper and the Edinburgh Review*. Edinburgh.

1821. *General statement of an argument on the subject of population in answer to Mr. Malthus's Theory*. Edinburgh.

1829. *An inquiry into the natural grounds of right to vendible property or wealth*. Edinburgh.

Bibliography

Bailey, S. 1825. *A critical dissertation on the nature, measures, and causes of value; chiefly in reference to the writings of Mr. Ricardo and his followers*. London: Printed for R. Hunter.

Senior, N. 1836. *An outline of the science of political economy*. London: Allen & Unwin.

Real and Nominal Quantities

R. O'Donnell

Adam Smith, in his *Inquiry into the Nature and Causes of the Wealth of Nations* of 1776, distinguished between the *nominal* price and the *real* price of commodities. He defined the nominal price of a commodity as its price in silver or gold and the real price as the quantity of labour which it can purchase or 'command'. In other words, he adopted the money wage of ordinary labour as his standard of value. Ever since the critical commentaries of Ricardo and Marx this choice of standard has been seen by many as evidence that Smith's measure and theory of value were confused or inconsistent (see Ricardo, *Works* I, pp. 14–15; Marx 1861–63, I, pp. 69–77; II, pp. 200, 369; and Douglas 1928). More recently there has been a rejection of the view that Smith confused labour embodied and labour commanded and, as Hollander says, 'the issue at hand, it is now generally recognized, corresponds to the modern "index number" problem of estimating changes in "real income" over space and time' (Hollander 1973, p. 127; see also Schumpeter 1954, p. 127 and Blaug 1978, pp. 51–3).

Undoubtedly some of this criticism and disagreement arose because of the obscurity of Smith's account – his chapter 'On the real and nominal Price of Commodities or of their Price in Labour, and their Price in Money' has recently been described as 'arguably . . . the most convoluted chapter ever to emerge from the pen of a great economist' (O'Brien 1975, p. 82). However, if the *purpose* for which Smith wanted this measure of value, and the *assumptions* upon which he based it, are clearly identified then no serious confusions or inconsistencies are found, and the now widely accepted view that he adopted labour command as an index of general purchasing power can be dismissed.

Smith's Labour Command Measure

In his treatment of value Smith focused on *changes* in relative price brought about in the process of technical change (see Bladen 1975). Consequently he was concerned to find a 'standard by which we can compare the values of different commodities at all times and at all places' (Smith 1776, I, v.17). He rejected gold and silver because their values vary due to changes in their method of production, and 'a commodity which is itself continually varying in its own value, can never be an accurate measure of the value of other commodities' (1776, I, v.7). It has seldom been noted that in choosing to measure the change in the value of a commodity by the change in the quantity of labour it can command Smith made a number of assumptions which served to lend a rational foundation to that choice: he adopted a set of assumptions which rendered changes in the labour *commanded* by a commodity roughly proportional to changes in the labour *embodied* in it.

First, Smith assumed that 'equal quantities of labour, at all times and places, may be said to be equal value to the labourer' (1776, I, v.7) – thus labour *time* is a good measure of ultimate dearness or difficulty of production. This assumption has indeed been noted by many commentators. Second, Smith assumed that the corn wage was constant over long periods (1776, I, v.15). Third, and most significant, was Smith's assumption that corn was produced at constant cost. He made only an oblique reference to this in chapter v of Book I (see 1776, I, v.16) but that scarcely excuses this crucial assumption being ignored in most commentaries (see, e.g. Bowley 1973, p. 116) and openly denied in others (see Hollander 1973, p. 130n). For, when he came to *use* his labour command or corn measure Smith made this assumption quite explicit –

> In every different stage of improvement, besides, the raising of equal quantities of corn in the same soil and climate, will, at an average, require nearly equal quantities of labour; or what comes to the same thing, the price of nearly equal quantities (1776, I, xi.e.28).

Furthermore, he stated clearly that this constant cost was the *basis* upon which his use of the labour or corn measure of value was founded (1776, I, xi.e.28).

The Operation of Smith's Measure

The assumptions outlined above lent a degree of logical validity to labour command or corn as a measure of *changes* in value. As Sylos-Labini points out, it is known 'that the variations of this standard correspond to those expressed by labour embodied if the distributive shares are constant' (Sylos-Labini 1976, p. 206). It should be clear that the use by Smith of a labour commanded or corn measure to examine changes in value due to changes in methods of production depended not only on a constant corn wage but also on the constant production cost of corn. For without this is a change in the *real price* (labour command) of any given commodity could indicate not only a change in *its* value but also a change in the value of *corn*.

A numerical example can be constructed which illustrates the primary use of the labour command measure of value in the *Wealth of Nations*. Consider a *manufactured commodity* in the production of which improved techniques have reduced the amount of labour required (H) from 2 to 1.

	H	W	$H.W.$	a	$P = (WH/a)$	$P.W$
Time 1:	2	10	20	0.5	40	4
Time 2:	1	10	10	0.5	20	2

A constant corn wage requires a constant money wage (W) of 10, given the unchanged production conditions of corn and an unchanged value of money. If the share of wages in the value of output (denoted a) is constant then the change (fall) in the value (of the manufacturer commodity) measured in labour commanded (P/W) will be proportional to its change measured in labour embodied (H). Both labour embodied and labour commanded will have been halved. Smith's major use of his measure of value in this fashion was in his long 'Digression concerning the variations in the value of silver' in chapter xi of Book I. There he challenged the conventional view that a high or low *money* price of goods in general – that is, a high or low value of silver – was an indicator of the level of economic development of a country (1776, I, xi. n. 2). Against this he argued that it was only from the high or low price of certain goods 'in proportion to that *of corn*' (i.e. their high or low *real price*) that 'we can infer, with a degree of probability that approaches almost to certainty, that it was rich or poor, that the greater part of its lands were improved or unimproved, and that it was either in a more or less barbarous state, or in a more or less civilized one' (1776, I, xi. n. 3).

In view of this demonstration that Smith's measure of value was designed precisely to deal with *changing* relative prices it is clear that the now widely accepted view that he saw the labour command value or real price of a given commodity or aggregate of commodities as an index of its *general purchasing power* must be dismissed. Indeed, it is a wonder that the spread of this interpretation was not long ago halted by Smith's own explicit statement that differential rates of productivity growth will sever any connection between changes in the value of an individual commodity as measured by labour commanded (or labour embodied) and changes in its purchasing power over other commodities in general (see Smith 1776, I, viii. 4).

See Also

▶ Index Numbers
▶ Real Income

Bibliography

Bladen, V.W. 1975. Command over labour: A study in misinterpretation. *Canadian Journal of Economics* 8(4): 504–519.
Blaug, M. 1978. *Economic theory in retrospect*, 3rd ed. Cambridge: Cambridge University Press.
Bowley, M. 1973. *Studies in the history of economic theory before 1870*. London: Macmillan.
Douglas, P. 1928. Adam Smith's theory of value and distribution. In *Adam Smith'*, ed. J.M. Clark et al., 1776–1926. Chicago: University of Chicago Press.
Hollander, S. 1973. *The economics of Adam Smith*. London: Heinemann.
Marx, K. 1861–63. *Theories of surplus value*. In 3 parts. London: Lawrence & Wishart, 1969–72.
O'Brien, D.P. 1975. *The classical economists*. Oxford: Clarendon Press.
Ricardo, D. 1821. Principles of political economy. In *Works and correspondence of David Ricardo*, Vol. I, ed. P. Sraffa with the collaboration of M.H. Dobb. Cambridge: Cambridge University Press, 1951.
Schumpeter, J. 1954. *History of economic analysis*. London: George Allen & Unwin.
Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations,* 2 vols, ed. R.H. Campbell and A.S. Skinner. Oxford: Clarendon Press, 1976.
Sylos-Labini, P. 1976. Competition: The product market. In *The market and the state: Essays in honour of Adam Smith*, ed. T. Wilson and A.S. Skinner. Oxford: Clarendon Press.

R

Real Balances

Don Patinkin

JEL Classifications
E5

By the term 'real balances' is meant the real value of the money balances held by an individual or by the economy as a whole, as the case may be. The emphasis on real, as distinct from nominal, reflects the basic assumption that individuals are free of 'money illusion'. It is a corresponding property of any well-specified demand function for money that its dependent variable is real balances. Indeed, Keynes in his *Treatise on Money* (1930, vol. 1, p. 222) designated the variation on the Cambridge equation that he had presented in his *A Tract on Monetary Reform* (1923, ch. 3: 1) as 'The "Real-Balances" Quantity Equation'.

Implicit – and sometimes explicit – in the quantity-theory analysis of the effect of (say) an increase in the quantity of money is the assumption that the mechanism by which such an increase ultimately causes a proportionate increase in prices is through its initial effect in increasing the real value of money balances held by individuals and consequently increasing their respective demands for goods: that is, through what is now known as the 'real-balance effect'. This effect, however, was not assigned a role in the general-equilibrium system of equations with which writers of the interwar period attempted to describe the workings of a money economy. In particular, these writers mistakenly assumed that in order for their commodity demand functions to be free of money illusion, they had to fulfil the so-called 'homogeneity postulate', which stated that these functions depended only on relative prices, and so were not affected by a change in the absolute price level generated by an equi-proportionate change in all money prices (Leontief 1936, p. 192). Thus they failed to take account of the effect of such a change on the real value of money balances and hence on commodity demands. This in turn led them to contend that there existed a dichotomy of the pricing process, with equilibrium relative prices being determined in the 'real sector' of the economy (as represented by the excess-demand equations for commodities), while the equilibrium absolute price level was determined in the 'monetary sector' (as represented by the excess-demand equation for money): (Modigliani 1944, sec. 13). This, however, is an invalid dichotomy, for it leads to contradictory implications about the determinacy or, alternatively, stability of the absolute price level (Patinkin 1965, ch. 8).

Nor was the real-balance effect taken account of in Keynes's *General Theory* and in the subsequent Hicks (1937)–Modigliani (1944) IS–LM exposition of this theory, which rapidly became the standard one of macroeconomic textbooks. According to this exposition, the only way in which a decline in wages and prices can increase employment is by its effect in increasing the real value of money balances, hence reducing the rate of interest, and hence (through its stimulating effect on investment) increasing the aggregate demand for goods and hence employment. A further and basic tenet of this exposition was that there was a minimum below which the rate of interest could not fall. So if the wage decline were to bring about a lowering of the rate of interest to this minimum before full employment were reached, any further decline in the wage rate would be to no avail. In brief, the economy would then be caught in the 'liquidity trap'. And even though Keynes had stated in the *General Theory,* 'whilst this limiting case might become practically important in the future, I know of no example of it hitherto' (p. 207), the Keynesian theory of employment was for many years interpreted in terms of this 'trap'.

It was against this background that Pigou (1943, 1947) pointed out that the increase in the real value of money holdings generated by the wage and price decline increased the aggregate demand for goods directly, and not only indirectly through its downward effect on the rate of interest. Pigou's rationale was that individuals saved in order to accumulate a certain amount of wealth relative to their income, and that indeed the savings function depended inversely on the ratio of wealth to income. Correspondingly, as wages and prices declined, the real value of the monetary component of wealth increased and with it the ratio of wealth to income, causing a decrease in savings, which means an increase in the aggregate demand for consumption goods. Pigou's argument (which was formulated for a stationary state) thus had the far-reaching theoretical implication that even if the economy were caught in the 'liquidity trap', there existed a low enough wage rate that would generate a full-employment level of aggregate demand. In this way Pigou (1943, p. 351) reaffirmed the 'essential thesis of the classicals' that 'if wage-earners follow a competitive wage policy, the economic system must move ultimately to a full-employment stationary state'.

In his exposition and elaboration of Pigou's argument (which inter alia brought out the significance of the argument for dynamic stability analysis), Patinkin (1948) labelled the direct effect on consumption of an increase in the real value of money balances as the 'Pigou effect'. However, in subsequent recognition of the fact that this effect is actually an integral part of the quantity theory – as well as the fact that Pigou had been anticipated in drawing the implications of this effect for the Keynesian system by Haberler (1941, pp. 242, 389, and 403) and Scitovsky (1941, pp. 71–2) – Patinkin (1956, 1965) relabelled it the 'real-balance effect' and presented it as a component of the wealth effect.

In an immediate comment on Pigou's article, Kalecki (1944) pointed out that the definition of 'money' relevant for the real-balance effect is not the usual one of currency *plus* demand-deposits: for example, in the case of a price decline, the increase in the real value of the demand deposits has an offset in the corresponding increase in the real burden on borrowers of the loans they had received from the banking system. Thus (emphasized Kalecki) the monetary concept relevant for the real-balance effect in a gold-standard economy is only the gold reserve of the monetary system.

More generally, the relevant concept is 'outside money' (equivalent to the monetary base, sometimes also referred to as 'high-powered money'), which is part of the net wealth of the economy, as distinct from 'inside money', which consists of the demand deposits created by the banking system as a result of its lending operations and which accordingly is not part of net wealth (Gurley and Shaw 1960). This distinction was subsequently challenged by Pesek and Saving (1967), who contended that banks regard only a small fraction of their deposits as debt, so that these deposits too should be included in net wealth. In criticism of this view, Patinkin (1969, 1972a) showed that if perfect competition prevails in the banking system, the present value of the costs of maintaining its demand deposits equals the value of these deposits, so that the latter cannot be considered as a component of net wealth. This is also the case if imperfect competition with free entry prevails in the system. On the other hand, if – because of restricted entry – the banking sector enjoys abnormal profits, then the present value of these profits should be included in net wealth for the purpose of measuring the real-balance effect.

There remains the question of whether – for the purpose of measuring the real-balance effect – one should include government interest-bearing debt, as contrasted with the non-interest-bearing debt (viz., government fiat money) which is a component of the monetary base. Clearly, in a world of infinitely lived individuals with perfect foresight, the former does not constitute net wealth and hence is not a component of the real-balance effect: for the discounted value of the tax payments which the representative individual must make in order to service and repay the debt obviously equals the discounted value of the payments on account of interest and principal that he will receive. Nor is the assumption of infinitely lived individuals an operationally meaningless one: for as Barro (1974) has elegantly shown, if in making his own consumption plans, the representative individual with perfect foresight is sufficiently concerned with the welfare of the next generation to the extent of leaving a bequest for it, he is acting as if he were infinitely lived.

More specifically, Barro's argument is as follows: assume that an individual of the present generation achieves his optimum position by consuming C_0 during his lifetime and leaving a positive bequest of B_0 for the next generation. Clearly, such an individual could have increased his consumption to $C_0 + \Delta C_0$ and reduced his bequest to $B_0 - \Delta C_0$ – but preferred not to do so. Assume now that the individual also holds government bonds payable by the next generation, and let the real value of these bonds increase as the result of a decline in the price level, expected to be permanent. The revealed preference of the present generation for the consumption-bequest combination C_o, B_o implies that this increase in the real value of its holdings of government interest-bearing debt will not cause it to increase its consumption at the expense of the next generation. In brief, government debt in this case is effectively not a component of wealth and hence of the real-balance effect.

Needless to say, the absence of perfect foresight, and the fact that individuals might not leave bequests (as is indeed assumed by the life-cycle theory of consumption) means that government interest-bearing debt should to a certain extent be taken account of in measuring the real-balance effect – or what in this context is more appropriately labelled the 'net-real-financial-asset effect' (Patinkin 1965, pp. 288–94).

If we assume consumption to be a function of permanent income, and if we assume that the rate of interest which the individual uses to compute the permanent income flowing from his wealth is 10 per cent and the marginal propensity to consume out of permanent income before 0.80, then the marginal propensity to consume out of wealth (and out of real balances in particular) is the product of these two figures, or 0.08. However, in the case of consumers' durables (in the very broad sense that includes – besides household appliances – automobiles, housing, and the like), the operation of the acceleration principle implies an additional real-balance effect in the short run. In particular, assume that when the individual decides on the optimum composition of the portfolio of assets in which to hold his real wealth, W, he also considers the proportion, q, of these assets that he wishes to hold in the form of consumers durables, K_d, so that his demand for the *stock* of consumer-durable goods is $K_d = qW$. Assume now that wealth increases solely as a result of an increase in real balances, M/p. This leaves the representative individual with more money balances in relation to his other assets than he considers optimal. As a result he will attempt to shift out of money and into these other assets until he once again achieves an optimum portfolio. In the case of consumers' durables, this means that in addition to his preceding demand for new consumer-durable goods, he has a demand for

$$
\begin{aligned}
C_d = \Delta K_d &= q[\Delta(M/p)] \\
&= q\left[(M/p)_t - (M/p)_{t-1}\right]
\end{aligned}
$$

units, where $(M/p)_t$ represents real balances at time t. In general, the individual will plan to spread this additional demand over a few periods. In any event, once an optimally composed portfolio is again achieved, this additional effect disappears, so that the demand for new consumers' durables (which in the case of a stationary state is solely a replacement demand) will once again depend only on the ordinary real-balance effect as described at the beginning of this paragraph (Patinkin 1967, pp. 156–62).

It is, of course, true that the process of portfolio adjustment generated by the monetary increase will cause a reduction in the respective rates of return on the other assets in the portfolio, so that the initial wealth effect of the monetary increase will be followed by substitution effects. Now, Keynes limited his analysis in the *General Theory* to portfolios consisting only of money and securities; hence (as indicated above) an increase in the quantity of money could increase the demand for goods only indirectly through the substitution effect created by the downward pressure on the rate of interest. But once one takes account of the broader spectrum of assets held by individuals, one must also take account of the direct real-balance effect on the purchase of these other assets as well.

Various empirical studies have shown that the real-balance effect as here defined (viz., as part of the wealth effect) is statistically significant (Patinkin 1965, note M; Tanner 1970). Other studies have demonstrated the statistical significance of yet another definition of this effect: namely, as the effect on the demand for commodities of an excess supply of money, defined as the excess of the existing stock of money over its 'desired' or 'long-run' level (Jonson 1976; Laidler and Bentley 1983; cf. also Mishan 1958). It seems to me, however, that such a demand function is improperly specified: for though (as indicated above) the excess supply of money has a role to play in the consumption function (and particularly in that for consumers' durables), the complete exclusion of the real-balance effect *cum* wealth effect from the aforementioned demand function implies that in equilibrium there is no real-balance effect – an implication that is contradicted by the form of demand functions as derived from utility maximization subject to the budget constraint (Patinkin 1965, pp. 433–8, 457–60; Fischer 1981).

Granted the statistical significance of the real-balance effect, the question remains as to whether it is strong enough to offset the adverse expectations generated by a price decline – including those generated by the wave of bankruptcies that might well be caused by a severe decline. In brief, the question remains as to whether the real-balance effect is strong enough to assure the stability of the system: to assure that automatic market forces will restore the economy to a full-employment equilibrium position after an initial shock of a decrease in aggregate demand (Patinkin 1948, part II; 1965, ch. 14: 1). On the assumption of adaptive expectations, Tobin (1975) has presented a Keynesian model with the real-balance effect which under certain circumstances is unstable. On the other hand, McCallum (1983) has shown that under the assumption of rational expectations, the model is generally stable.

In any event, no one has ever advocated dealing with the problem of unemployment by waiting for wages and prices to decline and thereby generate a positive real-balance effect that will increase aggregate demand. In particular, Pigou himself concluded his 1947 article with the statement that such a proposal had 'very little chance of ever being posed on the chequer board of actual life'. Thus the significance of the real-balance effect is in the realm of macroeconomic theory and not policy.

Correspondingly, recognition of the real-balance effect in no way controverts the central message of Keynes's *General Theory.* For this message – as expressed in the climax of that book, Chapter 19 – is that the only way a general decline in money wages can increase employment is through its effect in increasing the real quantity of money, hence reducing the rate of interest, and hence stimulating investment expenditures; but that even if wages were downwardly flexible in the face of unemployment, this effect would be largely offset by the adverse expectations and bankruptcies generated by declining money wages and prices, so that the level of aggregate expenditures and hence employment would not increase within an acceptable period of time. In Keynes's words: 'the economic system cannot be made self-adjusting along these lines' (ibid., p. 267). And there is no reason to believe that Keynes would have modified this conclusion if he had also taken account of the real-balance effect of a price decline (Patinkin 1948, part III; 1976, pp. 110–11).

The above discussion has considered only the real-balance effect on the demand for goods. In principle, this effect also operates on the supply of labour: for the greater the real balances and hence wealth of the individual, the greater his demand for leisure as well, which means the smaller his supply of labour. This influence, however, has received relatively little attention in the literature (but see Patinkin 1965, p. 204; Phelps 1972; Barro and Grossman 1976, pp. 14–16).

Another limitation of the discussion is that it deals only with a closed economy. In the analysis of an open economy, the real-balance effect plays an important role in some of the formulations of the monetary approach to the balance of payments.

See Also

▶ Money Illusion
▶ Quantity Theory of Money

Bibliography

American Economic Association. 1951. *Readings in monetary theory*. Philadelphia: Blakiston.

Barro, R.J. 1974. Are government bonds net wealth? *Journal of Political Economy* 82: 1095–1117.

Barro, R.J., and H.I. Grossman. 1976. *Money, employment and inflation*. Cambridge: Cambridge University Press.

Fischer, S. 1981. Is there areal-balance effect in equilibrium? *Journal of Monetary Economics* 8: 25–39.

Gurley, J.G., and E.S. Shaw. 1960. *Money in a theory of finance*. Washington, DC: Brookings Institution.

Hicks, J.R. 1937. Mr Keynes and the 'classics': A suggested interpretation. *Econometrica* 5: 147–159. Reprinted in *Readings in the theory of income distribution*, Philadelphia: Blakiston for the American Economic Association, 1946, 461–476.

Jonson, P.D. 1976. Money and economic activity in the open economy: The United Kingdom, 1880–1970. *Journal of Political Economy* 84: 979–1012.

Kalecki, M. 1944. Professor Pigou on 'The classical stationary state': A comment. *Economic Journal* 54: 131–132.

Keynes, J.M. 1923. *A tract on monetary reform*. London: Macmillan.

Keynes, J.M. 1930. *A treatise on money*, The pure theory of money, vol. I. London: Macmillan.

Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.

Laidler, D., and B. Bentley. 1983. A small macro-model of the post-war United States. *The Manchester School* 51: 317–340.

Leontief, W. 1936. The fundamental assumption of Mr Keynes' monetary theory of unemployment. *Quarterly Journal of Economics* 51: 192–197.

McCallum, B.T. 1983. The liquidity trap and the Pigou effect: A dynamic analysis with rational expectations. *Economica* 50: 395–405.

Mishan, E.J. 1958. A fallacy in the interpretation of the cash balance effect. *Economica* 25: 106–118.

Modigliani, F. 1944. Liquidity preference and the theory of interest and money. *Econometrica* 12: 45–88. Reprinted in American Economic Association (1951), 186–240.

Patinkin, D. 1948. Price flexibility and full employment. *American Economic Review* 38: 543–564. Reprinted with revisions and additions in American Economic Association (1951), 252–283.

Patinkin, D. 1956. *Money, interest, and prices*. Evanston: Row, Peterson.

Patinkin, D. 1965. *Money, interest, and prices*, 2nd ed. New York: Harper & Row.

Patinkin, D. 1967. *On the nature of the monetary mechanism*. Stockholm: Almqvist and Wicksell. Reprinted in Patinkin (1972b), 143–167.

Patinkin, D. 1969. Money and wealth: A review article. *Journal of Economic Literature* 7: 1140–1160.

Patinkin, D. 1972a. *Money and wealth*. In Patinkin (1972b), 168–94.

Patinkin, D. 1972b. Studies in monetary economics. New York: Harper & Row.

Patinkin, D. 1976. *Keynes' monetary thought: A study of its development*. Durham: Duke University Press.

Pesek, B.P., and T.R. Saving. 1967. *Money, wealth and economic theory*. New York: Macmillan.

Phelps, E.S. 1972. Money, public expenditure and labor supply. *Journal of Economic Theory* 5: 69–78.

Pigou, A.C. 1943. The classical stationary state. *Economic Journal* 53: 343–351.

Pigou, A.C. 1947. Economic progress in a stable environment. *Economica* 14: 180–188. Reprinted in American Economic Association (1951), 241–251.

Scitovsky, T. 1941. Capital accumulation, employment and price rigidity. *Review of Economic Studies* 8: 69–88.

Tanner, J.E. 1970. Empirical evidence on the short-run real balance effect in Canada. *Journal of Money, Credit and Banking* 2: 473–485.

Tobin, J. 1975. Keynesian models of recession and depression. *American Economic Review* 65: 195–202.

von Haberler, G. 1941. *Prosperity and depression: A theoretical analysis of cyclical movements*, 3rd ed. Geneva: League of Nations.

Real Bills Doctrine

Roy Green

Keywords

Banking School; Bullionist controversy; Cantillon, R.; Convertibility; Currency School; Fractional reserve banking; Fullarton, J.; Land bank; Law of circulation; Law of reflux; Law, J.; Mercantilism; Overissue; Physiocracy; Real bills doctrine; Ricardo, D.; Smith, A.; Steuart; Sir J.; Thornton, H.; Tooke, T

JEL Classifications

E6

The 'real bills doctrine' has its origin in banking developments of the 17th and 18th centuries. It received its first authoritative exposition in Adam Smith's *Wealth of Nations*, was then repudiated by Thornton and Ricardo in the famous Bullionist Controversy, and was finally rehabilitated as the 'law of reflux' by Tooke and Fullarton in the currency–banking debate of the mid-19th century. Even now, echoes of the real bills doctrine reverberate in modern monetary theory.

The central proposition is that bank notes which are lent in exchange for 'real bills', that is, titles to real value or value in the process of creation, cannot be issued in excess; and that, since the requirements of the non-bank public are given and finite, any superfluous notes would return automatically to the issuer, at least in the long run. The grounds for rejecting the real bills doctrine have been many and varied. The main counter-argument is that overissue is not merely possible but inevitable in the absence of any external principle of limitation; in this view, commercial wants are insatiable and excess notes would not return to the issuer but undergo depreciation in the exact proportion to their excess.

By the time the real bills doctrine appeared in the economic literature, fractional reserve banking was already well established, releasing unproductive hoards for trade and investment. This did not satisfy John Law, that 'reckless, and unbalanced but most fascinating genius' (Marshall 1923, p. 41n.). He outlined a primitive real bills doctrine in the course of his proposal for a land bank, which would issue paper money on 'good security'. He imagined, however, that the need for a metallic reserve was superseded by the abolition of legal convertibility, and that *economic* convertibility would *always* be maintained by conformity with the real bills doctrine (Law 1705, p. 89).

The problem was that, as a mercantilist, Law identified money with capital; he believed that creating paper money was equivalent to increasing wealth. It was his attempt to 'break through' the metallic barrier that gave him 'the pleasant character mixture of swindler and prophet' (Marx 1894, p. 441). The spectacular collapse of Law's 'System' set off a negative reaction against financial innovation, which was reflected in Cantillon's 'anti-System' (Rist 1940, p. 73) and in Hume's opposition to what he called 'counterfeit money' (1752, p. 168). A more positive effect was a shift in the focus of political economy itself to the production process. This shift was led by the Physiocrats and by Adam Smith, whose 'original and profound' (Marx 1859, p. 168) analysis of money and banking was developed in the context of classical value theory.

A decade before the *Wealth of Nations,* Sir James Steuart had attempted to revive Law's ideas from a 'neo-mercantilist' viewpoint (1767, book IV, pt. 2). For Smith, by contrast, the role of bank credit was to increase not the quantity of capital but its *turnover* (1776, pp. 245–6; also Ricardo, *Works,* III, pp. 286–7). Output was fixed by the level of accumulation, which for all the classical economists included the speed of its turnover. Credit had the effect both of reducing the magnitude of reserve funds which economic agents needed to hold and of allowing the money material itself – treated as an element of circulating capital and an unproductive portion of the social wealth – to be displaced by paper, thus providing 'a sort of wagon-way through the air'.

Smith followed Law and Steuart, however, in arguing that an overissue of bank notes could not take place if they were advanced upon 'real' bills of exchange, that is, those 'drawn by a real creditor upon a real debtor', as opposed to 'fictitious' bills, that is, those 'for which there was properly no real creditor but the bank which discounted it, nor any real debtor but the projector who made use of the money' (1776, p. 239; also p. 231). When a banker discounted fictitious bills, the borrowers were clearly 'trading, not with any capital of their own, but with the capital which he advances to them'. When, on the other hand, real bills were discounted, bank notes were merely substituted for a substantial proportion of the gold and silver which would otherwise have been idle, and therefore available for circulation (p. 231). The quantity of notes was thus equivalent to the maximum value of the monetary metals that would circulate in their absence at a given level of economic activity (p. 227).

This development of the classical law of circulation applied to credit and fiduciary money alike, with the difference that in the latter case overissue in the 'short run' might result in a permanent depreciation of the paper. By contrast, credit- money, that is, banknotes which were exchanged for real bills, could never be in long-run excess:

> The coffers of the bank . . . resemble a water-pond, from which, though a stream is continually running out, yet another is continually running in, fully equal to that which runs out; so that, without any further care or attention, the pond keeps always equally, or very near equally full. (p. 231)

Only what Smith called 'over-trading' would upset this balance, by promoting excessive credit expansion and an accompanying drain of bullion.

Although the real bills doctrine was accepted by the Bank of England Directors as a guide to monetary management, it was challenged in the bullion controversy following the suspension of cash payments in 1797 as 'the source of all the errors of these practical men' (Ricardo, *Works,* III, p. 362; also Thornton 1802, p. 244 and *passim).* In the view of the 'bullionists',

> The refusal to discount any bills but those for *bona fide* transactions would be as little effectual in limiting the circulation; because, though the Directors should have the means of distinguishing such bills, which can by no means be allowed, a greater proportion of paper currency might be called into circulation, not than the wants of commerce could employ but greater than what could remain in the channel of currency without depreciation. (Ricardo, *Works,* III, p. 219.

Indeed, there was no other limit to the depreciation, and corresponding rise in the price level, 'than the will of the issuers' *(Works,* III, p. 226).

Nevertheless, the bullionist argument itself was open to challenge, because it confused money with credit. The inconvertible paper of the Bank Restriction was issued not as forced currency but on loan; it was therefore responsible not for increasing the money supply but simply altering its *composition*, by substituting one financial asset for another in the hands of the public. Only when cash payments were restored, however, was any further attempt made to rehabilitate the real bills doctrine, this time as the 'law of reflux': provided notes were lent on sufficient security, 'the reflux and the issue will, in the long run, always balance each other' (Fullarton 1844, p. 64; Tooke 1844, p. 60). The 'Banking School' called this law 'the great regulating principle of the internal currency' (Fullarton 1844, p. 68). Their opponents, the 'Currency School' orthodoxy, 'never achieved better than this average measure of security'; and, after all, the average 'is not to be despised' (Marx 1973, p. 131). The real bills doctrine made its next appearance in the Federal Reserve Act of 1913. In banking at least, discretion has always been the better part of valour.

See Also

- ► Banking School, Currency School, Free Banking School

Bibliography

Cantillon, R. 1755. *Essai sur la nature du commerce en general.* Trans. H. Higgs. London: Macmillan, 1931.

Fullarton, J. 1844. *On the regulation of currencies*. London: John Murray.

Hume, D. 1752. *Essays, literary, moral and political*. London: Ward, Lock & Co. n.d.

Law, J. 1705. *Money and trade considered*. Edinburgh: Anderson.

Marshall, A. 1923. *Money, credit and commerce*. London: Macmillan.

Marx, K. 1859. *A contribution to the critique of political economy.* Moscow: Progress Publishers, 1970.

Marx, K. 1894. *Capital*, vol. 3. Moscow: Progress Publishers, 1971.

Marx, K. 1973. *Grundrisse*. Harmondsworth: Penguin.

Ricardo, D. 1951–73. *The works and correspondence of David Ricardo*, ed. P. Sraffa. Cambridge: Cambridge University Press.

Rist, C. 1940. *History of monetary and credit theory from John Law to the present day.* London: Allen & Unwin.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations.* London*:* Routledge, 1890.

Steuart, J. 1767. *An inquiry into the principles of political oeconomy.* Edinburgh: Oliver & Boyd, 1966.

Thornton, H. 1802. *An enquiry into the nature and effects of the paper credit of Great Britain.* London: LSE Reprint Series, 1939.

Tooke, T. 1844. *An inquiry into the currency principle.* London: LSE Reprint Series, 1959.

Real Bills Doctrine Versus the Quantity Theory

Timothy S. Fuerst

Abstract

The real bills doctrine and the quantity theory of money represent distinct theoretical models of price-level determination and consequently imply different prescriptions for the conduct of monetary policy. The real bills doctrine takes the price level as exogenous and recommends money supply movements that passively respond to the economy. In sharp contrast, the quantity theory insists that the only way to ensure price level stability is by constraining the money supply and not allowing the money supply to move passively in response to economic conditions.

Keywords

Central banks; Credit; Equation of exchange; Federal Reserve System; Fiat money; Fisher, I.; Gold standard; Great Depression; Hume, D.; Inflation; Interest rate policy; Law, J.; Monetary policy; Money supply; Neutrality of money; Nominal interest rates; Quantity theory of money; Real bills doctrine; Real bills doctrine versus the quantity theory; Ricardo, D.; Smith, A.; Thornton, H.; Velocity of circulation

JEL Classifications

E31; E52; E6

Drawing on two very different hypotheses about the link between nominal money and economic activity, the real bills doctrine and the quantity theory of money represent sharply divergent advice on the conduct of monetary policy. The quantity theory has many prominent advocates, but the real bills doctrine has had a dominant influence in the history and practice of central banking. Further, the real bills doctrine was at the core of the Congressional act creating the US Federal Reserve System so that its importance echoes down to the current day.

The real bills doctrine views money as playing a decidedly passive role, calling for monetary expansion in line with economic activity. According to this view, economic activity is linked to business trade credit and the issuance of short-term debt instruments. Banks should freely purchase these 'real bills' with banknote issue, where the modifier 'real' refers to short-term debt instruments used to finance productive activity as opposed to speculation. The doctrine dates to at least 1705 with the publication of *Money and Trade Considered* by John Law, who suggested that banknote issue should be secured by and thus linked to the nominal value of land. The most famous statement of the doctrine is by Adam Smith, whose linkage of note issue to bills of exchange gave the doctrine its name:

> When a bank discounts to a merchant a real bill of exchange drawn by a real creditor upon a real debtor, and which, as soon as it becomes due, is really paid by that debtor; it only advances to him a part of the value which he would otherwise be obliged to keep by him unemployed, and in ready money for answering occasional demands. The payment of the bill, when it becomes due, replaces to the bank the value of what it had advanced, together with the interest. The coffers of the bank, so far as its dealings are confined to such customers, resemble a water pond, from which, though a stream is continually running out, yet another is continually running in, fully equal to that which runs out; so that, without any further care or attention, the pond keeps equally, or very nearly full. (1776, p. 304)

Smith's water-pond metaphor illustrates the real-bills view that note issue would be self-regulating when tied to economic activity, that is, money issue could never be excessive when issued against short-term commercial bills.

The fundamental criticism of the real bills doctrine is that the value of commercial bills (or, in Law's case, the value of land) is tied proportionately to the price level. A commercial bill necessarily includes the dollar value of the goods or services to which it is linked. Thus, under the real bills doctrine, nominal note issue is tied to the nominal price level. If the price level is influenced by the money supply, then we have a circularity problem: nominal prices determine note issue, and note issue affects prices. Henry Thornton first noted the danger of this inflationary circle in his 1802 *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*. (David Ricardo was also a prominent opponent of the doctrine.) The thrust of Thornton's criticism was that the real bills doctrine provided no limit on banknote issue. Smith seems to have avoided Thornton's criticism because in Smith's system the gold standard provided an overall restraint on note issue. An excessive banknote issue would result in a bank losing its gold holdings, and see a drain on its 'coffers'. (See Laidler 1981, 1984, for a defence of Smith.) But in a world with an inconvertible paper currency Thornton's inflationary critique is devastating.

Humphrey (2001) provides an algebraic description of the real bills doctrine. Suppose that the needs for trade credit are proportional to nominal production, PY, where P denotes the price level and Y denotes real production. The real bills doctrine would imply that banknote issue and thus the money supply (M) should be proportionally linked to the needs of trade credit so that we have:

$$M = kPY$$

where k is the constant of proportionality between trade credit and nominal production. The Thornton inflationary critique is now obvious: even with an exogenous level of output (Y), there is no way of determining the two endogenous variables, the money supply (M) and price level (P). A real bills counter-argument would be that the price level is exogenous to money, that is, the money supply has no direct effect on prices. As discussed below, the quantity theory makes the exact opposite claim.

The Real Bills Doctrine and the Great Depression

Remarkably, the real bills doctrine survived Thornton and Ricardo's withering 19th century criticism to find a central place in 20th century US monetary history. In a fascinating account, Meltzer (2003) and Humphrey (2001) trace the flowering of the real bills doctrine into the US Federal Reserve Act of 1913. US Federal Reserve Banks existed for the purpose of 'accommodating commerce and business' and were supposed to discount only 'eligible paper', which the Act defined as 'notes, drafts, and bills of exchange arising out of actual commercial transactions'. Although, like Adam Smith, the Act presumed the existence of the gold standard, the real bills doctrine was deemed sufficient even in the absence of a specie constraint. For example, in the *Tenth Annual Report* (1924) of the Board of Governors of the Federal Reserve System, it is noted that 'there is little danger that the credit created and distributed by the Federal Reserve Banks will be in excessive volume if restricted to productive issues' (1924, p. 28). The Report further suggested no link between money and prices: 'The interrelationship of prices and credit is too complex to admit of any simple statement' (1924, p. 32). Adolph Miller, a founding member of the Federal Reserve Board and co-author of the Report, rejected the notion that 'changes in the level of prices are caused by changes in the volume of credit and currency.. .or that changes in the volume of credit and currency are caused by Federal Reserve policy' (quoted in Meltzer 2003, pp. 187–8).

Meltzer (2003) convincingly argues that it was this belief in the self-regulating nature of the real bills doctrine that led the Federal Reserve to stand idly by as the US economy spiralled into the Great Depression in the early 1930s. From a real-bills perspective, monetary policy was very loose during these years because Reserve Banks stood ready to discount bills at historically low nominal rates of interest. Meltzer (2003, p. 321) concludes that

> the real bills doctrine implied that the correct policy was a passive one. Most [Federal Reserve] governors had always held these views ... The economies of the United States and much of the rest of the world became victims of the Federal Reserve's adherence to an inappropriate theory and the absence of basic economic understanding such as that developed by [Henry] Thornton and [Irving] Fisher.

The Quantity Theory

In sharp contrast to the real bills doctrine, the quantity theory held as its fundamental principle that the quantity of nominal money (M) is largely exogenous and is the principal force determining the endogenous price level (P). This argument was first articulated by David Hume (1752). An immediate corollary is that changes in the price level, that is, inflation, are primarily determined by movements in the supply of money. In the words of the celebrated quantity theorist Milton Friedman (1956, pp. 20–1):

> there is perhaps no other empirical relation in economics that has been observed to recur so uniformly under so wide a variety of circumstances as the relation between substantial changes over short periods in the stock of money and in prices; the one is invariably linked with the other and is in the same direction; this uniformity is, I suspect, of the same order as many of the uniformities that form the basis of the physical sciences.

The quantity theory's causal link between M and P included the concept of long-run monetary neutrality: exogenous changes in M would eventually be exactly matched by proportional changes in P. This inference is grounded on the stability of real money demand. In the words of Friedman: 'The quantity theory is in the first instance a theory of the demand for money' (1956, p. 4); 'The quantity theorist accepts the empirical hypothesis that the demand for money is highly stable – more stable than functions such as the consumption function that are offered as alternative key relations' (1956, p. 16). If we let $L(R,Y)$ denote real money demand as a function of the nominal interest rate (R) and the level of real production (Y), we have a money market equilibrium condition given by:

$$L(R,\ Y) = \frac{M}{P}.$$

The proportionality hypothesis is then quite clear: for a stable level of L, exogenous changes in M must be matched by changes in P of the exact same magnitude.

The quantity theory also included the concept of short-run non-neutrality. In the words of Hume (1752, p. 38):

> When any quantity of money is imported into a nation, it is not at first disposed into many hands but is confined to the coffers of a few persons, who immediately seek to employ it to advantage ... It is easy to trace the money in its progress through the whole commonwealth, where we shall find that it must first quicken the diligence of every individual before it increase the price of labour.

'There is always an interval before matters be adjusted to their new situation' (1752, p. 40). Quantity theorists would argue that increases in M are initially met by increases in production (Y) and declines in interest rates (R), but that in the long run R and Y would return to their original levels and that P would thus fully reflect the new higher level of M.

The quantity theory is closely associated with the quantity equation which can be derived as follows. The previous money demand relationship can be re-written as

$$M\frac{Y}{L(R,Y)} = PY.$$

If we define the velocity of money as

$$V \equiv \frac{Y}{L(R,Y)}$$

then we can write this relationship as the celebrated quantity equation:

$$MV = PY.$$

This is Pigou's (1927) variant of Irving Fisher's (1922) classic equation of exchange. The quantity equation is a useful device for expositing the two central tenets of the quantity theory of money: (a) in the long run, output (Y) and velocity (V) are exogenous to money, so that exogenous movements in the money supply (M) are met by proportional movements in prices (P), and (b) in the short run, movements in the money supply are met by some combination of movements in velocity, prices and output, so that changes in M have non-neutral effects on output. The quantity equation can also be used to illustrate Thornton's inflationary critique of the real bills doctrine. For a given level of the nominal rate and an exogenous level of production, velocity is determined by the money demand function, but there is no restriction on the size of M or the size of P.

The Contemporary Policy Debate

From the vantage point of the outset of the 21st century, there is a sense in which the quantity theory has won numerous intellectual battles but lost the war. Most economists subscribe to the principles of long-run monetary neutrality and short-run non-neutrality. Most would also agree that the quantity equation can be a useful intellectual organizing device. Finally, a standard result in any monetary theory course is the nominal indeterminacy that arises under an exogenous interest-rate operating procedure (for example, Sargent 1987, ch. 4). This result is just the modern statement of Thornton's 1802 criticism of the real bills doctrine. Hence, it would appear that the quantity theory is in the ascendant.

But remnants of the real bills doctrine are pervasive in both monetary policy implementation and theoretical work. In terms of policy, essentially all central banks in the industrialized world typically ignore or downplay movements in monetary aggregates and instead conduct monetary policy according to an interest rate operating procedure, a close descendant of a real-bills policy. The rationale for such a policy choice is the assertion that the demand for money and thus velocity are unstable. Such a policy implies seasonal movements in monetary aggregates to accommodate movements in real activity, a passive money supply movement that is directly out of a real-bills playbook.

From a theoretical perspective, there have been two prominent recent contributions in favour of interest rate policy. First, Sargent and Wallace (1982) provide something of a rehabilitation of the real bills doctrine by developing a model in which fluctuating nominal interest rates are harmful, and in which a policy of pegging the nominal interest rate at zero is Pareto efficient. Second, Woodford (2003) has pioneered an effort to conduct monetary policy analysis in 'cashless' models – models in which the price level is well defined even though there is no money in the model and the central bank follows an interest-rate operating procedure. We review each of these contributions in turn.

Sargent and Wallace (1982) consider a two-period-lived overlapping-generations model in which fiat money is held even though nominal interest rates are positive because of a legal restriction on private real lending. There are three types of agents: poor savers, rich savers, and borrowers. Using their logarithmic preference specification, the two classes of savers have a constant desired level of savings, say, S^P for the poor and $S^R \gg S^P$ for the rich. The borrowers have a demand for loans given by

$$D^L = \frac{D}{1+r}$$

where r is the real interest rate, and $D > S^R$. (Sargent and Wallace 1982, consider the case in which the demand for loans fluctuates deterministically, but this is unimportant for their basic result.) The legal restriction is that borrowers cannot issue small-denomination notes. Hence, poor savers cannot lend directly to the borrowers, but can only save by accumulating fiat money. The equilibrium conditions for the money and credit markets are given by:

$$\text{Money market: } S^P \frac{M_t}{P_t}$$

$$\text{Credit market: } S^R = \frac{D}{1+r_t}$$

where M_t and P_t denote the time-t money supply and price level, and r_t is the real rate of interest. Under what Sargent–Wallace call a 'quantity-theory' regime, the central bank keeps the money supply fixed at some $M_t = M$. In this case, the price level and the real interest rate are constant and calculated from the above equilibrium conditions. This equilibrium is clearly not Pareto optimal as agents do not face the same inter-temporal rate of return – that is, rich savers earn a return of $r > 0$, while poor savers earn a zero real return on currency holdings.

Under a ‘real-bills’ regime the central bank stands ready to lend cash at a zero nominal rate of interest so that

$$P\,(1 + r_t) = \frac{P_t}{P_{t+1}}.$$

In particular, the central bank purchases the ‘real bills’ issued by the borrowers. To finance these purchases the central bank creates the new fiat money denoted by N_t. The borrowers can then use this cash to purchase goods from the poor savers. By purchasing the borrowers’ bonds with fiat money, the central bank is effectively opening up an avenue by which poor savers can lend to borrowers. Without this central bank intervention, the positive nominal rates in the credit market are symptoms of a problem – the inability of a fixed money stock to promote proper credit allocation. The real bills equilibrium conditions are given by:

$$\text{Money market: } S^P \frac{M_t \; + N_t}{P_t}$$

$$\text{Credit market: } \frac{N_t}{P_t} + S^R = \frac{D}{(P_t/P_{t+1})}.$$

Combining, we have that an equilibrium under the real-bills regime is defined by a price sequence that satisfies:

$$S^P + S^R = \frac{M_t}{P_t} + \frac{D}{(P_t/P_{t+1})}.$$

Solving, we have:

$$P_t = \left(\frac{D}{S^P + S^R}\right) P_{t+1} + \left(\frac{1}{S^P + S^R}\right) M_t.$$

Assuming $D < (S^P + S^R)$, the set of stationary equilibria are given by

$$P_t = \left(\frac{1}{S^P + S^R}\right) \sum_{j=0}^{\infty} \left(\frac{D}{S^P + S^R}\right)^j M_{t+j}$$

where the path of the money supply is free. In the special case in which the money supply grows at a constant rate g we have

$$P_t \; = \left(\frac{1}{S^P \; + S^R - D(1+g)}\right) M_t \; .$$

Note that, if g becomes large enough, the monetary equilibrium disappears.

Sargent and Wallace restrict the analysis to a particular equilibrium in which the beginning-of-period money supply is held fixed, $M_t = M$ for $t =$ 0; 1; 2; 3 .. .. However, the money supply grows and contracts *within* each period as the central bank accommodates the supply of one-period bonds issued by the borrowers (‘real bills’) with the passive expansion of N_t. In this equilibrium the price level is constant and the real return on savings is zero. This equilibrium is Pareto efficient, in contrast to the Pareto inefficiency of the quantity-theory regime. This is an argument in favour of the real bills doctrine and represents Sargent and Wallace’s rehabilitation of the doctrine.

There are difficulties with this conclusion. First, the real-bills equilibrium selected by Sargent and Wallace does not Pareto-dominate the quantity-theory regime (rich savers are worse off under the real-bills regime). Second, there is an infinite number of other real-bills equilibria, all defined by the behaviour of the money stock, and not all of these are Pareto efficient. For example, if the money supply grows at a constant rate $g > 0$ the real-bills equilibrium is not Pareto efficient. Finally, Thornton’s inflationary critique of the real-bills regime endures: since the money supply is entirely free, there are no restrictions on the short-term and long-term price level.

The second body of recent theoretical work that has a real-bills flavour is provided by Woodford (2003). The title of Woodford’s treatise is *Interest and Prices*, a title that makes clear a principal assertion in the work: the money supply is largely irrelevant to price-level determination. The key relationship in the work is the Fisher equation linking nominal rates (i_t) to inflation rates and real rates (r_t):

$$i_t = r_t + p_{t+1} - p_t$$

where p_t is the log of the price level. For simplicity let us suppose that the real rate is exogenous. If the central bank conducts policy according to an exogenous nominal interest rate policy, then the Fisher equation uniquely determines the growth rate of prices (the inflation rate), but not the level of prices. This is, again, the Thornton critique of the real bills doctrine. But Woodford assumes that the central bank follows an endogenous interest rate policy in which the nominal rate responds to movements in prices:

$$i_t = \alpha p_t.$$

Assuming that $\alpha > 0$, the unique stationary equilibrium is given by:

$$p_t = \sum_{j=0}^{\infty} \left(\frac{1}{1+\alpha} \right)^{j+1} r_{t+j}.$$

From a quantity-theory perspective this is a remarkable conclusion: the price level is determined without any mention being made of the money supply. Where is the money demand curve? Either it does not matter (as the money supply moves passively to hit the interest rate target) or it does not even exist (a 'cashless' world). Woodford's (2003) analysis thus rejects the quantity theory as a useful guide for policy, and at the same time provides a 21st-century response to Thornton's 19th-century critique of the real bills doctrine: the money supply should be adjusted passively to hit the interest-rate target (as under a real-bills policy), but the interest-rate target should be moved endogenously to ensure price-level stability.

In the intellectual clash of ideas there are typically no clear winners or losers, but instead a synthesis of the combatants. This is surely true of the debate between the real-bills doctrine and the quantity theory of money. Current monetary policy practice and theory has a notable real-bills flavour in the near-universal use of interest rates as the operating target. To repeat, the advantage of such a policy is that it allows the money supply to respond automatically to and thus accommodate natural movements in real economic activity. But Thornton and the quantity theorists provide a cautionary critique: under an exogenous interest rate policy, there is no way of limiting the inflationary circle between note issue and the price level. To respond to this quantity-theory critique, Woodford (2003) and others have proposed an endogenous interest-rate policy of the form outlined above. This is just one manifestation of the synthesis of the two combatants in this intellectual debate.

See Also

▶ Monetarism
▶ Quantity Theory of Money
▶ Real Bills Doctrine

Bibliography

Board of Governors of the Federal Reserve System. 1924. *Tenth annual report of the federal reserve board: Covering operations for the year 1923*. Washington, DC: Government Printing Office.

Fisher, I. 1922. *The purchasing power of money, reprinted*. 2 ed, 1963. New York: August M. Kelley.

Friedman, M. 1956. The quantity theory of money: A restatement. In *Studies in the quantity theory of money*. Chicago: University of Chicago Press.

Hume, D. 1752. Of interest; of money. In *Writings on economics*, ed. E. Rotwein, 1970. Madison: University of Wisconsin Press.

Humphrey, T.M. 1974. The quantity theory of money: Its historical evolution and role in policy debates. *Federal Reserve Bank of Richmond Economic Review* 1974-(May/June): 2–19.

Humphrey, T.M. 1982. The real bills doctrine. *Federal Reserve Bank of Richmond Economic Review* 1982-(September/October): 3–13.

Humphrey, T.M. 2001. Monetary policy frameworks and indicators for the federal reserve in the 1920s. *Federal Reserve Bank of Richmond Economic Quarterly* 87(1): 65–92.

Laidler, D. 1981. Adam Smith as a monetary economist. *Canadian Journal of Economics* 14: 185–200.

Laidler, D. 1984. Misconceptions about the real bills doctrine: A comment on Sargent and Wallace. *Journal of Political Economy* 92: 149–155.

Law, J. 1705. *Money and trade considered*. Edinburgh: Anderson.

Meltzer, A.H. 2003. *A history of the federal reserve, Volume I: 1913–1951*. Chicago: University of Chicago Press.
Pigou, A.C. 1917. The value of money. *Quarterly Journal of Economics* 32(November): 38–65.
Pigou, A.C. 1927. *Industrial fluctuations*. London: Macmillan.
Sargent, T.J. 1987. *Macroeconomic theory*. 2 ed. Orlando: Academic Press.
Sargent, T.J., and N. Wallace. 1982. The real-bills doctrine vs. the quantity theory: A reconsideration. *Journal of Political Economy* 90: 1212–1236.
Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*, 1976. Indianapolis: Liberty Press.
Thornton, H. 1802. *An enquiry into the nature and effects of the paper credit of Great Britain*, 1939. London: LSE Reprint Series.
Woodford, M. 2003. *Interest and prices*. Princeton: Princeton University Press.

Real Business Cycles

Ellen R. McGrattan

Abstract

Real business cycles are recurrent fluctuations in an economy's incomes, products, and factor inputs – especially labour – that are due to non-monetary sources. These sources include changes in technology, tax rates and government spending, tastes, government regulation, terms of trade and energy prices. Most real business cycle (RBC) models are variants or extensions of a neoclassical growth model. One such prototype is introduced. It is then shown how RBC theorists, applying the methodology of Kydland and Prescott (Econometrica 50:1345–1370, 1982), use theory to make predictions about actual time series. Extensions of the prototype model, current issues and open questions are also discussed.

Keywords

Competitive equilibrium; Depreciation; Great depression; Home production; Household budget constraint; International business cycles; Labour supply; Labour-market search; Markov processes; Monetary shocks; Productivity shocks; Real business cycles; Real exchange rates; Regulation; Research and development; Stabilization policies; Stochastic growth models; Technical change; Technology shocks; Terms of trade; Total factor productivity; Unemployment

JEL Classification

D4; D10

Real business cycles are recurrent fluctuations in an economy's incomes, products, and factor inputs – especially labour – that are due to non-monetary sources. Long and Plosser (1983) coined the term 'real business cycles' and used it to describe cycles generated by random changes in technology. Other real sources of fluctuations that have been studied include changes in tax rates and government spending, tastes, government regulation, terms of trade, and energy prices.

Kydland and Prescott (1982), who studied the *quantitative* predictions of a stochastic growth model with shocks to technology, found that covariances between model series and autocorrelations of model output were consistent with corresponding statistics for US data. These findings were viewed as surprising, for two reasons. First, the findings ran counter to the idea that monetary shocks are the driving force behind business cycle fluctuations. Second, the policy implication for Kydland and Prescott's model was that stabilization policies are counterproductive. Fluctuations arise when households optimally respond to changes in technology.

The methodology that Kydland and Prescott (1982) used in their study of business cycles transformed the way in which applied research in macroeconomics is done. For this reason, the term 'real business cycles' is often associated with a methodology rather than Kydland and Prescott's original findings. Indeed, the methods of their 1982 paper have been used to study many different sources of business cycles, including monetary shocks.

R

Most real business cycle (RBC) models are variants or extensions of a neoclassical growth model. One such prototype is introduced. It is then shown how RBC theorists, following Kydland and Prescott (1982), use theory to make predictions about actual time series. Extensions of the prototype model are discussed. Current issues and open questions follow.

Prototype Real Business Cycle Model

Households choose sequences of consumption and leisure to maximize expected discounted utility. When aggregated, preferences are defined for a stand-in household that maximizes the expected value of

$$\sum \beta^t u(c_t, 1 - h_t) N_t, \tag{1}$$

where u is the utility function, c_t is per capita consumption at date t, $1 - h_t$ is per capita leisure at date t, N_t is the population at date t which grows at rate η, and β is a discount factor.

The technology available in period t is $z_t F_t(K_t, H_t)$, where $z_t F_t$ is the output produced at date t with K_t units of capital and H_t hours. The function F_t has constant returns to scale so that doubling the inputs doubles the output. The variable z_t is a stochastic technology shock assumed to follow a Markov process. The variation in z modelled here is variation in the effectiveness of factor inputs, capital and labour, to produce final goods and services or *total factor productivity* (TFP). Fluctuations in TFP arise from many possible sources. For example, improvements in TFP can arise from new inventions or innovations in existing production processes. Reductions in TFP can arise from increased regulation on producers.

Households are endowed with time each period, normalized without loss of generality to 1, which they can allocate to work or to leisure. They can invest x_t (per capita) in new capital goods. Doing so yields

$$N_{t+1} k_{t+1} = N_t[(1 - \delta)k_t + x_t], \tag{2}$$

where k_t is per capita beginning-of-period t capital, k_{t+1} is per capita end-of-period t capital, and δ is the rate of per period depreciation.

Households face taxes on purchases of consumption and investment and on incomes to capital and labour. With taxation, the household budget constraint in period t is

$$\begin{aligned}(1 + \tau_{ct})c_t &+ (1 + \tau_{xt})x_t \\ &= r_t k_t - \tau_{kt}(r_t - \delta)k_t + (1 - \tau_{ht})w_t h_t \\ &\quad + \psi_t.\end{aligned} \tag{3}$$

Variables r_t and w_t are pre-tax payments to capital and labour, respectively. Variables τ_{ct}, τ_{xt}, τ_{kt}, and τ_{ht} are tax rates on consumption, investment, capital, and labour, respectively. These tax rates are assumed to be stochastic and follow a Markov process. Variable ψ_t is the per capita transfer payment at date t made by the government to each household. Total transfer payments are equal to tax revenues less total spending by the government. The per capita spending of the government at date t is g_t.

To derive explicit predictions about the behaviour of these households, it is necessary to first define and then compute an equilibrium for the economy. In doing so, it is convenient to de-trend any variables that grow over time and deal only with stationary processes. To be precise, assume that there is a constant rate of improvement in production processes over time so that $F_t(K_t, H_t) \equiv F\left(K_t, (1+\gamma)^t H_t\right)$ with F homogeneous of degree 1. If the per capita capital stock grows at rate γ and z_t and h_t are stationary, then output grows at rate γ. Certain assumptions on utility and the process for government spending also ensure that components of output grow at rate γ. Denote by $\tilde{v}_t$ the de-trended level of variable v_t, that is, $\tilde{v}_t = v_t/(1+\gamma)^t$.

A competitive equilibrium is defined as household policy functions for consumption $c\left(\tilde{k}, \tilde{K}, s\right)$, investment $x\left(\tilde{k}, \tilde{K}, s\right)$, and hours $h\left(\tilde{k}, \tilde{K}, s\right)$, where $\tilde{k}$ is the (de-trended) stock of capital for the household, $\tilde{K}$ is the (de-trended) aggregate stock of capital, and $s = (\log z, \tau_c, \tau_x, \tau_k, \tau_h, \log \tilde{g};)$ pricing functions $w\left(\tilde{K}, s\right)$ and $r\left(\tilde{K}, s\right)$; a function governing the evolution of the aggregate capital stock $\tilde{K}' = \overline{\psi}(\tilde{K}, s)$ that maps the current state into the capital stock next period $(\tilde{K})$, and a function $\Phi(s', s)$ governing the transition of the stochastic shocks from s to s' such that (*a*) households maximize the expected value of (1) subject to (2) and (3) with the initial capital stock $\tilde{k}_0$ and functions for prices, aggregate capital, and the transition of s taken as given; (*b*) productive factors are paid their marginal products; (*c*) expectations are rational so that $\tilde{k} = \tilde{K}$ and

$$\overline{\psi}(\tilde{k}, s) = \begin{array}{l}[(1-\delta)\tilde{k} + x(\tilde{k}, s)] \\ \times/[(1+\eta)(1+\gamma)];\end{array}$$

and (*d*) markets clear:

$$c(\tilde{k}, \tilde{k}, s) + x(\tilde{k}, \tilde{k}, s) + g(s) = z(s)F(\tilde{k}, h).$$

Note that, in forming expectations about the future, households take processes for prices, tax rates and transfers as given. If households behave competitively, they assume that their own choice of capital next period does not affect the economy-wide level of capital. Therefore, in computing optimal decision functions for the household, it is necessary to distinguish the household's holdings of capital and the aggregate holdings of capital.

Comparing Model Predictions with Data

Given equilibrium functions, properties of the model time series can be compared with data in a straightforward way. Starting with initial conditions on the state, the evolution of the state is determined by functions Ψ and Φ, resulting in sequences $\{\tilde{k}, s\}_{t=0}^{\infty}$ for the state. Equilibrium price and decision functions are then used with these sequences for the state to determine sequences of all prices and allocations.

A standard assumption for the transition $\Phi(s', s)$ is the vector autoregression

$$s_{t+1} = P_0 + Ps_t + Q\varepsilon_{t+1},$$

where each element of ε_t is a normally distributed random variable, independent of the other elements of ε and across time, with mean equal to zero and variance equal to 1. Allowing non-zero off-diagonals in the matrices P and Q allows for correlations in the elements of the vector s. For example, a standard assumption is that tax rates and spending are positively correlated.

If the elements of the matrix QQ' are not large, the equilibrium evolution of the capital stock is well approximated by the following function:

$$\log\tilde{k}_{t+1} = A_0 + A_k\log\tilde{k}_t + B_k s_t,$$

which is linear in the log of the de-trended, per capita capital stock and the stochastic states. Similarly, the logarithms of consumption, investment, output and hours of work can be well approximated as linear functions of log $\tilde{k}$ and s_t. (See Marimon and Scott 1999, for an introduction to log-linear methods and nonlinear methods.) Stacking the results in matrix form yields a system of equations

$$X_{t+1} = AX_t + B\varepsilon_{t+1}$$

$$Y_t = CX_t + \omega_t,$$

where X contains all variables of interest, some of which may not be observable, and Y is a vector of observables. This system can be easily simulated and lends itself nicely to standard methods of estimating model parameters. (See Anderson et al. 1996.)

An important feature of the analysis in Kydland and Prescott (1982) was the construction of the same statistics for the model and for the US data. Employing this methodology requires two necessary steps. The first concerns measurement: data series must be consistent with model series. For example, consumer durable expenditures are investments much like expenditures on new housing. National accountants treat expenditures on durables and housing differently, but the prototype model does not. Thus, revising the national accounts to include services, rents and depreciation of durables is necessary for data and model series to be consistent. The second step of Kydland and Prescott's (1982) methodology concerns reporting: the same statistics should be computed for the model and the revised data. Such comparisons are useful in highlighting similarities and deviations, which are both necessary ingredients to further the development of good theory.

Applying the two methodological tenets to the prototype model and US data yields a number of interesting results. Both the theory and the US data display pro-cyclical movements in consumption and investment, with the movements in investment being far greater in percentage terms. With tax rates and government spending fixed at mean US levels, the theory predicts fluctuations in per capita hours that are too smooth relative to US hours, and a correlation between hours worked and productivity that is too high relative to the correlation in US data. When fiscal shocks consistent with US policy are introduced, the theory predicts movements in per capita hours and a correlation between hours worked and productivity that are in line with the data.

Extensions of the Prototype

During the 1980s and 1990s, business cycle research was exploratory but methodologically rooted. Researchers investigated the effects of many different shocks, the mechanisms that propagate them, and the welfare implications – in a consistent way that made clear what factors were important and why. A brief history is provided here, but interested readers are referred to the volume edited by Cooley (1995) and to a summary of more recent work in King and Rebelo (1999) and Rebelo (2005).

Kydland and Prescott (1982) and Long and Plosser (1983) emphasize technology shocks as an important source of fluctuations. Greenwood et al. (1988) also explore the role of technology shocks for the business cycle but restrict attention to technological changes affecting the productivity of new capital goods and allow for accelerated depreciation of old capital. Mendoza (1995) includes shocks to the terms of trade in an international business cycle model and shows that responses of real exchange rates to productivity shocks and terms-of-trade shocks are quite different, both qualitatively and quantitatively. Braun (1994), Christiano and Eichenbaum (1992), and McGrattan (1994) add fiscal shocks which are important for movement in hours and labour productivity, as noted above. Kim and Loungani (1992) add shocks to energy prices and show that the addition has only a modest impact on the variability of output and hours. Cooley and Hansen (1989) include monetary shocks and a cash-in-advance constraint and show that these additions have negligible effects on business cycle predictions.

The original technology-driven business cycle models under-predicted fluctuations in observed hours and over-predicted the correlation between hours and productivity, leading to further investigations of the model of the labour market and alternative mechanisms for propagating shocks. High – possibly infinite – elasticities were required in the original RBC models to generate fluctuations in aggregate hours comparable to the data. Rogerson (1988) motivates an infinite aggregate elasticity of labour supply in a world with variation in the fraction of people working: individuals work a standard workweek or not at all. This idea is implemented in an RBC model by Hansen (1985), who finds a significant increase in hours fluctuations relative to Kydland and Prescott (1982).

Another factor affecting the labour market is explored by Benhabib et al. (1991) and Greenwood and Hercowitz (1991) who introduce home production. These researchers show that business cycle predictions depend crucially on the willingness and opportunity of households to substitute time in home work and market work. Under plausible parameterizations, the models do in fact generate greater variability of hours and lower correlations between hours and productivity.

The empirical performance of the RBC model is also improved when labour-market search frictions are introduced, as in Andolfatto (1996) and Merz (1995). Labour-market search models have also been used to study movements in unemployment and vacancies.

Current Research and Open Questions

RBC research has evolved beyond the study of business cycles. The methodology that

Kydland and Prescott (1982) introduced is now being applied to central questions in labour, finance, public finance, history, industrial organization, international macroeconomics and trade.

Within business cycle research, some open questions remain. What is the source of large cyclical movements in TFP? This question is especially interesting in the case of the US Great Depression, when TFP declined significantly (Cole and Ohanian 2004). Are movements in TFP primarily due to new inventions and processes that are, by the nature of research and development, stochastically discovered? Or are movements in TFP primarily due to changing government regulations that may alter the efficiency of production? Are they due to unmeasured investments that fluctuate over time? The answers matter for policymakers, and they matter for economists who calculate the welfare costs or gains of changing policies.

See Also

- ▶ Business Cycle Measurement
- ▶ International Real Business Cycles
- ▶ Monetary Business Cycle Models (Sticky Prices and Wages)
- ▶ Political Business Cycles
- ▶ Welfare Costs of Business Cycles

Acknowledgment *I wish to thank Gary Hansen, Lee Ohanian and Ed Prescott for their comments on an earlier draft. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.*

Bibliography

Anderson, E.W., L.P. Hansen, E.R. McGrattan, and T.J. Sargent. 1996. Mechanics of forming and estimating dynamic linear economies. In *Handbook of computational economics*, vol. 1, ed. H. Amman, D. Kendrick, and J. Rust. Amsterdam: North-Holland.

Andolfatto, D. 1996. Business cycles and labor-market search. *American Economic Review* 86: 112–132.

Benhabib, J., R. Rogerson, and R. Wright. 1991. Homework in macroeconomics: Household production and aggregate fluctuations. *Journal of Political Economy* 99: 1166–1187.

Braun, R.A. 1994. Tax disturbances and real economic activity in the postwar United States. *Journal of Monetary Economics* 33: 441–462.

Christiano, L., and M. Eichenbaum. 1992. Current real-business-cycle theories and aggregate labor-market fluctuations. *American Economic Review* 82: 430–450.

Cole, H.L., and L.E. Ohanian. 2004. New deal policies and the persistence of the great depression: A general equilibrium analysis. *Journal of Political Economy* 112: 779–816.

Cooley, T.F. 1995. *Frontiers of business cycle research*. Princeton: Princeton University Press.

Cooley, T.F., and G.D. Hansen. 1989. The inflation tax in a real business cycle model. *American Economic Review* 79: 733–748.

Greenwood, J., and Z. Hercowitz. 1991. The allocation of capital and time over the business cycle. *Journal of Political Economy* 99: 1188–1214.

Greenwood, J., Z. Hercowitz, and G.W. Huffman. 1988. Investment, capacity utilization, and the real business cycle. *American Economic Review* 78: 402–417.

Hansen, G.D. 1985. Indivisible labor and the business cycle. *Journal of Monetary Economics* 16: 309–327.

Kim, I.-M., and P. Loungani. 1992. The role of energy in real business cycle models. *Journal of Monetary Economics* 29: 173–189.

King, R.G., and S. Rebelo. 1999. Resuscitating real business cycles. In *Handbook of macroeconomics*, vol. 1B, ed. J. Taylor and M. Woodford. Amsterdam: North-Holland.

Kydland, F.E., and E.C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50: 1345–1370.

Long Jr., J.B., and C.I. Plosser. 1983. Real business cycles. *Journal of Political Economy* 91: 39–69.

Marimon, R., and A. Scott (eds.). 1999. *Computational methods for the study of dynamic economies*. New York: Oxford University Press.

McGrattan, E.R. 1994. The macroeconomic effects of distortionary taxation. *Journal of Monetary Economics* 33: 573–601.

Mendoza, E.G. 1995. The terms of trade, the real exchange rate, and economic fluctuations. *International Economic Review* 36: 101–137.

Merz, M. 1995. Search in the labor market and the real business cycle. *Journal of Monetary Economics* 36: 269–300.

Rebelo, S. 2005. Real business cycle models: Past, present, and future. *Scandinavian Journal of Economics* 107: 217–238.

Rogerson, R. 1988. Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics* 21: 3–16.

Real Cost Doctrine

J. Maloney

Keywords
Alternative cost doctrine; Austrian economics; Edgeworth, F. Y.; Jevons, W. S.; Labour supply; Marginal utility theory; Marshall, A.; Real cost doctrine; Subjective theory of value

JEL Classifications
D2

Real cost doctrine is the doctrine that the supply price of a good is the price required to overcome the disutility involved in producing it. The worker, in other words, produces output up to the point at which his (decreasing) marginal utility of income equals his (increasing) marginal disutility of labour. The real cost doctrine can be seen as a half-way house inhabited by economists who had adopted a subjective theory of value but stopped short of the 'alternative cost' doctrine whereby the supply price of a resource is equal to its potential earning in its next most productive use. Much of the discussion which took place between English and Austrian economists concerned whether, and to what extent, the two doctrines logically came to the same thing.

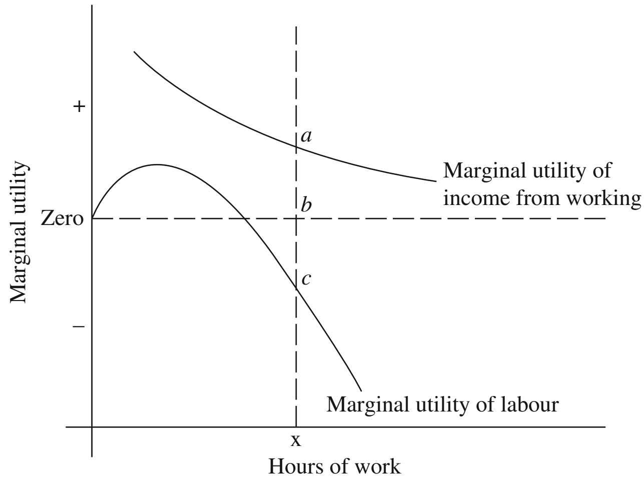
Jevons (1871) formulated the real cost doctrine in terms of the diagram in Fig. 1. Jevons assumes here (no such assumption is strictly necessary) that the worker at the start of the day not only enjoys his work but that, for a while, his enjoyment increases as he warms up to it. But, as the hours pass, the fatigue and boredom come to predominate over pleasure at an ever-increasing rate. The worker will maximize his surplus of utility over disutility by stopping at point $X\,(ab = bc.)$

The idea that subjective disutility of labour is central in determining output and price is, perhaps, Jevons's most unquestionably original idea. Not only is it absent from the work of Walras and Menger, but its prefigurations in the classical period are rare and rudimentary when compared with the pre-1871 analyses of marginal utility theory. (Jennings 1855, points out that marginal disutility of labour increases as the working day progresses but fails to build anything upon it.)

Marshall's theory of price determination, unveiled in his *Principles of Economics* (1890), differs little from Jevons's. Yet what looked radical in Jevons appears almost backward-looking in Marshall. This has something to do with the extension and dissemination of neoclassical principles in the intervening 20 years. But it also stems from a difference of presentation grounded in the contrast between Jevons's impatience with and Marshall's deference towards the Ricardian tradition. Much of Marshall's frequent praise for the English classical economists deftly sidesteps the question of how far they had actually been right. In the *Principles*, however, not only are cost and utility considerations given equal importance when determining price, but the fact that Marshall's conception of cost is ultimately a Jevonian 'subjective disutility' one is played down. It receives the strongest emphasis when Marshall argues that the capitalist as well as the worker undergoes real costs in the productive process, the capitalist's cost being that of 'waiting' rather than consuming his wealth immediately. (Nassau Senior had invoked Marx's sarcasm by speaking of capitalist 'abstinence' in the same context: Marshall tried both to circumvent the ridicule by renaming abstinence 'waiting' and to defend Senior from a neoclassical perspective, pointing out that *at the margin* of *aggregate* saving, considerable immediate sacrifice was undoubtedly involved.)

The rival doctrine, that of alternative cost, was espoused principally by the Austrians Wieser and Böhm-Bawerk and advertised in Britain by Wicksteed. All three denied the existence of any such thing as a supply curve, 'supply' simply being reverse demand. Böhm-Bawerk cited a horse fair: the buyer's utility from acquiring a horse and the seller's utility from keeping his horse played not just an equal but an identical role in determining price. Hence only a demand

Real Cost Doctrine, Fig. 1



curve need be drawn; at the equilibrium price, it crosses the vertical line representing the fixed stock of horses. Both Marshallian and Austrian analysis predict the same price.

But, of course, the fixed stock of horses makes this a very simple case: we are ignoring the cost of producing them. Such considerations, however, were no problem to the Austrians, who proclaimed that the costs of factors of production and raw materials ultimately depended on utilities from alternative uses forgone. Thus, as regards the labour market, the wage in a particular industry was governed by the demand for labour in other industries. Each worker had to be paid enough to keep him out of his next best paid available job. The Jevonian notion of disutility of labour dropped out of the picture, Böhm-Bawerk (1894) arguing against it on the *empirical* ground that few workers had the chance to make fine adjustments to the length of their working day. To this Edgeworth retorted that the Austrian doctrine implied that individuals made the choice to work or not to work once and for all at the beginning of their careers – it could not handle variations in labour supply due to variations in the wage rate.

The debate as a whole thus seemed to imply that the choice between real cost and alternative cost depended on whether *flexible* labour supply at the *individual* level (assumed by Jevons) or *inflexible* labour supply at the *aggregate* level (implied by the Austrians) was the more objectionable violation of reality. Yet logically the two theories come to exactly the same thing, and are seen to do so as long as the two 'sides' make one clarification apiece.

Austrians must make it clear that 'forgone utility' includes not only forgone income but also forgone leisure (when you work at all) and forgone non-pecuniary benefits (when you choose a less pleasant but better-paid job in preference to a more pleasant but worse-paid one). Böhm-Bawerk (1894) did spell this out.

Real cost theorists must make it clear that when a baker ponders whether to work another hour, what matters is not the disutility of the work as compared with doing nothing, but the disutility of work as compared with what he would choose to do (it might still be nothing!) if he were not baking. Equally it is not the 'gross' marginal utility of income which matters but the marginal utility of the *additional* income gained from spending another hour at the bakery rather than doing something else (other paid work, some leisure activity, or nothing). Edgeworth (1894) *failed* to spell this out; had he done so, a number of economists might have realized sooner than they actually did that both theories ultimately come to the same thing. (See Hobson 1926, for an example of confusion persisting well into the 20th century.)

See Also

- ► Marshall, Alfred (1842–1924)
- ► Opportunity Cost

Bibliography

Blaug, M. 1985. *Economic theory in retrospect*. 4th ed. Cambridge: Cambridge University Press.

Edgeworth, F.Y. 1894. Professor Böhm-Bawerk on the ultimate standard of value. *Economic Journal* 4: 518–521.

Hobson, J.A. 1926. *Free thought in the social sciences*. London: G. Allen & Unwin.

Jennings, R. 1855. *Natural elements of political economy*. London: Longman, Brown, Green, Longmans.

Jevons, W.S. 1871. *Theory of political economy*. Ed. and with an introduction by R.D. Collison Black. Harmondsworth: Penguin. 1970.

Marshall, A. 1890. *Principles of economics*. London: Macmillan.

von Böhm-Bawerk, E. 1894. One word more on the ultimate standard of value. *Economic Journal* 4: 719–724.

Wicksteed, P.H. 1910. *The common sense of political economy*. London: Macmillan.

Real Exchange Rates

Menzie D. Chinn

Abstract

The real exchange rate plays a central role in the open economy. This article describes the various ways in which the real exchange rate has been defined in the literature. It also examines the theoretical and empirical determinants of this variable.

Keywords

Exchange rate dynamics; Nominal exchange rates; Productivity; Purchasing power parity; Real exchange rates; Sticky prices; Terms of trade; Tradable and non-tradable goods; Unit roots

JEL Classifications

F31

The real exchange rate plays a crucial role in models of the open economy. How the real exchange rate should be defined, how it behaves over time, and what determines it at various time horizons are all questions that have been posed over the years. They have taken on heightened importance in recent years, as the scope of international transactions has expanded and more and more economic activity is either directly or indirectly affected by economic activity in other countries.

The most common definition of the real rate is the nominal exchange rate adjusted by price levels,

$$q_t \equiv s_t - p_t + p_t^* \tag{1}$$

where s is the log exchange rate defined in units of home currency per unit of foreign, and p and p^* are log price levels. If purchasing power parity (PPP) holds, then q is always unity (or a constant if price indices are used). One should expect PPP to hold in a world where transportation and transactions costs were negligible, consumption baskets were identical, and no arbitrage profits existed. Absent these conditions, the real exchange rate will vary.

One way of thinking about the determinants of movements in the real exchange rate is to appeal to a decomposition. Suppose the price index is a geometric average of traded and non-traded good prices:

$$p_t = \alpha p_t^N + (1-\alpha) p_t^T \tag{2}$$

where the lower-case letters denote logged values. Then substituting (2) into (1) yields:

$$q_t \equiv \left(s_t - p_t^T + p_t^{T*}\right) + \left[-\alpha\left(p_t^N - p_t^T\right) + \alpha^*\left(p_t^{N*} - p_t^{T*}\right)\right] \tag{3}$$

$$q_t \equiv q_t^T + [\omega_t] \tag{3$'$}$$

Equation (3) indicates that the real exchange rate can be expressed as the sum of two components: (i) the relative price of tradables q^T, (ii) the intercountry relative price of non-tradables in terms of tradables in the home country ω.

The Determinants of the Real Exchange Rate

If PPP holds only for tradable goods, then only the second term in Eq. (3′) can be non-zero, and the relative tradables–non-tradables price is the determining factor in the value of the real exchange rate. Another possibility is that all goods are tradable but not perfectly substitutable; then the imperfect substitutes model results, and q^T is equivalent to q. More generally, both terms on the right hand side of Eq. (3′) can take on non-zero values. In either of these cases, there are a large number of variables that could influence each relative price. And of course, there is nothing to rule out both relative price channels as being operative. In popular discussion, all three definitions of 'the real exchange rate' are used, sometimes leading to considerable confusion.

Most models of the real exchange rate can be categorized according to which specific relative price serves as the object of focus. If the relative price of non-tradables is key, then the resulting models – in a small country context – have been termed 'dependent economy' (Salter 1959; Swan 1960) or 'Scandinavian' model. In the former case, demand-side factors drive shifts in the relative price of non-tradables. In the latter, productivity levels and the nominal exchange rate determine the nominal wage rate and hence the price level, and thence the relative price of non-tradables. In this latter context, the real exchange rate is a function of productivity (Krueger 1983, p. 157). Consequently, the two sets of models both focus on the relative non-tradables price but differ in their focus on the source of shifts in this relative price. Since the home economy is small relative to the world economy (hence, one is working with a one-country model), the tradable price is pinned down by the rest-of-the-world supply of traded goods. Hence, the 'real exchange rate' in this case is $(p^N - p^T)$.

The relative price of tradables definition is most appropriate when considering the relative price that achieves external balance in trade in goods and services. This variable is also what macroeconomic policymakers refer to as 'price competitiveness'; hence, anything that affects the markup of price over cost – including both the level of demand, input costs, and market structure – can determine the real exchange rate.

Notice the dichotomy between the relative price of tradables and the relative price of non-tradables breaks down when countries specialize in the production of goods. Then the real exchange rate is the same as the terms of trade; purchasing power parity would occur only if the two goods were perfect substitutes (see Lucas 1982; Stockman 1980).

Empirical Modelling and Results

Real Exchange Rate Dynamics

In one special case, there is no need to model the real exchange rate. If relative PPP is *assumed* to hold, then q is a constant. Empirically, this is clearly not true in the short run but could be in the long run. Consequently, tremendous effort has been invested in investigating whether q is trend stationary, even though trend stationarity is not the same as purchasing power parity holding (the stronger condition of mean stationarity is required). Numerous studies have evaluated the trend stationarity of q directly by application of unit root tests, or indirectly by assessing whether the component series of q exhibit common long-term trends. Broadly speaking, the conclusions in this literature are mixed. Generally, panel methods, long time samples, and the use of producer or wholesale price indices provide more evidence in favour of a trend stationarity q than do pure time series methods, short samples, and the use of consumer price indices (see Rogoff 1996; Taylor and Taylor 2004). These results leave open the possibility that economic variables affect the movement of exchange rates over the short as well as the long run.

Modelling Real Exchange Rate Movements as a Function of Economic Variables

The modelling of the real exchange rate determinants can be divided into two main categories.

The first category includes models of the nominal exchange rate which, by virtue of the assumption of sticky prices, become models of the real exchange rate. First and foremost among these are sticky price monetary models that incorporate exchange rate overshooting, such as Dornbusch (1976) and Frankel (1979). In the long run, purchasing power parity holds, so that these models are only short-run models.

The second category includes models that focus on the determinants of the long-run real exchange rate. By far dominant in this category are those that centre on the relative price of non-tradables. These include the specifications based on the approaches of Balassa (1964) and Samuelson (1964) that model the relative price of non-tradables as a function of sectoral productivity differentials, including Hsieh (1982), Canzoneri et al. (1999) and Chinn (1999, 2000). They also include those models that search more broadly and include demand-side determinants of the relative price, such as DeGregorio and Wolf (1994). Engel (1999) has cast doubt upon the relevance of the relative non-tradables price. He demonstrates that for the G-7 economies, the variability of q^T as proxied by the tradable components of the CPI is comparable to the variability of ω even at horizons of 15 years.

More recently, some version of the portfolio balance model has been resurrected. Lane and Milesi-Ferretti (2002) have forwarded a model wherein the real rate depends upon net foreign assets. Early panel evidence in favour of the importance of this factor is to be found in Gagnon (1996).

Some methodological approaches do not fall neatly into one or the other category. The analysis by Mark and Choi (1997) is one instance. They compare the usefulness of monetary and real factors in predicting exchange rate changes over long horizons, and find – surprisingly – that monetary factors have persistent effects on the real exchange rate. Using a different methodology, namely, a structural vector autoregression, Clarida and Gali (1995) find that monetary and demand-side factors dominate in the determination of exchange rates. Also relying upon a structural (permanent-transitory) decomposition involving the real exchange rate and the current account, Lee and Chinn (2006) find that positive permanent shocks (interpreted as productivity innovations) tend to appreciate the currency and (at least for the United States) have an impact comparable in magnitude to those of temporary shocks.

See Also

▶ Cointegration
▶ Exchange Rate Dynamics
▶ Monetary Business Cycle Models (Sticky Prices and Wages)
▶ Nominal Exchange Rates
▶ Purchasing Power Parity
▶ Real Exchange Rates
▶ Terms of Trade
▶ Tradable and Non-tradable Commodities
▶ Unit Roots

Bibliography

Balassa, B. 1964. The purchasing power parity doctrine: A reappraisal. *Journal of Political Economy* 72: 584–596.

Canzoneri, M.B., R.E. Cumby, and B. Diba. 1999. Relative labor productivity and the real exchange rate in the long run: Evidence for a panel of OECD countries. *Journal of International Economics* 47: 245–266.

Chinn, M. 1999. Productivity, government spending and the real exchange rate: Evidence for OECD countries. In *Equilibrium exchange rates*, ed. R. MacDonald and J. Stein. Boston: Kluwer.

Chinn, M. 2000. The usual suspects: Productivity and demand shocks and Asia-Pacific real exchange rates. *Review of International Economics* 8: 20–43.

Clarida, R., and J. Gali. 1995. Sources of real exchange rate movements: How important are nominal shocks? *Carnegie-Rochester Conference Series on Public Policy* 41: 9–66.

DeGregorio, J., and H. Wolf. 1994. Terms of trade, productivity, and the real exchange rate. Working Paper No. 4807. Cambridge, MA: NBER.

Dornbusch, R. 1976. Expectations and exchange rate dynamics. *Journal of Political Economy* 84: 1161–1176.

Engel, C. 1999. Accounting for US real exchange rate changes. *Journal of Political Economy* 107: 507–538.

Frankel, J. 1979. On the mark: A theory of floating exchange rates based on real interest differentials. *American Economic Review* 69: 610–622.

Gagnon, J. 1996. Net foreign assets and equilibrium exchange rates: Panel evidence. International Finance Discussion Papers No. 574. Washington, DC: Board of Governors of the Federal Reserve System.
Hsieh, D. 1982. The determination of the real exchange rate: The productivity approach. *Journal of International Economics* 12: 355–362.
Krueger, A.O. 1983. *Exchange-rate determination*. Cambridge: Cambridge University Press.
Lane, P.R., and G.M. Milesi-Ferretti. 2002. External wealth, the trade balance, and the real exchange rate. *European Economic Review* 46: 1049–1071.
Lee, J., and M. Chinn. 2006. Current account and real exchange rate dynamics in the G-7 countries. *Journal of International Money and Finance* 25: 257–274.
Lucas, R. 1982. Interest rates and currency prices in a two-country world. *Journal of Monetary Economics* 10: 335–359.
Mark, N., and D.Y. Choi. 1997. Real exchange rate prediction over long horizons. *Journal of International Economics* 43: 29–60.
Rogoff, K. 1996. The purchasing power parity puzzle. *Journal of Economic Literature* 34: 647–668.
Salter, W.A. 1959. Internal and external balance: The role of price and expenditure effects. *Economic Record* 35: 226–238.
Samuelson, P. 1964. Theoretical notes on trade problems. *Review of Economics and Statistics* 46: 145–154.
Stockman, A. 1980. A theory of exchange rate determination. *Journal of Political Economy* 88: 673–698.
Swan, T. 1960. Economic control in a dependent economy. *Economic Record* 36: 51–66.
Taylor, A.M., and M.P. Taylor. 2004. The purchasing power parity debate. *Journal of Economic Perspectives* 18(4): 135–158.
Williamson, J. 1994. *Estimating equilibrium exchange rates*. Washington, DC: Institute for International Economics.

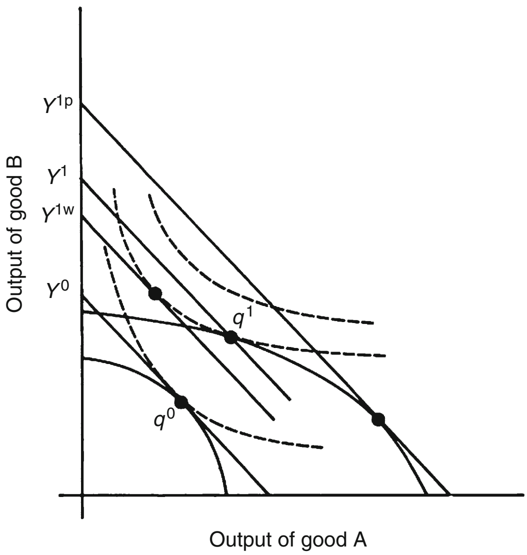
Real Income

D. Usher

Real income can be defined at two levels. As a statistic, it is money income corrected for changes in prices. If a person's income rose from $20.000 in 1980 to $25,000 in 1981, we say that the growth in his *money* income was 25 per cent, (25–20)/20 expressed as a percentage. However, if the price level rose from 100 in 1980 to 110 in 1981, we say that the growth in his *real* income was only 13.64 per cent [(25/110)–(20/100)]/(20/100). But statistics are not just manipulations of data, and real income is not completely defined until we identify the characteristic, property or aspect of the economy the statistic is intended to reflect. Real income must be defined as a concept as well as a statistic, to indicate what the statistic is for and to serve as a guide in choosing and manipulating the original data from which the statistic is compiled. In particular, the rules for constructing a price index to deflate money income into real income can only be established with reference to the information the statistic is designed to convey.

As a concept, real income is intrinsically comparative. It makes sense to say that real income is 3 per cent higher this year than last, or that the real income of one country is 30 per cent higher than the real income of another. It makes no sense to say that the real income of a country is such and such today, except as an implied comparison with another time or place. The concept of real income is usually, though not necessarily, applied to a country or a region rather than to a person or family.

What then is being compared? There are two standard answers to this question, giving rise to two distinct but related concepts of real income. Real income may be an indicator of welfare or of productive capacity. In both cases, real income is measured in dollars worth when quantities of goods are valued at an arbitrarily chosen set of prices, usually the market prices in some base year. When real income is looked upon as an indicator, of welfare, an increase in real income signifies that the representative consumer, whose indifference curves are assumed to remain invariant, is becoming better off over time. To construct a time series of real income as an indicator of welfare, one would, ideally, require a complete set of indifference curves, a set of time series of all goods consumed, and a set of base year prices. One could then measure real income each year as the least amount of money required at base year prices to purchase a bundle of goods on the indifference curve attained with the bundle of goods that the representative consumer actually consumes in that year. When real income is looked upon as a measure of productive capacity, an



Real Income, Fig. 1

increase in real income signifies that there has been an outward shift in the production possibility frontier for the economy as a whole. For any set of base year prices, real income each year is the maximum value at those prices of any bundle of goods on the production possibility frontier for that year. Both concepts of income give rise to a family of measures, each member of which corresponds to a different set of price weights.

The distinction between these concepts of real income is illustrated on Fig. 1 for a comparison of two years, 0 and 1, in an economy with two goods, A and B. Outputs per head are measured on the horizontal and vertical axes. The dashed curves are indifference curves. The solid curves are production possibility frontiers in the years 0 and 1. Amounts produced and consumed are represented by the points q^0 and q^1. Relative prices may be represented as slopes of lines. Money income corresponding to any bundle of goods and any relative price may be represented as a distance on the vertical axis if we adhere to the convention, which is harmless in this context, that the money price of the good B is always equal to 1.0. The common slope of the parallel straight lines represents the relative price of the two goods in the year 0 which is the chosen base year. Money income in the year 0 is Y^0, the value of quantities produced in year 1 at prices in the year 0 is Y^1, real income in year 1 as a measure of welfare is Y^{1w} and real income as a measure of productive capacity is Y^{1P}. By construction, money income, real income as a measure of welfare and real income as a measure of productive capacity are all the same in the base year. Note that $Y^{1W} < Y^1 < Y^{1P}$ as long as indifference curves and the production possibility frontiers have their usual shapes. The corresponding rates of economic growth are $(Y^{1P}/Y^0 - 1)$ and $(Y^{1W}/Y^0 - 1)$. For each definition of real income and each choice of a base year, the appropriate price index may be constructed.

In practice, matters are at once simpler and more complex. They are simpler because the different measures of real income may not be too far apart and because users of the national accounts, impatient with the niceties of concepts, want an all-purpose measure of real income that gives a rough idea of what is happening to the economy. Matters are more complex because the world refuses to conform to the set of concepts within which income is defined: There are more than two goods. The decision to apportion the flow of income into amounts of a finite set of goods is somewhat arbitrary. The nature and quality of goods is changing over time, so that it is by no means certain how much the quantity of what the statistician calls a good has increased. Society consists of many people each with his own unique set of indifference curves. Even when people have the same indifference curves, their response to price changes is not independent of income. All that can be observed in practice are prices and quantities, not the underlying indifference curves or production possibility frontiers.

The welfare interpretation is better suited to the measurement of real consumption than to the measurement of real income because indifference curves are defined over consumption goods alone and items such as newly built factories or aircraft can only be accounted for as surrogates for the flow of consumption goods they will eventually bring forth. In practice, formulae appropriate for averaging prices of consumption goods are applied to prices of all goods, capital goods and consumption goods alike.

The productive capacity interpretation cannot be made to generate a time series of real income unless the production possibility frontier for each year in the series is a well-defined function of quantities of a set of goods that remains invariant over time. The frontiers in years 0 and 1 are comparable if actual and potential outputs in both years consist of certain amounts of apples and oranges. The frontiers are not comparable, and real income as a measure of productive capacity is ill-defined, if the economy produces only apples and oranges in year 0 and only grapes and lemons in year 1.

To measure real income, statisticians have to squeeze the flow of innumerable items each year into a set of quantities of a relatively small number of commodities, such that the nature of the commodities is assumed to remain the same over time. Each of the many different kinds of stereo sets must one way or another be represented as a definite amount of one homogeneous commodity. This may be done directly, or it may be done indirectly by deflating the value of stereo sets by a price index. The process by which time series of the supposedly homogeneous and invariant quantities of the different goods are forced into an index of aggregate quantity to reflect the chosen concept of real income as closely as possible (or, equivalently, time series of prices of different goods are forced into a price index to deflate money income) is described in the entry on Index Numbers.

See Also

▶ Hedonic Functions and Hedonic Indexes
▶ Index Numbers

Bibliography

For a survey of the concept of real income with an extensive bibliography see A.K. Sen, 'The welfare basis of real income comparisons. 1979. *Journal of Economic Literature*: 1–45.

On the measurement of prices in the presence of quality change, see George Stigler and James Kindahl, *The behavior of industrial prices*. New York: Columbia University Press, 1970, and Zvi Griliches (ed.), *Price indexes and quality change: Studies in new methods of measurement*. Cambridge, MA: Harvard University Press, 1971.

The distinction between the welfare and productive capacity interpretations of real income is due to John Hicks. 1940. The valuation of social income. *Economica* 7: 105–124.

Real Rigidities

David Romer

Abstract

Real rigidities are forces that reduce the responsiveness of firms' profit-maximizing prices to variations in aggregate output resulting from variations in aggregate demand. Real rigidities make firms less inclined to take actions that dampen movements in aggregate output, and so increase the responsiveness of output to disturbances. They appear essential to any successful explanation of short-run macroeconomic fluctuations. As a result, various forms of real rigidity pervade modern models of business cycles.

Keywords

Adjustment costs; Aggregate demand; Business cycles; Capital-market imperfections; Cyclical markups; Efficiency wages; Elasticity of substitution; Imperfect competition; Input–output analysis; Labour mobility; Labour supply; Menu costs; Nominal rigidities; Real business cycles; Real rigidities; Staggered price setting; Sticky prices; Strategic complementarity

JEL Classifications

D4; D10; E12; E32

'Real rigidities' is the name given to a large class of business cycle propagation mechanisms. Real rigidities appear essential to any successful explanation of business cycles.

The Definition of Real Rigidities

Although the term 'real rigidities' appears vague, it in fact refers to a precise concept. Consider an economy of symmetric price-setting firms that is at its flexible-price equilibrium, and suppose that the money supply increases with prices unchanged, so that aggregate output increases. Now ask by how much a representative firm would want to increase its price if it faced no barriers to nominal price adjustment. By definition, the smaller the amount the firm would want to increase its price in response to a given increase in aggregate output, the greater the degree of real rigidity.

One can see the meaning of 'real rigidity' more formally by observing that, in the experiment described above, the profits of the firm that is considering changing its price, neglecting any costs of price adjustment, typically can be written in the form $V(p_i\text{-}p, y)$, where p_i is the firm's price, p is the aggregate price level, and y is the departure of output from its flexible-price level (all in logs). In most models of this type, $V(\cdot)$ is a smooth function. The first-order condition for the profit-maximizing price is $V\left(p_i^* - p, y\right) = 0$ (subscripts denote partial derivatives). At the flexible-price equilibrium, $p_i^* = p$ and $y = 0$. Starting from that equilibrium, the derivative of the representative firm's desired relative price, $p_i^* = p$ with respect to y is thus

$$\frac{d\left(p_i^* - p\right)}{dy}\Big|_{p_i^* - p = 0,\, y=0} = \frac{V_{12}(0,0)}{V_{11}(0,0)} \equiv \varphi$$

For the flexible-price equilibrium to be stable, ϕ must be positive. By definition, a lower value of ϕ corresponds to greater real rigidity. Note that real rigidity is defined entirely in terms of relations among real variables: it refers to the (lack of) responsiveness of desired real prices to aggregate real output.

The definition of real rigidity in models without symmetric price-setting firms is analogous: any force that reduces the amount that price setters would change their relative prices in response to movements in aggregate output that are the result of changes in aggregate demand is a real rigidity.

Real Rigidities and Business Cycles

Real rigidities are crucial to business cycles. At a general level, real rigidities make firms less inclined to take actions that dampen movements in aggregate output. As a result, they increase the responsiveness of output to disturbances.

The importance of real rigidities is easiest to see in a static model where firms face fixed costs of changing prices. Consider the model sketched above, with two extensions. First, replace the profit function with a second-order approximation around the flexible-price equilibrium. This implies that the representative firm's loss in profits from failing to charge its profit-maximizing price (neglecting costs of price adjustment) is $K\left(p_i - p_i^*\right)^2$, where $K \equiv -V_{11}$. It also implies that the representative firm's profit-maximizing price is given by $p_i^* - p = \varphi y$, where φ is as defined before. Second, assume that each firm faces a fixed cost $Z > 0$ of changing its nominal price.

The economy begins at its flexible-price equilibrium. We want to know by how much output can change in response to a change in aggregate demand before firms change their prices. Non-adjustment is an equilibrium as long as the representative firm's losses from failing to adjust are less than Z. Prior to the shock, the representative firm's price equals the aggregate price level, p. If the firm adjusts its price, it sets it to the new profit-maximizing level, $p + \varphi y$. Thus the condition for nonadjustment to be an equilibrium is $K[p - (p + \varphi y)]^2 < Z$, or $|y| < 1/\left(\varphi\sqrt{Z/K}\right)$. Thus, when φ is lower – that is, when real rigidity is greater – the range over which aggregate demand shocks affect real activity is greater.

Real rigidities are not just important to models with nominal rigidity, however. Consider, for example, the following minimalist real business cycle model. The markets for labour and goods are perfectly competitive, and the representative firm's production function is $y_i = a + n_i$ (y is output, a is productivity, and n is labour input, again all in logs). Labour supply is $n = \gamma w$, $\gamma > 0$, where w is the (log) real wage. In this model, the elasticity of profit-maximizing relative prices to demand-driven output fluctuations (that is, to variations in y with a fixed, which must come from variations in n) equals the elasticity of the real wage with respect to y, which is $1/\gamma$. Thus a larger value of γ corresponds to greater real rigidity.

The production function implies that labour demand is perfectly elastic at $w = a$. Labour-market equilibrium therefore requires that $n = \gamma a$. A larger value of γ therefore implies that productivity shocks have a larger impact on employment, and thus that the output effects of the shocks are magnified to a greater extent. Thus, even in this purely Walrasian model of fluctuations, real rigidity acts as a propagation mechanism.

Real rigidities also act as amplification mechanisms in dynamic models of price adjustment. Consider an economy with barriers to price adjustment where output is above its flexible-price level, and suppose that some firms have an opportunity to change their prices, and that their new prices will be in effect for more than one period. Greater real rigidity increases the persistence of the departure of output from its flexible-price level, for three reasons. First, as in the static model of price adjustment, it reduces the benefits of price adjustment, and so makes firms more inclined not to adjust at all. Second, the fact that only some firms have the opportunity to adjust means that output will continue to be above its flexible-price level. Greater real rigidity then implies that the firms that adjust will respond by less, thus drawing out the period of above-normal output. Third, the fact that other firms will be in the same situation when they adjust their prices means that they will adjust by less, which in turn dampens the adjustments of the firms that adjust immediately.

There is a close link between real rigidity and strategic complementarity in profit-maximizing prices. If we assume the stylized aggregate demand curve $y = m - p$ (where m reflects factors that shift aggregate demand), then the expression for the representative firm's profit-maximizing relative price, $p_i^* - p = \varphi y$, implies $p_i^* = \varphi m + (1 - \varphi)p$. Thus greater real rigidity corresponds to greater strategic complementarity in desired prices: when φ is lower, each firm wants its price to move more closely with other prices.

Real rigidity and strategic complementarity in desired prices are not identical, however. To see this, suppose the aggregate demand equation is instead $y = \beta\,(m - p)$, $\beta > 0$. Then $p_i^* = \varphi\beta m + (1 - \varphi)p$. Thus β affects strategic complementarity but not real rigidity. And it is real rigidity that is key to cyclical fluctuations. Nonetheless, because of the close link between the two concepts, and because many business cycle models assume $y = m - p$, the terms real rigidity and strategic complementarity in prices are often used interchangeably.

Types of Real Rigidities

Since any force that reduces the responsiveness of profit-maximizing relative prices to demand-driven output fluctuations is a real rigidity, there are many possible real rigidities. Some might not be commonly thought of as 'rigidities'. For example, as the simple real business cycle model shows, more elastic labour supply is a type of real rigidity.

Such neoclassical sources of real rigidity, however, are almost surely not strong enough to generate output fluctuations of the size and nature that we observe. In Walrasian models, the real wage is likely to rise sharply with the quantity of labour. For this not to occur, either the long-run elasticity of labour supply must be high or the intertemporal elasticity of substitution in labour supply must be high and short-run aggregate fluctuations must have a large transitory component. Neither of these conditions appears to hold in practice. In models of nominal disturbances and barriers to nominal price adjustment, the result is large incentives for price changes, and thus little nominal rigidity. In productivity-driven real business cycle models, the result is small movements in labour input, so that the dynamics of aggregate output largely mimic the dynamics of the underlying productivity shocks. Researchers have therefore turned their attention to non-Walrasian sources of real rigidity.

It appears difficult to understand substantial employment fluctuations without non-Walrasian real rigidities in the labour market. At a general level, what is needed is for some force causing workers to be off their labour supply curves, at least in the short run, so that the cyclical behaviour of the real wage is not governed by the elasticity of labour supply. For example, if there is equilibrium unemployment because of efficiency wages, the cyclical behaviour of the real wage depends on how the efficiency wage varies with aggregate output. As a result, the real wage can be (though it need not be) less procyclical than in a Walrasian labour market, with the result that fluctuations in employment and output are greater.

A more subtle real rigidity in the labour market arises if labour is imperfectly mobile in the short run (because of search frictions, for example), so that each firm faces an upward-sloping labour supply curve rather than perfectly elastic supply at the economy-wide wage. Consider, for example, a firm contemplating cutting its price and increasing production in a recession. With imperfectly mobile labour, this requires paying a higher real wage. Thus the amount the firm wants to reduce its price is smaller – that is, real rigidity is greater.

There can also be important real rigidities in other markets. In the goods market, forces making desired markups countercyclical act as real rigidities. When desired markups are more countercyclical, then, for a given degree of procyclicality of real marginal costs, desired movements in relative prices are smaller. Countercyclical desired markups can stem from a variety of sources. One simple but potentially important possibility is that, when economic activity is greater, firms' incentives to disseminate information and consumers' incentives to acquire it are greater, and so demand is more elastic.

Another feature of goods markets that can act as a real rigidity is input-output links among firms. If the prices charged by intermediate suppliers are sticky, the costs that firms face for intermediate inputs tend to rise by less than the suppliers' costs in a boom, thereby reducing the amount that firms would raise their prices if they were free to do so.

Capital-market imperfections can also create real rigidities. Capital-market imperfections can cause financing costs to be countercyclical, as higher output increases cash flow (and hence firms' ability to use internal finance) and raises asset values (and hence increases collateral and reduces the cost of external finance). With one component of costs countercyclical, desired prices are less procyclical. To give another example, financial difficulties in recessions can increase the importance of short-term profits to firms relative to expanding their customer base, and so can make desired markups countercyclical.

There is an important distinction between two broad categories of real rigidities. One category consists of forces, such as limited short-run labour mobility among firms, that increase real rigidity by affecting what happens when one firm changes its prices and others do not. The other consists of forces, such as factors that reduce the procyclicality of the real wage, that increase real rigidity by affecting what happens when all firms' output moves together. In terms of the definition of real rigidity as $V_{12}(0,0)/[-V_{11}(0,0)]$, the first category consists of forces raising $-V_{11}(0,0)$, and the second consists of forces reducing $V_{12}(0,0)$.

The distinction between these two categories is important for two reasons. First, real rigidities that result from forces that affect what happens when one firm changes its price with other firms' prices fixed are not relevant to the properties of business cycle models with identical firms and fully flexible prices. Second, the two types of real rigidities have different microeconomic implications. Most importantly, factors that increase real rigidity by affecting what happens when one firm changes its price and others do not increase the costs of departures from the profit-maximizing price; as a result, they typically predict smaller movements in firms' relative prices in response to many types of shocks.

Selected Literature

Ball and Romer (1990) establish that in a static setting, imperfect competition and barriers to nominal adjustment alone are unlikely to generate substantial nominal rigidity. They show the general importance of real rigidities to static menu-cost models and stress that forces making desired real wages relatively unresponsive to output fluctuations are likely to be essential to generating substantial nominal rigidity (see also Blanchard and Fischer 1989, ch. 8). Earlier work by Akerlof and Yellen (1985) incorporates substantial real rigidity in a model of price stickiness, although it does not explicitly analyse the importance of real rigidity to the results. Haltiwanger and Waldman (1989) show that strategic complementarity magnifies the impact of non-responders on equilibrium outcomes, a result that is closely related to the finding that real rigidities magnify the effects of barriers to price adjustment.

Kimball (1995) establishes the central role of real rigidities to the persistence of output fluctuations in models of staggered price adjustment, and stresses the importance of the distinction between forces that affect what happens when all firms' outputs move together and forces that affect what happens when one firm changes its price with other firms' prices fixed (see also Blanchard 1987). Klenow and Willis (2006) show the differing microeconomic implications of the two categories of real rigidities. Romer (2006, ch. 6) provides a general discussion of the importance of real rigidities to static and dynamic models of nominal rigidity, catalogues many specific real rigidities and provides numerous references.

Real rigidities pervade modern business cycle models. In real business cycle models, for example, such common features as indivisible labour supply, variable capital utilization and labour hoarding, and learning-by-doing (see, for example, Rogerson 1988; Burnside and Eichenbaum 1996; and Chang, Gomes and Schorfheide, 2002) magnify the effects of disturbances precisely because they are real rigidities. In models with price stickiness, some important recent analyses where real rigidities play a central role include Mankiw and Reis (2002), Gertler and Leahy (2006) and Carvalho (2006).

See Also

- ▶ Cyclical Markups
- ▶ Monetary Business Cycle Models (sticky prices and wages)
- ▶ New Keynesian Macroeconomics
- ▶ Real Business Cycles
- ▶ Sticky Wages and Staggered Wage Setting

Bibliography

Akerlof, G.A., and J.L. Yellen. 1985. A near-rational model of the business cycle, with wage and price inertia. *Quarterly Journal of Economics* 100: 823–838.

Ball, L., and D. Romer. 1990. Real rigidities and the non-neutrality of money. *Review of Economic Studies* 57: 183–203.

Blanchard, O.J. 1987. Aggregate and individual price adjustment. *Brookings Papers on Economic Activity* 1987(1): 57–109.

Blanchard, O.J., and S. Fischer. 1989. *Lectures on macroeconomics*. Cambridge: MIT Press.

Burnside, C., and M. Eichenbaum. 1996. Factor-hoarding and the propagation of business-cycle shocks. *American Economic Review* 86: 1154–1174.

Carvalho, C. 2006. Heterogeneity in price stickiness and the real effects of monetary shocks. *Frontiers of Macroeconomics* 2(1), Article 1.

Chang, Y., J.F. Gomes, and F. Schorfheide. 2002. Learning-by-doing as a propagation mechanism. *American Economic Review* 92: 1498–1520.

Gertler, M. and Leahy, J. 2006. A Phillips curve with an Ss foundation. Working Paper No. 11971. Cambridge, MA: NBER.

Haltiwanger, J., and M. Waldman. 1989. Limited rationality and strategic complements: The implications for macroeconomics. *Quarterly Journal of Economics* 104: 463–483.

Kimball, M.S. 1995. The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit, and Banking* 27: 1241–1277.

Klenow, P.J. and Willis, J.L. 2006. Real rigidities and nominal price changes. Working Paper No. 06-03. Kansas City, MO: Federal Reserve Bank of Kansas City.

Mankiw, N.G., and R. Reis. 2002. Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics* 117: 1295–1328.

Rogerson, R. 1988. Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics* 21: 3–16.

Romer, D. 2006. *Advanced Macroeconomics*. 3rd ed. New York: McGraw-Hill.

R

Real Wage Rates (Historical Trends)

Robert C. Allen

Abstract

Historical studies of the real wage allow us to track the divergence in the world of economics since the Middle Ages and changes in the distribution of income during the Industrial Revolution. Before the 19th century, the real wage moved inversely to the population. Since then it has increased dramatically.

Keywords

Black death; Capital accumulation; Capitalism; Consumer price index; Economic development; Factor prices; Great divergence; Index numbers; Industrial Revolution; Inequality in wages; International migration; International trade; Malthus's theory of population; National income accounting; Price histories; Productivity; Real wage; Solow, R.; Standards of living; Subsistence; Surplus labour; Technical change; Wage ladder

JEL Classifications

N3

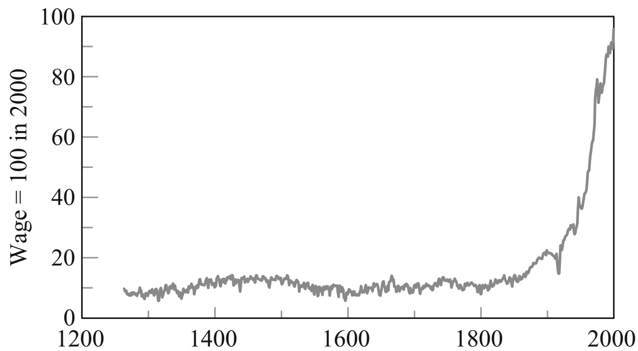
The real wage indicates the purchasing power of a worker's income. The real wage is the ratio of the nominal wage (what someone is actually paid) to a price or price index. Sometimes that price is the price of the product of a competitive firm, in which case the real wage is the marginal product of labour and has a productivity interpretation. In the more common case, however, and the one this article deals with, the price deflator is a consumer price index. In this case, the real wage measures the standard of living of the worker. Since that bears on central questions of economic growth and distribution, real wages have been an important tool for measuring and interpreting economic growth and stagnation over the last millennium.

Measuring real wages raises practical problems that are particularly acute in historical investigations. First, one needs information about wages, prices and spending shares to perform any calculations. Data sources for these have to be developed, which ultimately involves extensive archival work. There are conceptual problems as well since many people in the past received some income in kind as well as cash and many were also employed on piece rates that must be converted to earnings before an assessment of their purchasing power can be made. Second, an index number must be chosen to compute the consumer price index. While theorists have advanced many useful arguments about why some formulae are better than others, data limitations often force compromises. One of the most extreme was the once common practice of deflating wages by the price of grain. Other products were ignored, as well as the inconvenient fact that most people in the West ate bread, not grain. Most recent studies have avoided this practice. Third, new products and improvements in the quality of old products bedevil historical studies as they do modern ones. Although product innovation was less common in the past, the creation of the global economy led to the introduction into Europe or mass availability of maize, potatoes, tomatoes, chilli peppers, sugar, tobacco, cotton cloth, tea, coffee and porcelain. Also, comparisons of real wages between continents with radically different diets raise the question of new products in a cross-sectional context. How do you compare the standard of living of an English worker eating bread, beef and beer with a Chinese worker eating fish and rice?

Real Wages and Economic Growth in Developed Countries

Economic theorists have divided the history of the world into two phases. Before the onset of modern economic growth around 1800, income per head grew very slowly, if at all. Increases in productivity simply resulted in more people. The real wage moved inversely with the population and remained constant in the long run. This was the Malthusian phase of history.

The second phase began in about 1800. Technology improved steadily raising income per head. Population growth was restrained, so an increase in the labour force did not swamp the increase in labour demand. Consequently, the real wage rose in step with productivity. This has been called the Solow phase in view of Solow's (1956) growth model. While these models can be nuanced, as we will see, they provide a starting point for real wage history: is it consistent with these models?



Real Wage Rates (Historical Trends), Fig. 1 Real wage in London, 1200–2000 (*Sources:* Data before 1914: Allen (2001)). Later wage and price indices: Phelps Brown and Hopkins (1955), Mitchell (1998), ILO (various years)

We can measure the real wage over the past 800 years thanks to the accumulated research of historians who have written 'price histories' of cities since the mid-19th century. The price historian finds an institution like a college or hospital that has existed for centuries and examines its accounts to abstract the prices of the things it bought and the wages it paid. Oxford and Cambridge colleges were the first to be studied (Rogers 1866–92) and since then many European cities have been investigated. Phelps Brown and Hopkins (1955, 1956) were the first to take advantage of this material and construct a real wage index for English building workers from 1264 to 1954. More recently, Allen (2001) and Clark (2005) have reworked this material and added new evidence to compute new real wage series. While there are differences among these authors on issues like real wage change in the Industrial Revolution, the broad outlines of the real wage story are the same (Fig. 1).

The Malthusian and Solow phases stand out in this figure. Before the 19th century, England followed the Malthusian pattern. There was no long-run trend in the real wage, although there were fluctuations. These coincided with population swings; in particular, the real wage rose after the Black Death in 1348–9, which killed about one-third of the population, and fell in the 16th century as the population started to rebound. Real wages only rose above the pre-industrial peak once the Industrial Revolution was well under way or completed. The rise since then has been spectacular by comparison, and today the real wage is ten times its level in the pre-industrial world. This is the Solow phase of economic history.

The period of the Industrial Revolution, roughly 1770–1860, is something of a problem. Were real wages rising then or falling? The classical economists, who were writing in the early 19th century, were pessimistic about the prospects of workers. While they agreed that capitalism was likely to cause output per capita to rise, they also believed that real wages would remain constant at 'subsistence'. For Malthus, Ricardo and other mainstream economists, the reason was demographic: wages were the income of the bulk of the population, and a higher wage would lead them to have more children and live longer. The result would be an increase in the workforce that would push wages back to subsistence. While radicals like Marx and Engels rejected the demographic model, they agreed that wages would not rise under capitalism. Their explanation, however, turned on the demand for labour rather than its supply. Marx and Engels believed that technological progress would be so rapid and so labour-saving that the demand for labour would always fall short of the supply – again forcing wages back to subsistence. Only collective action or state interference would prevent this. None of the classical economists, in other words, expected the 'Malthusian economy' to transmute into the 'Solow economy'.

By the 20th century, it was clear that these arguments were wrong, for living standards were far higher than they had been 100 years before, as Fig. 1 shows. Kuznets (1955) raised the possibility that inequality went through an 'invented U' patterned during economic development. In his model, this worked through the wage structure itself. At the outset, workers were in

R

low-productivity, low-wage sectors. As the modern sector grew, it attracted workers by offering higher wages. Inequality increased as workers moved to that sector since those employed there were earning more than their counterparts in agriculture. Inequality in wages declined as the process of labour reallocation was completed, for all workers were then earning the higher wage paid in the modern sector.

The problem of economic development in poor countries provoked Lewis (1954) to propose a model of growth and distribution that was more classical in spirit and that emphasised the differential movements of output per head and the real wage. Lewis divided the economy into two sectors. In the traditional sector, consisting of peasant agriculture and the urban 'informal' economy, the main inputs were land and labour, and the latter was in surplus. In the modern, industrial sector, output was produced with capital and labour, and the former was scarce. Growth occurred as capital was accumulated and the modern sector expanded. Its labour force was drawn from the traditional sector. Surplus labour in that sector meant that the marginal product of labour was low, perhaps even zero, and income was shared and at a subsistence level. An elastic supply of labour kept the real wage in the modern sector at subsistence even though output per worker was rising. This process of rising inequality would continue, in Lewis's view, until the modern sector had expanded to absorb all the surplus labour. Only then would the real wage rise in step with output per worker.

How well do these theories fare in practice? The question has been extensively researched and vigorously debated in the case of the British Industrial Revolution. Lindert and Williamson (1983) were the first to apply modern economic methods to the question. They computed economy-wide average earnings and a consumer price index founded on budget surveys and corresponding prices. Their conclusion was guardedly 'optimistic' in that they found the average real wage rose sharply after 1815. This conclusion was not universally accepted. Feinstein (1998) computed an alternative price index that significantly reduced the rate of real wage growth leading to his title 'pessimism perpetuated'. Clark (2005), on the other hand, proposed yet another price index that tilted the conclusions back in a Lindert–Williamson direction. Most recently, Allen (2007a) has plumped for 'pessimism preserved'.

These disagreements reflect the limitations of the data, which are only a poor reflection of the ideal information discussed above. There were no comprehensive and representative samples of consumer spending, and even the annual series of individual prices are problematic. Quality change, in particular the growing use of cotton rather than wool in clothing, is only imperfectly grasped with the available information. There is considerable scope for contradictory – yet plausible – readings of the evidence.

The impact of economic development on wage rates has been pursued for many other countries with mixed results. Over the long term, real wage change in Western Europe has followed a pattern like that for England shown in Fig. 1 (Scholliers and Zamagni 1995). The United States has been repeatedly studied, and revisions to price and wage indices have been as thoroughgoing for the USA, as they have been for Britain. For instance, Douglas's (1930) conclusion that real wages only rose by eight per cent during the boom from 1890 to 1914 was overturned by Rees (1961), who found that real wages increased by 40 per cent. Over the long term, of course, real wages have risen dramatically in America, but the real wage lagged behind GDP per head during early industrialisation, according to Margo's (2000) study of the period 1800–60.

Outside the advanced Organisation for Economic Co-operation and Development (OECD) countries, the link between economic growth and real wage advance is much weaker. The economic boom experience by Tsarist Russia, for instance, was not reflected in urban or rural real wages (Allen 2003a). Latin America enjoyed a substantial rise in GDP per head in the 20th century, with only an elusive impact on real wages. In Mexico, which has been studied more than most countries, there were periods when real wages surged and others when real wages collapsed. The declines look about as big as the gains, but the uncertainty arising from the introduction of new goods makes definitive conclusions hazardous (Bortz and Aguila 2006).

Real Wages and the Great Divergence

The difference in real wages can be computed between two places as well as between two times, and the geographical dimension has allowed real wages to play an important role in the 'great divergence' debate. Since 1800, incomes have grown most rapidly in the most prosperous countries, so their lead over the poor countries has increased. Charting this divergence is the first step in explaining it. Real GDP per head may be the best indicator, and economists have tried to extrapolate it far into the past. Errors, however, accumulate, and the estimates become increasingly problematic the further back one goes. Real wages provide an alternative, and simpler, approach to the problem. The real wage is individual in its focus – what could a particular worker buy with his or her income? – and so avoids the economy-wide assumptions of national income accounting. The real wage also requires fewer data, and it can be computed directly for dates in the past without the need to extrapolate backwards.

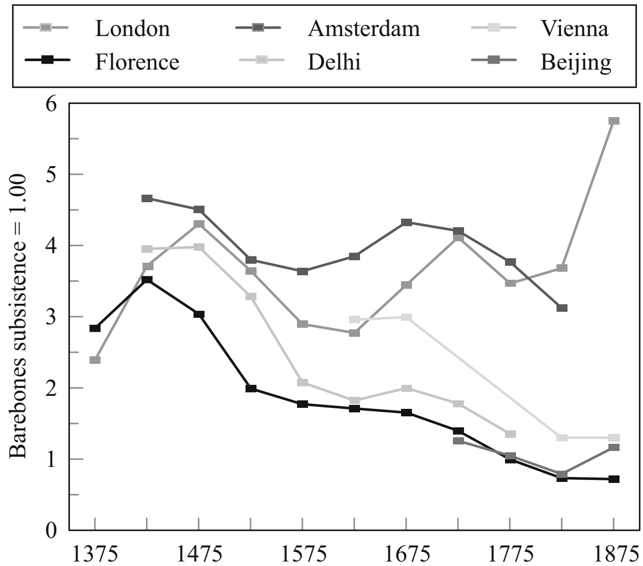
Indeed, the classical economists, who established a long-standing view on the subject, expressed Europe's lead in terms of real wages. Adam Smith (1776, pp. 74–5, 91, 187, 206) saw the world in terms of a wage ladder: workers in England and the Netherlands had a higher standard of living than workers in France or elsewhere on the European continent. Workers in Asia lagged behind Europeans. 'The real price of labour, the real quantity of the necessaries of life which is given to the labourer. . .is lower both in China and Indostand than it is through the greater part of Europe'.

This view has recently been challenged by scholars of Asia who have argued that Asia was as prosperous as Europe at the time Smith wrote. According to Pomeranz (2000, p. 49), 'It seems likely that average incomes in Japan, China, and parts of South-East Asia were comparable to (or higher than) those in western Europe even in the late eighteenth century'.

How does Smith's wage ladder stand up in terms of modern evidence? The price histories of European cities provide a start, for they allow us to compute real wage differences across Europe from the late Middle Ages to the 19th century (van Zanden 1999; Allen 2001). While today real wages are similar across Western Europe, the calculations show that the last time this was even approximately true was in the late 15th century. Between 1500 and 1750, real wages in Amsterdam and London, the booming maritime cities of north-western Europe, were trendless, while they fell sharply in other parts of Europe under the impact of population growth not offset by economic expansion (Allen 2003b). Incomes had diverged in Europe, in the manner described by Smith, before the Industrial Revolution. Indeed, it was decades, if not a century, before modern economic growth was perceptible in the real wage data. So far as Europe was concerned, the great divergence preceded the Industrial Revolution rather than being its sequel.

What about Europe and Asia? It is only very recently that comparisons have been made across the continents. Parthasarathi's (1998) study of England and India supported the revision view, but a broader collection of data supports Smith's assessment (Allen 2007b). Comparisons with Japan and China also show that real wages there were like those of the backward parts of Europe. Even the Yangtze Delta, the most advanced region in China, had real wages on a par with those in Milan, not London or Amsterdam (Allen et al. 2007). While the Ottoman Empire has not received as much attention as east Asia in the revisionist historiography, Özmucur and Pamuk (2002) have shown that the real wage in Istanbul was also like that in Italy.

The worldwide conclusions require comparisons across regions with radically different diets. The comparisons are made by computing the cost of Smith's 'quantity of necessaries of life which is given to the labourer'. Figure 2 shows full-time, full-year earnings for a labourer deflated by the cost of maintaining a family on a mainly carbohydrate diet yielding 1,920 calories per adult male



Real Wage Rates (Historical Trends), Fig. 2 Subsistence ratio for labourers, various world cities, 1375–1875. Income/cost of subsistence basket (*Sources:* Allen (2001, 2007b); Allen et al. (2007))

equivalent. In each region, the cheapest available carbohydrate is used for the calculation. A value of 1 indicates that the labourer's wage equalled this 'bare bones' subsistence, and, indeed, that was the case in the 18th century in much of Europe and Asia. Living standards were higher, however, in Amsterdam and London.

Real Wages and Globalisation

The history of the global economy has attracted attention and real wages have played an important role in exposing its properties. Research on the 19th and 20th centuries has aimed to establish trends in real wages across countries as well as over time (Allen 1994; Williamson 1995). O'Rourke and Williamson (1999) argued that international trade and migration tightly bound economies and determined their relative factor prices. Trends in real wages, in other words, were determined by the evolution of the global economy rather than by the internal forces of capital accumulation and technical change that most previous theories have emphasised. In a study of the British economy, O'Rourke and Williamson (2005) argued that international factors determined factor prices from 1850 onwards and perhaps from as early as 1750. The relative importance of internal and external factors in determining the real wage is a lively area of current research.

See Also

- ► Economic Growth in the Very Long Run
- ► Industrial Revolution

Bibliography

Allen, R.C. 1994. Real incomes in the English-speaking world, 1879–1913. In *Labour market evolution: The economic history of market integration, wage flexibility and the employment relation*, ed. G. Grantham and M. MacKinnon. London: Routledge.

Allen, R.C. 2001. The great divergence in European wages and prices from the middle ages to the First World War. *Explorations in Economic History* 38: 411–447.

Allen, R.C. 2003a. *Farm to factory: A reassessment of the Soviet Industrial Revolution*. Princeton: Princeton University Press.

Allen, R.C. 2003b. Poverty and progress in early modern Europe. *Economic History Review* 56: 403–443.

Allen, R.C. 2007a. *Pessimism preserved: Real wages in the British industrial revolution*. Working Paper No. 314. Department of Economics, Oxford University.

Allen, R.C. 2007b. India in the great divergence. In *The new comparative economic history: Essays in Honor of Jeffery G. Williamson*, ed. T.J. Hatton, K.H. O'Rourke, and A.M. Taylor. Cambridge, MA: MIT Press.

Allen, R.C., J.-P. Bassino, D. Ma, C. Moll-Murata, and J.-L. van Zanden. 2007. *Wages, prices, and living* *standards in China, Japan, and Europe, 1738–1925*. Working Paper No. 316. Department of Economics, Oxford University.

Beveridge, L. 1965. *Prices and wages in England from the twelfth to the nineteenth century*. London: Frank Cass.

Bortz, J., and M. Aguila. 2006. Earning a living: A history of real wage studies in twentieth-century Mexico. *Latin America History Review* 41: 112–138.

Boskin, M.J., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson. 1996. *Toward a more accurate measure of the cost of living. Final report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index*. Washington, DC: Government Printing Office.

Clark, G. 2005. The condition of the working class in England, 1209–2004. *Journal of Political Economy* 113: 1307–1340.

Diewert, W.E. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4: 115–145.

Douglas, P.H. 1930. *Real wages in the United States: 1890–1926*. Boston: Houghton Mifflin.

Feinstein, C.H. 1998. Pessimism perpetuated: Real wages and the standard of living in Britain during and after the Industrial Revolution. *Journal of Economic History* 58: 625–658.

ILO (International Labour Organization). Various years. *Yearbook of labour statistics.* Geneva: ILO.

Kuznets, S. 1955. Economic growth and income inequality. *American Economic Review* 45: 1–28.

Lewis, W.A. 1954. Economic development with unlimited supplies of labour. *Manchester School of Economics and Social Studies* 22: 139–191.

Lindert, P.H., and J.G. Williamson. 1983. English workers' living standard during the Industrial Revolution: A new look. *Economic History Review* 36: 1–25.

Margo, R.A. 2000. *Wages and labor markets in the United States 1820–1860*. Chicago: University of Chicago Press.

Mitchell, B.R. 1998. *International historical statistics: Europe, 1750–1993*. London: Macmillan Reference.

O'Rourke, K.H., and J.G. Williamson. 1999. *Globalization and history.* Cambridge, MA: MIT Press.

O'Rourke, K.H., and J.G. Williamson. 2005. From Malthus to Ohlin: Trade, industrialisation and distribution since 1500. *Journal of Economic Growth* 10: 5–34.

Özmucur, S., and S. Pamuk. 2002. Real wages and standards of living in the Ottoman Empire, 1489–1914. *Journal of Economic History* 62: 293–321.

Parthasarathi, P. 1998. Rethinking wages and competitiveness in the eighteenth century: Britain and south India. *Past & Present* 158: 79–109.

Phelps Brown, E.H., and S.V. Hopkins. 1955. Seven centuries of building wages. *Economica NS* 22: 195–206.

Phelps Brown, E.H., and S.V. Hopkins. 1956. Seven centuries of the prices of consumables, compared with builders' wage rates. *Economica NS* 23: 296–314.

Pomeranz, K. 2000. *The Great Divergence: China, Europe, and the making of the modern world*. Princeton: Princeton University Press.

Rees, A. 1961. *Real wages in manufacturing, 1890–1914*. Princeton: Princeton University Press.

Rogers, J.E.T. 1866–92. *A history of agriculture and prices in England.* Vols. 7. Oxford: Clarendon Press.

Scholliers, P., and V. Zamagni. 1995. *Labour's reward: Real wages and economic change in 19th- and 20th-century Europe*. Aldershot: Edward Elgar.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*, ed. E. Cannan. New York: The Modern Library, 1937.

Solow, R.M. 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics* 70: 65–94.

van Zanden, J.L. 1999. Wages and the standard of living in Europe, 1500–1800. *European Review of Economic History* 3: 175–197.

Williamson, J.G. 1995. The evolution of global labor markets since 1830: Background evidence and hypotheses. *Explorations in Economic History* 32: 141–196.

Realization Problem

P. Kenway

The realization problem was first considered by classical economists such as Ricardo and Sismondi. Keynes's theory of effective demand has a bearing on it too. But it was Marx who gave it its most rounded – and controversial – treatment. At its simplest, the realization problem amounts to this: is there sufficient monetary demand for the commodities which have been produced to be sold, and sold at their value?

It is by no means obvious that there is really any problem at all. Why is the very act of production itself not enough to guarantee that there will be sufficient demand to ensure that all commodities will be sold? This was the view held strongly by Ricardo. His argument amounted to this: nobody produces except to sell and nobody sells except to buy something else. Marx showed that such arguments were wrong because they overlooked the specific nature of capitalist production (*see* CRISES).

The realization problem arises therefore because production under capitalism is but a phase within the circulation of capital, $M - C\ldots P\ldots\text{–}C'\text{–}M'$. Here, money is firstly

R

converted into means of production and labour-power ($M - C$). Production then takes place ($C\ldots P\ldots C'$). The produced commodities must then be sold ($C'–M'$), they must be reconverted into money, their value must be realized (Marx 1885, p. 709). This must happen if the circuit of capital is to be completed. That this must happen, and yet that it may not, is the realization problem.

Some of the features of the problem must be emphasized. The commodity has a value before it arrives on the market, this value being made up of the constant and variable capital consumed in its manufacture, along with the surplus value produced. By the time the question of realization arises, a certain level of output is presupposed, which depends particularly on the amount of capital thrown into production; and for the realization problem to be overcome, a certain level of monetary demand must be found in the sphere of circulation.

These aspects are derived from an analysis of the individual capital only. Whilst an investigation of the realization problem must take all of them into account, the problem can only fully be analysed in the context of the reproduction of the total social capital. This Marx did in his discussion of the reproduction schemes, in part three of Volume Two of *Capital*.

The Reproduction Schemes

The reproduction schemes can be viewed as abstract, two-sector models of the production and circulation of capital. Department one produces means of production. The value of its output is made up of $C_1 + V_1 + S_1$, where C_1 is the constant capital and V_1 the variable capital used up in production. S_1 is the surplus value produced. Department two produces means of consumption and the value of its product is likewise made up of $C_2 + V_2 + S_2$.

Marx considered two situations, simple and expanded reproduction. Simple reproduction is where capitalists devote all their surplus value to the purchase of consumption goods and seek only to produce in the next period at the same level as this. Expanded reproduction is where capitalists must accumulate some of their surplus value in order to obtain a larger stock of constant and variable capital, for use in the next period.

The point of the schemes was to investigate how the circulation must proceed in order for capital successfully to reproduce itself. This involves circulation both within and between the two departments. For example, simple reproduction requires that capitalists in department two acquire means of production to the value C_2 from department one in order to be able to produce again.

Two points should be noted. Firstly, when considering the reproduction of the total social capital, account must be taken of both value and use-value. This had not been necessary when considering the individual capital only. There, Marx had simply assumed that within the sphere of circulation would be found all the commodities necessary both to transform the capital value into new elements of production and commodities to satisfy workers' and capitalists' consumption (Marx 1885, p. 470).

Secondly, the scheme for expanded reproduction requires capitalists to accumulate (rather than consume) value out of this year's surplus value. It is important to emphasize that the amount accumulated must be a sufficient value to cover the *entire* amount of extra capital needed, both the extra constant capital and the extra variable capital.

This means that in department one $(1 - a_1)S_1$ must be equal in value to $dC_1 + dV_1$. Likewise, in department two $(1 - a_2)S_2$ must be equal in value to $dC_2 + dV_2$. (a denotes the portion of surplus value devoted by capitalists to consumption whilst the prefix d denotes the additional capital required). These, combined with the requirement that the supply of means of production must equal the demand for them:

$$\begin{aligned} C_1+V_1+S_1 &= C_1+C_2(\text{replacing what has been used up}) \\ &\quad +dC_1+dC_2(\text{the extra required for next year}) \end{aligned}$$

are sufficient to construct workable examples of capitalist reproduction. (See, for example, the numerical examples given in Marx 1885, pp. 586–91.)

Other relationships can be derived from these which must hold if reproduction is to proceed successfully. One such, the 'Bukharin condition' for expanded reproduction (Rosdolsky 1968, p. 449) is:

In other words, what department two needs to buy from department one ($C_2 + dC_2$) must equal what department one needs to buy from department two ($V_1 + dV_1 + a_1S_1$). In the case of simple reproduction this reduces to the more familiar expression: $C_2 = V_1 + S_1$.

Interpretation of the Reproduction Schemes

Analysis of the schemes shows that accumulation and the circulation of values and use-values can take place in such a way as to permit the successful resolution of the realization problem. The expansion of capital is shown to be possible. The theory must demonstrate this in view of the history of capitalist development. In so doing, Marx was refuting economists such as Sismondi who thought that expanded reproduction was impossible.

But one must be careful not to conclude too much from this result. The 'Austrian Marxists' for example concluded that the schemes showed that the reproduction cycle of capital need never break down. Hilferding went so far as to argue that the schemes proved that Marx had never been a supporter of the breakdown theory (cited in Rosdolsky 1968, p. 451).

This view is mistaken. The schemes cannot just be interpreted as if they are a model of the 'real world'. They are at a particular level of abstraction and leave out of account, for example, not only technical progress but also any impact of changes in either the organic composition of capital or the rate of surplus value.

More importantly however, the fact that the simultaneous consideration of value, use-value and accumulation does not uncover insurmountable difficulties is by no means the same thing as proving that the circuit of capital need never be broken or that the realization problem is never going to manifest itself as a real difficulty.

What the schemes show – or more properly illustrate, for it is a result of the *method* of Marx's argument – is something rather different: the realization problem can be solved this year, but that solution creates anew all the conditions which will ensure that the problem arises again next year. To solve the problem this year, values must once more be tied-up as capital which must next year be put to use to produce surplus value. These value must subsequently be realized. This year's solution is the seed from which next year's problem springs.

The Realization Problem and Gluts

The schemes also show the close connection between the realization problem and the potential, within the reproduction of capital, for general gluts of capital and commodities.

From the formulation of the reproduction schemes, it is easy to see that this year finishes up with a stock of means of production to be carried over to next year. This is not all that is carried over, however. For the value of output in department two ($C_2 + V_2 + S_2$) exceeds the value of consumption out of this year's income, wages ($V_1 + V_2$) plus capitalist consumption ($a_1S_1 + a_2S_2$). The excess amounts to the value of the additional variable capital to be accumulated ($dV_1 + dV_2$). This result is caused by the requirement that value be produced and accumulated to cover the *entire* amount of additional capital needed for production on an expanded scale, not just to cover the additional constant capital alone.

Thus both stocks of means of production and means of consumption are carried forward. Both grow in an orderly way if production grows smoothly and their value can be realized so long as this continues. But these stocks bear testimony to the fact that the process of reproduction contains the potential for a general glut, which in the first place can take the form of unused means of production and unsold consumption goods. This potential will not manifest itself so long as the realization problem is overcome. But the constant recurrence of the realization problem means that the potential of the general glut is constantly renewed.

The Realization Problem and Theory of Effective Demand

Finally, what is the relationship between this analysis and Keynes's theory of effective demand? The schemes certainly include the result that the level of output at which all output can be sold is the one at which net investment equals that part of surplus value not devoted to capitalist consumption (assuming that workers do not save). This comes over clearly, for example in the discussion of the difficulties posed for simple reproduction by depreciation, that is, where capital is not fully exhausted within the one year (Marx 1885, p. 528 et seq.).

There is, however, a significant difference between Marx and Keynes here. Whereas Keynes was investigating the 'theory of what determines the *actual employment* of the available resources (Keynes 1936, p. 4), Marx was concerned with the "theory" of what enables a given level and structure of output to be realized, to be sold, in order that production may begin anew'.

The theory of effective demand certainly sheds an interesting light on the realization problem. But Marx's investigation of the realization problem is part of a coherent whole. The fact that his analysis is firmly rooted in a theory of value shows this. In contrast, Keynes's theory was developed in opposition to the orthodox theory of value and output (which are of course one and the same theory). The theory of effective demand is beset with the difficulty of explaining *why* the monetary level of demand matters. Marx's analysis of the nature of capitalist production provides this (see Kenway 1980).

An explanation of what determines the actual employment of the available resources is most pertinent, especially during a slump. But the investigation of how, why and whether capitalism can produce and reproduce itself is surely the more profound and more general question.

See Also

► Crises
► Marx, Karl Heinrich (1818–1883)

Bibliography

Kenway, P. 1980. Marx, Keynes and the possibility of crisis. *Cambridge Journal of Economics* 4(1): 23–36.
Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan, 1973.
Marx, K. 1885. *Capital*, vol. II. Harmondsworth: Penguin, 1978.
Rosdolsky, R. 1968. *The making of Marx's 'Capital'*. London: Pluto, 1977.

Realized Volatility

Torben G. Andersen

Abstract
Realized volatility is a fully nonparametric approach to *ex post* measurement of the actual realized return variation over a specific trading period. It encompasses specific empirical procedures and an associated continuous-record asymptotic theory for arbitrage-free jump diffusions. It provides the ideal model-free benchmark for volatility model performance evaluation, and it has numerous natural areas of application within financial economics.

Keywords
Affine models of the term structure; Conditional return variance; Derivative prices; Econophysics; Forecasting; Historical volatility; Integrated variance; Jumps; Martingales; Microstructure noise; Parametric time series models; Realized volatility; Return volatility; Stochastic volatility models; Yield curve

JEL Classifications
C10

Return volatility is critical for a range of issues in financial economics. In theory, an asset price reflects its return covariation with economy-wide risk factors, often captured through its covariance with returns of factor-replicating financial portfolios, including the broad (stock) market. Hence, assessments of asset pricing, fund performance and portfolio allocation are all directly linked to expectations of the future volatility and covariability of financial assets. Likewise, individual asset volatilities are key inputs to derivatives pricing and risk management. Finally, in recent years volatility *realizations* have become the object of direct contracting as the payoff on so-called volatility swaps is determined by the future value attained by a specified measure of return volatility over the contract horizon.

As a consequence, return volatility has been studied extensively in the literature. Until recently, there were two dominant paradigms. One uses parametric time series models within the GARCH (Engle 1982; Bollerslev 1986) or (genuine) stochastic volatility (Shephard 2005) class to obtain conditional return variance estimates and forecasts. A second exploits market prices of volatility sensitive contracts, such as options, to back out the (implied) expected future volatility. Although the latter approach also conditions on a specific pricing model, it embodies a wider information set as market prices reflect the views of market participants, not just historical returns. However, derivative prices carry premiums for bearing volatility risk and thus provide a less direct measure of expected (physical) volatility. Hence, these measures are complementary and each likely provides independent information. More importantly in this context, they both focus on an a priori concept of return volatility, largely synonymous with the conditional variance. This is appropriate for many purposes as financial decisions are made subject only to current expectations regarding the future market environment. However, such measures are identified only through specific parametric representations. Moreover, the *ex ante* expectations differ from the subsequent (random) volatility realizations. The latter may be assessed only from *ex post* model-free measurements of the actual return variation. Such measures are obviously useful for assessing (volatility) model performance. In addition, if accurate measures of realized volatility are available it is natural to exploit these directly for modelling and forecasting. With increasing availability of intra-day tick-by-tick trade and quote data, this perspective has gained in popularity and a voluminous literature is evolving on the approach. This article presents a brief overview of these developments and associated empirical applications.

Historical Volatility

The concept of realized volatility refines and extends the *historical volatility* measure which has a fairly long precedent in the literature. To make the argument transparent, we initially consider an extremely simplified environment. Assume we are given the daily closing logarithmic asset price, denoted p_t. The associated daily continuously compounded return is then $r_t = p_t - p_{t-1}$. Moreover, assume the returns are conditionally mean zero with i.i.d. standardized residuals, that is,

$$r_t = \sigma_t \cdot z_t, \quad \text{with} \quad z_t \sim iid(0,1) \\ \text{and} \ \operatorname{Var}\left(z_t^2\right) = \omega. \tag{1}$$

The goal of realized volatility measurement is to provide a model-free estimator for the return variation or volatility, σ_t^2, given only the concurrent return observations. Obviously, if volatility is time-varying, this is problematic. We have, at the daily level, conditional on current volatility,

$$E\left[r_t^2\right] = \operatorname{Var}(r_t) = \sigma_t^2, \quad \operatorname{Var}\left(r_t^2\right) \\ = \omega \cdot \sigma_t^4, \ \text{and} \ E\left[r_t^2\right] / \left[\operatorname{Var}\left(r_t^2\right)\right]^{1/2} \\ = \omega^{-1/2}. \tag{2}$$

Hence, the concurrent squared return is an unbiased estimator of the underlying return variance. Unfortunately, it is extraordinarily noisy. The signal-to-noise ratio, defined as the mean of the estimator relative to the standard deviation,

R

equals $\omega^{-1/2}$. Invariably, $\omega > 1$ at daily (and lower) frequencies, so the standard deviation of (estimated) realized volatility exceeds the expected value. Matters improve if we assume constant volatility, at σ_t^2, over the month representing, say, K daily returns. Letting $r_{t:K} = r_t + \ldots + r_{t+K}$ denote the monthly return, we may exploit the historical volatility indicator, $\left[r_t^2 + \ldots + (r_{t+K})^2\right]$ rather than simply $(r_{t:K})^2$. We then have,

$$\begin{aligned} \mathrm{Var}(r_{t:K}) &= E\left[r_t^2 + \ldots + (r_{t+K})^2\right] \\ &= K \cdot \sigma_t^2, \text{ and } \mathrm{Var}\left[r_t^2 + \ldots + (r_{t+K})^2\right] K \cdot \omega \cdot \sigma_t^4. \end{aligned} \tag{3}$$

The signal-to-noise ratio for the monthly realized volatility is $(K/\omega)^{1/2}$ or a factor $K^{1/2}$ larger than for Eq. (2). Equation (3) may also readily be converted into an estimator for daily volatility based on the sample mean of the daily squared return over the month. This estimator is consistent with convergence rate $K^{1/2}$. However, given the simplifying assumptions, this measure is best viewed as an informal gauge of the underlying level of volatility. It was applied for computation of annual realized volatility from monthly data by Officer (1973) and monthly volatility series from daily data by, amongst others, Merton (1980), and French et al. (1987).

The two equivalent interpretations of Eq. (3) have distinct properties within a more realistic time-varying volatility setting. Since it is untenable to assume constant volatility for a month, or even one day, estimation of daily volatility from a surrounding set of daily returns covering, say, one month is inherently problematic. Nonetheless, the monthly realized volatility measure is robust to variation in volatility as explained more formally below. Of course, given the rapidly evolving markets, we would often need to assess the time variation in realized volatility at a daily level. The above reasoning suggests this will require access to high-frequency intra-day data. This is the starting point for the modern realized volatility literature.

Realized Volatility as an *Ex Post* Return Variability Measure

Given the round-the-clock activity on financial markets, volatility is naturally seen as evolving stochastically in continuous time. A complete characterization of the volatility realization then consists of a full specification of the actual sample path. However, at most one price is observed at each point in time so *instantaneous* volatility cannot be assessed without relying on adjacent observations, which is justified only under auxiliary assumptions – just as daily volatility cannot be estimated from a single return. In contrast, realized volatility seeks to measure the temporally cumulated (instantaneous) volatility so the target is the (average) realization of volatility over a non-negligible interval, allowing for a feasible and robust estimator with desirable properties.

As a benchmark for analysis, let the logarithmic asset price, $p(t)$, be a continuous time stochastic process observed in a frictionless market. For brevity and clarity, the formal exposition is cast in a univariate setting, but all results for realized volatility generalize readily to the multivariate case. To avoid arbitrage, and subject only to weak auxiliary conditions, the price process constitutes a (special) semi-martingale (Harrison and Kreps 1978; Back 1991). The price process may then quite generally be represented as follows,

$$\begin{aligned} dp(t) = \mu(t)dt + \sigma(t)dB(t) \\ + j(t)dJ(t), \quad t \in [0, T], \end{aligned} \tag{4}$$

where $\mu(t)$ is a predictable, continuous process with bounded variation, the volatility process $\sigma(t)$ is strictly positive, $B(t)$ denotes a standard Brownian motion, $J(t)$ is a jump indicator taking the values zero (no jump) or unity (jump) and, finally, the $j(t)$ indicates the jump in the return process if a jump occurs at time t and $j(t) \equiv 0$ otherwise. We assume the jump intensity, denoted $\lambda(t)$, to be bounded so there is a finite number of jump in the price path per time period. This is standard in the asset pricing literature although it does rule out some valid Lévy representations.

Equation (4) implies an instantaneous expected return of $\{\mu(t) + \lambda(t)E[j(t)]\}dt$, which is an order smaller than the instantaneous innovation, $\sigma(t)(dt)^{1/2} + j(t)$ namely order dt versus $dt^{1/2}$. Hence, for short horizons, the return variability is dominated by the unpredictable martingale component and the mean return is negligible. These features are captured formally by the notion of return quadratic variation defined below.

We denote the discretely observed continuously compounded return at time t, based on price observations at times t and $t - h$, for $h > 0$, by $r(t, h) = p(t) - p(t - h)$. From Eq. (4), the h-period return then has the representation,

$$\begin{aligned} r(t,h) &= p(t) - p(t-h) \\ &= \int_{t-h}^{t} \mu(\tau)d\tau + \int_{t-h}^{t} \sigma(\tau)dB(\tau) \\ &\quad + \sum_{t-h\le\tau<t} j(\tau). \end{aligned} \tag{5}$$

Formally, the sample path variation of the return process over $[t - h, t]$ is given by the *quadratic variation* of the logarithmic price process,

$$\begin{aligned} QV(t,h) &\equiv \int_{t-h}^{t} \sigma^2(s)ds + \sum_{t-h<s\le t} j^2(s) \\ &\equiv IV(t,h) + \sum_{t-h<s\le t} j^2(s). \end{aligned} \tag{6}$$

Equation (6) attributes the return variation to the diffusive volatility and the cumulative squared jumps. The first term is denoted the *integrated variance*. This quantity is often the focus of the broader realized volatility literature as many studies ignore jumps. The integrated variance is also of direct relevance for option pricing under stochastic volatility (Hull and White 1987). This is in part due to the following result for the special case with neither jumps nor correlation between the return and volatility processes, that is, $B(t)$ is independent of $\sigma(s)$ for all $0 \le s; t \le T$. Conditional on the mean component and integrated variance, we then have,

$$\begin{aligned} r(t,h)/\{\mu(t,h), IV(t,h)\} \\ \sim N(\mu(t,h), IV(t,h)), \end{aligned} \tag{7}$$

where $\mu(t,h) = \int_{t-h}^{t} \mu(\tau)d\tau$. Since (innovations to) the mean component is of smaller order than the integrated variance for low values of h, the dominant feature is the time-varying second moment given by the realizations of integrated variance. Hence, the return distribution is a normal mixture governed by the integrated variance process. If return and volatility innovations are correlated, the distribution is no longer mixed normal, but the interpretation of the integrated variance as a return variability measure is maintained.

Of course, short of having a continuum of price observations available, the relevant quantities in Eqs. (6) or (7) are not directly observable. However, in theory, the quadratic variation can be approximated closely by the corresponding cumulative squared return process, motivating the following definition. Let $[t - h, t]$ be split into $M = h/\Delta$ sub-intervals of length Δ, with $0 < \Delta \ll h$, and define the *realized volatility* constructed from the equally spaced Δ-period returns as

$$RV(t,h;\Delta) = \sum_{m=1}^{M} r^2(t-h+m\cdot\Delta, \Delta). \tag{8}$$

The basic theory for semi-martingales ensures that realized volatility is consistent for quadratic variation in the sense that, for finer and finer sampling of intra-day returns, Eq. (8) will, in the limit, provide a perfect measure of the realizations of the latent quadratic variation, that is,

$$\begin{aligned} RV(t,h;\Delta) \to QV(t,h), \quad \text{as} \\ \Delta \to 0\,(\text{and } M = h/\Delta \to \infty). \end{aligned} \tag{9}$$

This provides a formal basis for *ex post* measurement of realized volatility without parametric assumptions. It is a model-free measure of actual realizations while standard approaches provide parametric model forecasts of future volatility. We have the following approximate relationship between parametric forecasts and realized volatility, for small h,

$$\begin{aligned} Var[r(t,h)|\,\mathcal{F}_{t-h};\theta] &\simeq E[RV(t,h;\Delta)|\,\mathcal{F}_{t-h}] \\ &\simeq E[QV(t,h;\Delta)|\,\mathcal{F}_{t-h}], \end{aligned} \tag{10}$$

where the left most expression denotes the conditional variance over $[t - h, t]$ conditional on the available information at time $t - h, \mathcal{F}_{t-h}$, and the (true) model parameter vector, θ. Andersen, Bollerslev and Diebold (ABD) (2008) provide an in-depth discussion of the approximation behind Eq. (10). The relation shows that realized volatility is *the* natural benchmark for assessing volatility forecast performance.

Realized volatility can in principle be used to estimate the instantaneous volatility of a pure diffusion consistently. Ruling out jumps in the price process, but allowing volatility to be caglad (left-continuous, right limit sample path) and thus having potential discontinuities, we have

$$QV(t,h) \to \sigma^2(t), \quad \text{as} \quad h \to 0. \tag{11}$$

This insight is certainly not new. Merton (1976, 1980) discusses the result explicitly, and Foster and Nelson (1996) develop asymptotic results. However, this limiting operation is impractical. Equation (10) merely requires that $M = h/\Delta \to \infty$. For Eq. (11) to hold, with quadratic variation replaced by a feasible realized volatility estimator, Δ must converge to zero at a faster rate than h. This requires a double limiting procedure with ever more data sampled within an ever shrinking neighbourhood of t. Intensive sampling over short intervals magnifies the microstructure effects stemming from price discreteness, bid-ask bounce, temporary order-driven dependencies and other institutional features affecting returns at the highest frequencies. The issue of how to deal with such complications has inspired an extensive literature, summarized succinctly in Hansen and Lunde (2006). The practical complications induced by microstructure noise can be illustrated through the asymptotic theory for the realized volatility estimator.

For a purely diffusive price process, it follows from Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (BNS) (2002a, b, 2004a), that asymptotically, as $\Delta \to 0$,

$$\begin{aligned}\sqrt{h/\Delta}\big[RV(t,h;\Delta) - IV(t,h)\big] \\ \sim N\big(0.2 \cdot IQ(t,h)\big),\end{aligned} \tag{12}$$

so the ratio of the integrated quarticity, $IQ(t,h) = \int_{t-h}^{t} \sigma^4(s)ds$, to the number of intra-day returns, $M = h/\Delta$, determines the precision of the realized volatility estimator, generalizing and improving the result discussed below Eq. (3). A simple way to convey the implications of microstructure noise is to impose an exogenous bound, say Δ^*, below which no useful information from sampling is available as the semi-martingale assumption is blatantly violated at the highest frequencies. Hence, over an interval of length h, only $M^* = h/\Delta^*$ observations can be exploited for inference, and h must be of a certain size for realized volatility measures to possess meaningful precision. Equation (11) requires Δ to vanish at a rapid rate, so the bound Δ^* is quickly binding. If only a handful of returns is available within a few minutes of t, the sampling scheme behind Eq. (11) breaks down. In contrast, for Δ^* fixed at one or five minutes, it is feasible to estimate the quadratic variation with reasonable accuracy for h equal to one trading day.

Alternative Return Variation Measures

For a pure diffusion in a frictionless market, the basic realized volatility estimator exploiting all available observations is optimal for the quadratic variation. However, once price jumps and microstructure noise are introduced, matters become more complex. A number of issues may be addressed through alternative return variation measures, including the ability to disentangle the effect of jumps from the diffusive volatility, to estimate quantities needed for feasible inference about realized volatility, and to develop more robust-to-noise measures of integrated variance.

First, absent jumps and under appropriate regularity, one may extend the theory for the integrated variance to include the integrated variation of arbitrary powers, that is, for $\Delta \to 0$ the *realized power variation of order p* consistently estimates the p'th order integrated power variation ($p \le 1/2$),

$$\begin{aligned}&(E|Z|^p)^{-1}(h/\Delta)^{1-p/2}\sum_{m=1}^{M}|r(t-h+m\cdot\Delta,\Delta)|^p \\ &\to \int_{t-h}^{t} \sigma^p(s)ds\end{aligned} \tag{13}$$

where Z denotes a standard normal variable. For $p = 4$, this result provides a simple estimator for the integrated quarticity which may be plugged into Eq. (12) to yield a feasible distribution theory for realized volatility. An asymptotic distribution theory akin to Eq. (12) is available for realized power variation (see Barndorff-Nielsen et al. 2006b). Another extension, the *realized k-skip bipower variation*, is also consistent for the integrated variance,

$$BV(t,h;k;\Delta) \equiv \frac{\pi}{2} \sum_{m=k+1}^{M} |r(t-h+m\cdot\Delta,\Delta)$$

$$\| r(t-h+(m-k)\cdot\Delta,\Delta)| \rightarrow IV(t,h) \equiv \int_{t-h}^{t} \sigma^2(s)ds. \quad (14)$$

For $k = 1$, $BV(t,\ h;\ 1,\ \Delta) \equiv BV(t,\ h;\ \Delta)$, is termed the *realized bipower variation*. These measures have convenient robustness properties. First, Eq. (14) remains valid even if the return process follows a jump-diffusion. Hence, the bipower measures annihilate the jumps asymptotically and thus provide simple consistent estimators for the integrated variance. This allows for separation of the jump and diffusive contributions to the realized return variation, as

$$\begin{aligned} &RV(t,h;\Delta) - BV(t,h;\Delta) \\ &\quad \rightarrow QV(t,h) - IV(t,h) \\ &\quad = \sum_{t-h<s\leq t} j^2(s), \quad \text{as} \quad \Delta \rightarrow 0. \end{aligned} \quad (15)$$

Combined with the asymptotic theory for realized power and bipower variation, the result renders formal statistical tests for the presence and impact of jumps feasible. This is applied for separate analysis of the diffusive and jump components (see BNS 2004b, 2006; Huang and Tauchen 2005; ABD 2007a, b). Alternative jump tests have recently been developed by Jiang and Oomen (2005), Andersen et al. (2007), and Lee and Mykland (2006a).

The bipower variation measures also display robustness against microstructure noise. To first order, when not sampling at the very highest frequencies, the impact of noise may be mimicked by adding an i.i.d. process to the 'efficient' prices to generate noisy observations. This noise component induces negative return correlation which inflates the realized volatility estimator, resulting in a potentially strong upward bias. This may be alleviated by sampling more sparsely although this uses less data and reduces efficiency, pointing towards a bias-variance trade-off: one should sample sparsely enough to avoid a significant bias but frequently enough that efficiency is not compromised. An informal bias diagnostic is to apply the realized volatility estimator for different underlying frequencies over the full sample, that is, h is on the order of years or a decade. The long horizon minimizes sampling variability so that, absent the microstructure bias, all the measures should centre closely on the sample realized return volatility. A *volatility signature plot* depicts those realized volatility measures against the underlying sampling frequency with Δ ranging from seconds to a full day. For liquid financial markets, signature plots typically indicate inflated values at the highest frequencies which then decay quite smoothly to a stable level for sampling between 5 and 40 minutes. Andersen, Bollerslev, Diebold and Labys (ABDL) (2000a) suggest that the shortest sampling intervals not displaying a significant bias may be desirable choices. Of course, this criterion rewards unbiasedness over efficiency. Bandi and Russell (2005a) explicitly trade off the microstructure bias with the efficiency gains from more data. An alternative is to apply skip-k bipower variation measures. Huang and Tauchen (2005) document that these tend to work well when applied to noisy observations from jump diffusions. Andersen, Bollerslev, Frederiksen and Nielsen (ABFN) (2006a) extend the volatility signature plots to include both power variation and skip-k bipower variation measures. Such generalized plots provide insights into the robustness of the realized quarticity and other quantities used for jump tests and may guide the choice of an adequate sampling frequency for analysis of a range of different issues in the presence of market frictions.

Finally, procedures have been developed to correct for microstructure bias while utilizing more of the available data. The proposals include the subsampling idea alluded to in Zhou (1996), and extended and formalized by Zhang et al. (2005) and Aït-Sahalia et al. (2006) as well as the kernel based methods of Barndorff-Nielsen, Hansen, Lunde and Shephard (2006a, b), and numerous filtering approaches discussed in Hansen and Lunde (2006).

Empirical Applications

Hsieh (1991) is perhaps the first to apply intra-day returns for historical volatility measurement of the daily return variation. Closely related work appears in publications by the Olsen & Associates group. This is surveyed in Dacorogna et al. (2001). Zhou (1996) offers the first systematic study of the realized volatility estimator combining theoretical and empirical issues. Interestingly, he discusses contamination by market microstructure noise as well as ideas for a variety of feasible corrections. Comte and Renault (1998) also comment on estimating diffusive spot volatility through the empirical counterpart to quadratic variation.

In parallel work, Andersen and Bollerslev (AB) (1997, 1998b) explore the dynamics of high-frequency return volatility, documenting the striking effectiveness of cumulative absolute and squared returns as daily volatility measures. These findings inspired theoretical inquiries and are followed by the statement of consistency of realized volatility for the quadratic variation for a general multivariate jump-diffusion setting in ABDL (2001a, b). In concurrent work, BNS (2001, 2002a) and Meddahi (2002) provide initial asymptotic theory for the realized volatility estimator, with the diffusive multivariate case treated in BNS (2004a).

Empirical work almost invariably operates with h equal to one trading day (or more). This is due to the pronounced intra-day volatility pattern which induces systematic shifts in the quadratic return variation over different segments of the trading day. As noted in AB (1997, 1998b) this type of largely deterministic effects are alleviated at the daily frequency, rendering this the natural basis for analysis. The list of empirical applications is growing rapidly. A brief overview of the topics explored along with selective, but not exhaustive, references are provided below.

The most common use of realized volatility is as a basis for volatility forecasting and evaluation. AB (1998a) document the potential of realized volatility for assessment of standard volatility forecasts, a theme further explored in Andersen et al. (1999) and rationalized more formally in Andersen et al. (2004) using powerful analytical techniques developed in Meddahi (2001). An integrated approach to measurement, modelling and forecasting of realized volatility is developed in ABDL (2003a, b). Ghysels et al. (2006) show that a combination of realized volatility measures for different frequencies and horizons may enhance forecast performance. Engle and Gallo (2006) follow a similar strategy but with a different modelling approach. ABD (2007a, b) improve performance by separating the jump and diffusive volatility components in the forecasting procedure. Other studies on the presence and importance of price jumps using realized volatility related jump statistics are ABFN (2006b), Tauchen and Zhou (2006), and Fleming and Paye (2006), Andersen et al. (2007), Jiang and Oomen (2005), and Lee and Mykland (2006). In the same spirit, Liu and Maheu (2005) find that more jump-robust power variation measures are preferred to realized volatility for forecasting purposes. Earlier studies of forecast performance include Blair et al. (2001) and Martens (2002). Some initial studies of the role of microstructure noise and discretization error for forecasting and forecast evaluation is provided by Aït-Sahalia and Mancini (2006), ABM (2005a, b, 2006a, b), and Ghysels and Sinko (2006). The issue of how to include (noisy) overnight return information into the volatility measures and forecasts is addressed by Fleming et al. (2003) and Hansen and Lunde (2005).

The evidence for long-range persistence in volatility is particularly striking when analysed via realized volatility rather than via daily return observations. AB (1998a) demonstrate that return series spanning only a couple of years are sufficient to identify a distinct hyperbolic decay in the realized power variation series. Moreover, the implied degree of fractional integration appears to be stable across subsamples, at around 0.35–0.45, implying a stationary volatility series. This finding is confirmed by virtually all subsequent studies of realized volatility exploring the issue, including ABDL (2001a, b, 2003a, b), Areal and Taylor (2002), Martens (2002), Zumbach (2004), Deo et al. (2005), and Deo et al. (2006). Moreover, related early work by the Olsen & Associates group also note the presence of scaling laws in volatility measures obtained from high frequency returns, see, for example, Müller et al. (1990) and the review in Dacorogna et al. (2001). This inspired an extensive amount of empirical work in the 'econophysics' area on volatility scaling laws which is summarized in Mantegna and Stanley (2000). These robust empirical results suggest that the long-memory property is not driven by occasional structural breaks but is present in the data generating process at high frequencies. As such, it sheds new light on a contentious issue which is not readily resolved without an effective volatility measure that improves the signal-to-noise ratio, allowing for more effective inference. Explicit estimation of the long memory features may be circumvented through a model with multiple volatility components, each governed by an autoregressive process, as such structures approximate the long-memory dependencies very well. This approach has been applied for realized volatility series by ABD (2007a, b), BNS (2001), and Corsi (2003), among others.

The no-arbitrage implications for the return dynamics expressed in Eqs (4) and (5), are necessarily quite weak and general, but do nonetheless have distributional implications which are potentially testable. As highlighted by Eq. (7), auxiliary assumptions produce a mixture of normals result akin to the mixture-of-distributions theory, originating from work by Clark (1973) and Tauchen and Pitts (1983). The novelty of Eq. (7) is the potential *ex post* observability of the mixing variable, the integrated variance, which enables direct inference regarding the distributional implications without any parametric assumptions. In practice there will be some discretization error and microstructure distortions, but the properties of the realized volatility estimator should facilitate powerful tests. ABDL (2000b) confirm that returns standardized by realized volatility are much closer to normal than the standardized residuals from the usual volatility models based on daily data, even if the normality is not exact. Thomakos and Wang (2003) reach the identical conclusion for a different set of assets. Of course, Eq. (7) is not valid in the presence of price jumps or dependence between the volatility and return innovation processes. The former issue may be addressed through jump identification procedures which seek to annihilate the impact of the jumps. The latter issue is accommodated by sampling the return process in 'financial time', consisting of calendar periods representing equal increments to the integrated variance process, as noted by Peters and de Vilder (2006) for a diffusive return process. This approach is extended to a jump-diffusive setting by Andersen et al. (2007) who also explore practical implementation issues for the associated distributional tests in detail. More extensive data sets and alternative jump identification techniques are considered in ABFN (2006b). They quite generally obtain jump-adjusted, financial-time sampled returns that are indistinguishable from i.i.d. standard Gaussian variates through realized volatility based empirical procedures. In sum, the results corroborate the general framework, and the tools developed for jump identification and measurement of the quadratic variation deliver empirically meaningful series of jumps and quadratic variation which are fully consistent with the theoretical underpinnings. In the process, direct evidence of the importance of jumps and the asymmetric return-volatility relationship is provided. Jumps are present for all asset classes and constitute a non-negligible fraction of overall return variation. For equities, negative returns tend to induce higher volatility than corresponding positive returns, a feature broadly recognized in the prior literature. Interestingly, there are also signs of significant asymmetric relationships for other asset classes, although both magnitude and sign may change over time. Similar issues are studied by Fleming and Paye (2006) and closely related topics are explored by Maheu and McCurdy (2002).

Multivariate applications of the realized volatility estimator are in principle straightforward. They are used to study the broad correlation patterns among individual stocks in ABDE (2001a, b), for volatility timing in portfolio allocation in Fleming et al. (2003), for estimation of systematic market risk exposure (time-varying market betas) in Andersen et al. (2005a, b), for assessment of a broader set of risk loading coefficients in Bollerslev and Zhang (2003), and for dynamic portfolio choice in Bandi et al. (2008). In spite of these initial explorations, it is clear that the multivariate setting introduces additional practical complications as there is evidence of significant delays in the reaction of one security price to movements in another related asset. Sheppard (2006) explores these issues in some depth even if general prescriptions for practice do not directly follow. The favoured approach in current work is to include temporal cross-correlation patterns through measurement of the relations between the return of one asset with lead and lag returns for the other asset – in the spirit of the corrections for non-trading effects in the estimation of betas from daily data in Scholes and Williams (1977). Hayashi and Yoshida (2005, 2006) and Griffin and Oomen (2006) study such realized covariance and correlation estimators. Bandi and Russell (2005b) and Zhang (2006) seek to trade off bias and efficiency optimally. An alternative approach is proposed in Bauer and Vorkink (2006).

A natural comparison for realized volatility based forecasts is with forecasts implied by traded financial contracts such as options. In fact, an intriguing parallel exists between expected future realized volatility and the pricing of volatility swaps, with the latter reflecting the expected future integrated variance under the risk-neutral (pricing) measure if jumps in the price path are absent. Hence, systematic differences between realized volatility and implied volatility measures reflect the market prices of volatility risk (see, for example, Britten-Jones and Neuberger 2000; Carr and Madan 1998; Carr and Wu 2005). Moreover, Bondarenko (2004) shows that the implied volatility result remains approximately valid for jump diffusions and returns sampled at discrete intervals only. Recent empirical papers on the performance of realized volatility forecasts versus implied volatility measures include, for example, Andersen et al. (2006), Bollerslev et al. (2005), Bollerslev and Zhou (2006), Busch et al. (2006), Chan et al. (2006) and Pong et al. (2004). The findings confirm that realized volatility forecasts contain information for future return variability over-and-beyond implied volatility forecasts, while standard volatility forecasts obtained from models utilizing only daily data are fully encompassed by the market based measures.

Future Directions for Research

Most of the empirical work associated with the realized volatility concept has focused directly on the measurement precision and forecast performance. In order for the approach to enter routinely in more mainstream applications within asset pricing, risk management and portfolio allocation the empirical studies must broaden in scope. As reviewed above, this has begun to happen, but much work remains. One recent example of using the concept for model specification testing is Andersen and Benzoni (2005). The study documents a serious deficiency in the ability of affine term structure models to accommodate the observed dynamics of realized yield volatility for the US Treasury market. Theoretically, the concurrent yield curve should span yield volatility, both *ex post* and *ex ante*, but this property is systematically violated as the yield variation at every maturity displays genuine stochastic features not associated with simultaneous directional shifts in the yield curve. The result extends earlier findings based on *ex ante* volatility measures at the monthly frequency. For generalizations of term structure models operating within the popular and tractable affine setting, the realized volatility measures promise to be valuable diagnostic tools.

The applications of realized volatility measures will surely continue to grow and broaden as the advantages of the enhanced precision in the measurement of the return variability are much too large to be ignored and many important questions await thorough analysis from this perspective. Of particular interest is the development of practical approaches to generate reliable measures for the high-dimensional case involving a large set of assets. On the theoretical front, work remains in terms of understanding the relative advantages of the different robust alternatives to the basic realized volatility measure. One potentially promising avenue is to further develop the locally constant volatility technique developed by Mykland (2006), seeking to provide a formal, yet simple and powerful, tool for asymptotic theory while allowing for time-varying price dynamics.

See Also

- ▶ ARCH Models
- ▶ Continuous and Discrete Time Models
- ▶ Kernel Estimators in Econometrics
- ▶ Law(s) of Large Numbers
- ▶ Long Memory Models
- ▶ Martingales
- ▶ Mean-Variance Analysis
- ▶ Measurement Error Models
- ▶ Mixture Models
- ▶ Options
- ▶ Stochastic Volatility Models
- ▶ Wiener Process

Bibliography

Aït-Sahalia, Y., and L. Mancini. 2006. *Out of sample forecasts of quadratic variation*. Manuscript, Princeton University and University of Zürich.

Aït-Sahalia, Y., P.A. Mykland, and L. Zhang. 2006. *Ultra high frequency volatility estimation with dependent microstructure noise*. Manuscript, Princeton University.

Andersen, T.G., and L. Benzoni. 2005. *Can bonds hedge volatility risk in the U.S. Treasury market? A specification test for affine term structure models*. Manuscript, Kellogg School, Northwestern University and University of Minnesota.

Andersen, T.G., and T. Bollerslev. 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance* 52: 975–1005.

Andersen, T.G., and T. Bollerslev. 1998a. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39: 885–905.

Andersen, T.G., and T. Bollerslev. 1998b. Deutschemark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53: 219–265.

Andersen, T.G., T. Bollerslev, and F.X. Diebold. 2007a. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89.

Andersen, T.G., T. Bollerslev, and F.X. Diebold. 2008. Nonparametric volatility measurement. In *Handbook of financial econometrics*, ed. L.P. Hansen and Y. Aït-Sahalia. Amsterdam: North-Holland.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics* 61: 43–76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2000a. Great realizations. *Risk* 13: 105–108.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2000b. Exchange rate returns standardized by realized volatility are (nearly) Gaussian. *Multinational Finance Journal* 4: 159–179.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2001b. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96: 42–55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2003a. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and C. Vega. 2003b. Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review* 93: 38–62.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and G. Wu. 2005a. A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review* 95: 398–404.

Andersen, T.G., T. Bollerslev, and D. Dobrev. 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics* 138: 125–180.

Andersen, T.G., T. Bollerslev, P.H. Frederiksen, and M.Ø. Nielsen. 2006c. Comment on realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24: 173–179.

Andersen, T.G., T. Bollerslev, P.H. Frederiksen, and M.Ø. Nielsen. 2006d. *Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns*. Manuscript, Kellogg School, Northwestern University.

Andersen, T.G., T. Bollerslev, and S. Lange. 1999. Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* 6: 457–477.

Andersen, T.G., T. Bollerslev, and N. Meddahi. 2004. Analytic evaluation of volatility forecasts. *International Economic Review* 45: 1079–1110.

Andersen, T.G., T. Bollerslev, and N. Meddahi. 2005b. Correcting the errors: volatility forecast evaluation

R

using high-frequency data and realized volatilities. *Econometrica* 73: 279–296.

Andersen, T.G., T. Bollerslev, and N. Meddahi. 2006a. *Market microstructure noise and realized volatility forecasting*. Manuscript, Kellogg School, Northwestern University.

Andersen, T.G., P.H. Frederiksen, and A.D. Staal. 2006b. *The information content of realized volatility forecasts*. Manuscript, Kellogg School, Northwestern University.

Areal, N.M.P.C., and S.J. Taylor. 2002. The realized volatility of FTSE-100 futures prices. *Journal of Futures Markets* 22: 627–648.

Back, K. 1991. Asset prices for general processes. *Journal of Mathematical Economics* 20: 371–395.

Bandi, F., and J.R. Russell. 2005a. *Microstructure noise, realized volatility, and optimal sampling*. Manuscript, Graduate School of Business, University of Chicago.

Bandi, F., and J.R. Russell. 2005b. *Realized covariation, realized beta and microstructure noise*. Manuscript, Graduate School of Business, University of Chicago.

Bandi, F., J.R. Russell, and Y. Zhou. 2008. Using high-frequency data in dynamic portfolio choice. *Econometric Reviews*, forthcoming.

Barndorff-Nielsen, O.E., S.E. Graversen, J. Jacod, and N. Shephard. 2006a. Limit theorems for bipower variation in econometrics. *Econometric Theory* 22: 677–720.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2006b. *Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise*. Manuscript, Nuffield College, Oxford.

Barndorff-Nielsen, O.E., and N. Shephard. 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B* 63: 167–241.

Barndorff-Nielsen, O.E., and N. Shephard. 2002a. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64: 253–280.

Barndorff-Nielsen, O.E., and N. Shephard. 2002b. Estimating quadratic variation using realised variance. *Journal of Applied Econometrics* 17: 457–477.

Barndorff-Nielsen, O.E., and N. Shephard. 2003. Realized power variation and stochastic volatility models. *Bernoulli* 9: 243–265.

Barndorff-Nielsen, O.E., and N. Shephard. 2004a. Econometric analysis of realised covariation: High-frequency covariance, regression and correlation in financial econometrics. *Econometrica* 72: 885–925.

Barndorff-Nielsen, O.E., and N. Shephard. 2004b. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2: 1–37.

Barndorff-Nielsen, O.E., and N. Shephard. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4: 1–30.

Bauer, G.H., and K. Vorkink. 2006. *Multivariate realized stock market volatility.* Manuscript, Bank of Canada and Sloan School, Massachusetts Institute of Technology.

Blair, B.J., S.-H. Poon, and S.J. Taylor. 2001. Forecasting S&P 100 volatility: The incremental information content of implied volatility and high-frequency index returns. *Journal of Econometrics* 105: 5–26.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Bollerslev, T., M. Gibson, and H. Zhou. 2005. *Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities*. Manuscript, Department of Economics, Duke University.

Bollerslev, T., and H.O. Mikkelsen. 1996. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73: 151–184.

Bollerslev, T., and B.Y.B. Zhang. 2003. Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance* 10: 533–558.

Bollerslev, T., and H. Zhou. 2006. Volatility puzzles: A simple framework for gauging return-volatility regressions. *Journal of Econometrics* 131: 123–150.

Bondarenko, O. 2004. *Market price of variance risk and performance of hedge funds*. Manuscript, University of Illinois at Chicago.

Britten-Jones, M., and A. Neuberger. 2000. Option prices, implied price processes, and stochastic volatility. *Journal of Finance* 55: 839–866.

Busch, T., B.J. Christensen, and M.Ø. Nielsen. 2006. *The role of implied volatility in forecasting realized volatility and jumps in foreign exchange, stock and bond markets*. Manuscript, University of Aarhus and Cornell University.

Carr, P., and D. Madan. 1998. Towards a theory of volatility trading. In *Volatility: New estimation techniques for pricing derivatives*, ed. R. Jarrow. London: Risk Books.

Carr, P., and L. Wu. 2005. *Variance risk premia*. Manuscript, Courant Institute, New York University.

Chan, W.H., M. Kalimipulli, and R. Sivakumar. 2006. *The economic value of using realized volatility in the index options market*. Manuscript, Wilfrid Laurier University and University of Waterloo.

Clark, P.K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41: 135–155.

Comte, F., and E. Renault. 1998. Long memory in continuous time stochastic volatility models. *Mathematical Finance* 8: 291–323.

Corsi, F. 2003. *A simple long-memory model of realized volatility.* Manuscript, University of Southern Switzerland.

Dacorogna, M.M., R. Gencay, U. Müller, R.B. Olsen, and O.V. Pictet. 2001. *An introduction to high-frequency finance*. San Diego, CA: Academic Press.

Deo, R., M. Hsieh, and C.M. Hurvich. 2005. *Tracing the source of long memory in volatility.* Manuscript, Stern School, New York University.

Deo, R., C.M. Hurvich, and Y. Lu. 2006. Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131: 29–58.

Engle, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50: 987–1008.

Engle, R.F., and G.M. Gallo. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131: 3–27.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. The economic value of volatility timing using realized volatility. *Journal of Financial Economics* 67: 473–509.

Fleming, J., and B.S. Paye. 2006. *High-frequency returns, jumps and the mixture-of-normals hypothesis*. Manuscript, Jones School, Rice University.

Foster, D., and D.B. Nelson. 1996. Continuous record asymptotics for rolling sample estimators. *Econometrica* 64: 139–174.

French, K.R., G.W. Schwert, and R.F. Stambaugh. 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19: 3–29.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2006. Predicting volatility: How to get the most out of returns data sampled at different frequencies. *Journal of Econometrics* 131: 59–95.

Ghysels, E., and A. Sinko. 2006. *Volatility forecasting and microstructure noise*. Manuscript, University of North Carolina, Chapel Hill.

Griffin, J.E., and R.C.A. Oomen. 2006. *Covariance measurement in the presence of non-synchronous trading and market microstructure noise*. Manuscript, University of Warwick.

Hansen, P.R., and A. Lunde. 2005. A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* 3: 525–554.

Hansen, P.R., and A. Lunde. 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24: 127–161.

Harrison, J.M., and D. Kreps. 1978. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20: 381–408.

Hayashi, T., and N. Yoshida. 2005. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11: 359–379.

Hayashi, T., and N. Yoshida. 2006. *Estimating correlations with nonsynchronous observations in continuous diffusion models*. Manuscript, Columbia University and University of Tokyo.

Hsieh, D.A. 1991. Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance* 46: 1839–1877.

Huang, X., and G. Tauchen. 2005. The relative contribution of jumps to total price variation. *Journal of Financial Econometrics* 3: 456–499.

Hull, J., and A. White. 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42: 281–300.

Jacod, J., and P. Protter. 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* 26: 267–307.

Jiang, G.J, and R.C.C. Oomen. 2005. *A new test for jumps in asset prices*. Manuscript, Eller College of Management, University of Arizona and Warwick Business School.

Lee, S.S, and P.A. Mykland. 2006a. *Jumps in real-time financial markets: a new nonparametric test and jump dynamics*. Manuscript, University of Chicago.

Liu, C., and J.M. Maheu. 2005. *Modeling and forecasting realized volatility: The role of power variation*. Manuscript, University of Toronto.

Maheu, J.M., and T.H. McCurdy. 2002. Nonlinear features of realized FX volatility. *Review of Economics and Statistics* 84: 668–681.

Mantegna, R.N., and H.E. Stanley. 2000. *An introduction to econophysics – correlations and complexity in finance*. Cambridge, MA: Cambridge University Press.

Martens, M. 2002. Measuring and forecasting S&P500 index-futures volatility using high-frequency data. *Journal of Futures Markets* 22: 497–518.

Meddahi, N. 2001. *An Eigen function approach for volatility modeling*. Manuscript, Department of Economics, University of Montreal and CIRANO.

Meddahi, N. 2002. A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics* 17: 479–508.

Merton, R.C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3: 125–144.

Merton, R.C. 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8: 323–361.

Müller, U.A., M.M. Dacarogna, R.B. Olsen, O.V. Pictet, M. Schwarz, and C. Morgenegg. 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance* 14: 1189–1208.

Mykland, P.A. 2006b. *A Gaussian calculus for inference from high frequency data*. Manuscript, University of Chicago.

Officer, R.R. 1973. The variability of the market factor of New York stock exchange. *Journal of Business* 46: 434–453.

Peters, R.T., and R.G. de Vilder. 2006. Testing the continuous semimartingale hypothesis for the S&P 500. *Journal of Business & Economic Statistics* 24: 444–454.

Pong, S., M.B. Shackleton, S.J. Taylor, and X. Xu. 2004. Forecasting currency volatility: A comparison of implied volatilities and ARFIMA models. *Journal of Banking & Finance* 28: 2541–2563.

Protter, P. 1992. *Stochastic integration and differential equations: A new approach*. 2nd ed. New York: Springer.

Scholes, M., and J. Williams. 1977. Estimating betas from nonsynchronous data. *Journal of Financial Economics* 5: 309–327.

Shephard, N. 2005. *Stochastic volatility. selected readings*. Oxford: Oxford University Press.

Sheppard, K. 2006. *Realized covariance and scrambling*. Manuscript, University of Oxford.

Tauchen, G.E., and M. Pitts. 1983. The price variability-volume relationship on speculative markets. *Econometrica* 51: 485–505.
Tauchen, G.E., and H. Zhou. 2006. *Identifying realized jumps on financial markets*. Manuscript, Duke University and Federal Reserve Board of Governors.
Taylor, S.J. 1986. *Modeling financial time series*. Chichester, UK: Wiley.
Thomakos, D.D., and T. Wang. 2003. Realized volatility in the futures markets. *Journal of Empirical Finance* 10: 321–353.
Zhang, L. 2006. *Estimating covariation: Epps effect, microstructure noise*. Manuscript, University of Chicago at Illinois.
Zhang, L., P.A. Mykland, and Y. Aït-Sahalia. 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100: 1394–1411.
Zhou, B. 1996. High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics* 14: 45–52.
Zumbach, G. 2004. Volatility processes and volatility forecasts with long memory. *Quantitative Finance* 4: 70–86.

Rebound Effects

Kenneth Gillingham

Abstract

In environmental and energy economics, rebound effects may influence the energy savings from improvements in energy efficiency. When the energy efficiency of a product or service improves, it becomes less expensive to use, income is freed-up for use on other goods and services, markets re-equilibrate and there may even be induced innovation. These effects typically reduce the direct energy savings from energy efficiency improvements, but lead to improved social welfare as long as there are not sufficiently large externality costs. There is strong empirical evidence that rebound effects exist, yet estimates of the different effects range widely depending on context and location.

Keywords

Backfire; Climate policy; Derived demand; Emissions; Energy efficiency; Greenhouse gases; Take-back effect; Welfare

JEL Classification

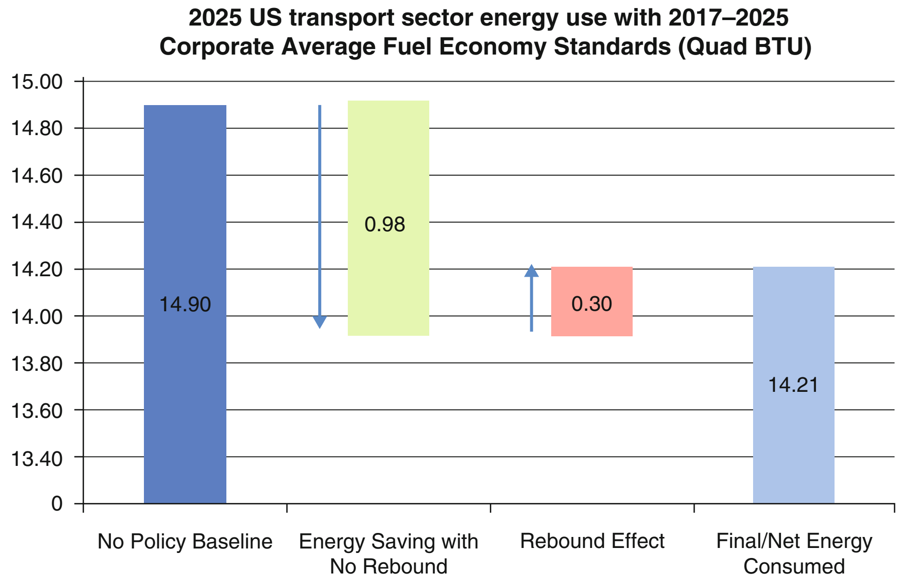
H23; Q38; Q41

Introduction

Energy efficiency policies are among the most common environmental policies around the world. Holding consumer, producer and market responses constant, an increase in energy efficiency for an energy-using durable good, such as a vehicle or refrigerator, will unambiguously save energy. Rebound effects are consumer, producer and market responses to an increase in energy efficiency that typically reduce the energy savings that would have occurred had these responses been held constant. The use of the term 'rebound'is intuitive: the responses lead to a rebounding of energy use back towards the energy use prior to the energy efficiency improvement. For this reason rebound effects are sometimes also called 'take-back'effects, for some of the energy savings are 'taken back' by the responses.

Often rebound effects are referred to in the singular, as 'the rebound effect', but it is widely understood that there are actually several effects at work. At one extreme, these rebound effects can lead to *additional* energy use above the amount used prior to the energy efficiency improvement. This is often called 'backfire', referring to the energy efficiency improvement 'backfiring' in terms of saving energy. At the other extreme, *negative* rebound effects, whereby the responses increase the energy savings, may be possible as well. In economics, rebound effects occur most often in reference to energy use, but of course rebound effects can also be described in terms of greenhouse gas emissions or other measures of environmental impact. Moreover, rebound effects are possible in other areas as well, such as materials or water. This article follows the convention and focuses on rebound effects in energy use.

Figure 1 illustrates the importance of rebound effects for the energy savings that can be expected from the 2017–2025 Corporate Average Fuel Economy Standards in the USA. This illustration assumes a 30% total rebound effect, which would 'take back' 0.30 quadrillion BTU of the energy



Rebound Effects, Fig. 1 2025 US transport sector energy use with the 2017–2025 CAFE standards illustrating how rebound effects may influence energy savings (Gillingham et al. 2013). The figure assumes a 30% rebound effect for illustrative purposes

savings that would have been expected from the policy.

The first mention of rebound effects goes back to the English economist William Stanley Jevons in 1865. Jevons lived at a time when coal-fired steam engine technology was dramatically improving in England. Yet, despite improvements in engine efficiency, coal use was not declining, but rather was increasing. Jevons attributed this to the improved productivity of coal use, leading to more investment and growth in the sectors of the economy that used coal. Jevons famously stated, ‘It is a confusion of ideas to suppose that the economical use of fuel is equivalent to diminished consumption. The very contrary is the truth’ (Jevons 1865). The backfire that Jevons was positing has more recently been called the ‘Jevons Paradox’, for it seems paradoxical to have an energy efficiency improvement lead to more energy use.

More recently, policymakers and academics have been increasingly interested in rebound effects from energy efficiency policies. If there are large rebound effects leading to backfire, energy efficiency policies may not save energy at all. Moreover, larger rebound effects would be expected to widen the welfare difference between first-best policies to address market failures, such as price policies to correct for externalities, and energy efficiency policies, which are generally considered second-best policies (unless there are behavioural failures, as described in Gillingham and Palmer (2014) and many other papers). Such observations have led to a vibrant academic and policy debate over the magnitude of the effects.

Rebound effects are often broadly classified into their microeconomic and macroeconomic effects. We proceed by discussing each and then turn to quantification and policy implications.

Microeconomic Rebound Effects

The microeconomic rebound effect may occur for both consumers and firms when there is an energy efficiency improvement. Most of the economic literature focuses on consumers, so we begin here with the microeconomic rebound effect from the consumer perspective.

The microeconomic rebound effect for consumers captures the change in the consumption bundle of all goods and services when there is an energy efficiency improvement in one product.

For consumers, it stems from the classic substitution and income effects of a price change, as in the Slutsky equation (Gillingham et al. 2015). Unfortunately, the literature is inconsistent in the terminology used, so it is instructive to begin with some simple microeconomic theory. The following exposition loosely follows Borenstein (2015).

Suppose we have an energy efficiency improvement in an energy-using good 0. For instance, this could be a vehicle, light or air conditioner. Let the original energy intensity (i.e. the reciprocal of energy efficiency) of the good be e_0 and the new energy intensity be e_0' With greater energy efficiency, the cost or price of using good 0 drops (i.e. the energy cost of the energy service drops). Denote the price of using good 0 as p_0 and the change in price with the efficiency improvement as Δp_0. Similarly, let the demand for good 0 be q_0. At the same time that the price decreases, there may also be a cost (C) associated with the energy efficiency improvement.

The combination of such a price decline and change in income leads to a consumer response in four ways as they re-optimise their consumption bundle.

First, there is a substitution effect, whereby the consumer substitutes from other goods and services to good 0 along the Hicksian (compensated) demand curve of the use of good 0. The amount of increased consumption is just the marginal change in the Hicksian demand with a price change $\left(\frac{\partial q_0^H}{\partial p_0}\right)$ times the change in the price of usage Δp_0. The increased consumption of course uses energy. The change in energy use from the substitution effect is the new energy intensity of the good times the increased consumption: $e_0' \frac{\partial q_0^H}{\partial p_0} p_0$.

Second, there is an income effect. Since the consumer is no longer spending as much on using good 0, the consumer may have income to re-spend ($q_0 \Delta p_0$). For example, if a consumer is now spending less to drive a mile, they have increased purchasing power. However, the energy efficiency improvement may come with a cost, so the total change in income is $q_0 \Delta p_0 - C$. This change of income may be positive or negative depending on how costly the efficiency improvement is. Furthermore, it can be expected to influence the demand for using good 0. Let the demand for using good 0 be given by q_0 and income be given by I, so the marginal change in demand with a change in income is $\frac{\partial q_0}{\partial I}$. Thus, the change in demand for using good 0 with the energy efficiency improvement is given by $\frac{\partial q_0}{\partial I}(q_0 \Delta p_0 - C)$and the change in energy use due to the income effect is then $e_0' \frac{\partial q_0}{\partial I}(q_0 \Delta p_0 - C)$.

Third, there may be a substitution effect for every other good and service. For goods that are substitutes for good 0, there would be a decrease in consumption with a price decrease of good 0. For complements, the opposite. Let the Hicksian (compensated) demand for good i be given by q_i^H, and the marginal change in Hicksian demand for good i with a change in p_0 be given by $\frac{\partial q_i^H}{\partial p_0}$. Then, for any good i, the change in energy use from the energy efficiency improvement in good 0 is $e_i \frac{\partial q_i^H}{\partial p_0} \Delta p_0$, where ei is the energy intensity of good i (i.e. the amount of energy used in providing the energy service). Thus the aggregate change in consumption for all other goods besides good 0 is $\sum_{i \neq 0} e_i \frac{\partial q_i^H}{\partial p_0} \Delta p_0$. Since some goods are substitutes and others are complements, and goods differ in energy intensity, the sign of this term is ambiguous (Chan and Gillingham 2014; Borenstein 2015; Berkhout et al. 2000). In general, one might expect it to be negative, for there is a general substitution in consumption towards good 0 when its efficiency increases.

Fourth, there may be an income effect for every other good and service. If there is additional income freed-up from the energy efficiency improvement, it can be re-spent on other goods and services as the consumer re-optimises consumption. As mentioned above, the change in income associated with the energy efficiency improvement is $q_0 \Delta p_0 - C$. Thus the change in consumption for any good i when good 0 has an energy efficiency improvement is simply the marginal change in demand of good i with a change in income $\left(\frac{\partial q_i}{\partial I}\right)$ times the change in income: $\frac{\partial q_i}{\partial I}(q_0 \Delta p_0 - C)$. The change in energy use for good i is then $e_i \frac{\partial q_i}{\partial I}(q_0 \Delta p_0 - C)$, and the aggregate change in energy use for all goods besides good

0 is $\sum_{i \neq 0} e_i \frac{\partial q_i}{\partial I}(q_0 \Delta p_0 - C)$. The sign of this income rebound effect is ambiguous as well. It depends on the change in income, as well as the relative energy intensity of normal goods versus inferior goods. If nearly all goods are normal goods and the change in income is positive, one would expect a positive sign for this effect.

The sum of these four responses forms the basis of the microeconomic rebound effect, which quantifies the change in energy use with a change in energy efficiency:

$$R = e_0' \frac{\partial q_0^H}{\partial p_0} \Delta p_0 + e_0' \frac{\partial q_0}{\partial I}(q_0 \Delta p_0 - C) + \sum_{i \neq 0} e_i \frac{\partial q_i^H}{\partial p_0} \Delta p_0 \sum_{i \neq 0} e_i \frac{\partial q_i}{\partial I}(q_0 \Delta p_0 - C).$$

The first two terms (substitution and income effects for good 0) are nearly always defined as the **direct rebound effect**, for they capture the direct consumer response in good 0 to the energy efficiency improvement (Sorrell and Dimitropoulos 2008). Assuming a positive change in income and usage of good 0 being a normal good, one would expect a positive sign for the direct rebound effect.

However, the other terms are defined in various ways in the literature, potentially leading to confusion (Turner 2013). In particular, the indirect rebound effect is a term widely used in the literature, yet its usage is inconsistent (Azevedo 2014; Gillingham et al. 2015). Its name indicates the more indirect nature by which energy savings are reduced. Many studies refer to the **indirect rebound** effect as the sum of terms three and four (Chan and Gillingham 2014). Other studies recognise that the indirect rebound effect includes both terms, but focus on only estimating the income effect on other goods and services (the fourth term) as a measure of the rebound effect (Chitnis et al. 2014). Others simply define the indirect rebound as the fourth term (Borenstein 2015). Still others either explicitly or implicitly use a much broader definition for the indirect rebound, which includes both the third and fourth terms as well as additional rebound effects.

One of these additional rebound effects is the **embodied energy rebound effect**, which captures the energy used to create the energy efficiency improvement. The sign of this effect is context-dependent. A more energy-efficient product may take more or less energy to produce. If the process of building a more efficient product is more energy-intensive, then the embodied energy rebound would be expected to be positive. Of course, there may be energy embodied in other goods and services as well, so a broader definition of the embodied energy rebound would also include the change in energy use from embodied energy from other goods and services.

It is common to include the embodied energy rebound as part of the indirect rebound effect. For example, Azevedo (2014) and Thomas and Azevedo (2013b) define the indirect rebound effect as the sum of terms three and four above, and use an energy intensity ei that includes the embodied energy in both good 0 and all other goods and services. Sorrell (2007) defines the total economy-wide rebound effect as the sum of the direct and indirect rebound effects. Under this definition, the indirect rebound effect is a residual that includes the third term, fourth term, all embodied energy effects and macroeconomic rebound effects.

Another proposed definition is to call the first three terms in the equation above the 'net direct rebound effect' for they account for the direct rebound as well as the change in energy use from all other goods and services, including the ones being substituted away from (Borenstein 2015). If this third term is negative, we would expect a smaller net direct rebound than direct rebound effect.

For the net energy savings from an energy efficiency improvement (abstracting from any macroeconomic rebounds), we can compare the microeconomic rebound effect (R) to the upfront energy savings from the efficiency improvement. Thus, the net energy savings after the microeconomic rebound would be given by the energy savings, $q_0(e_0 - e_0')$, minus R and minus any embodied energy effect (E):

$$\text{Net Savings} = q_0(e_0 - e_0') - R - E.$$

R

As mentioned above, the microeconomic rebound effect may also occur for firms. Consider a firm that is using capital and labour to produce a good or service. When there is an energy efficiency improvement, there is a factor-substitution effect: capital becomes relatively more productive, so more (energy-using) capital and less (non-energy-using) labour is included in the optimal production factor mix. Moreover, the marginal cost of production may decline, increasing the optimal amount of production. Thus, both the switch to more energy-using inputs and an increase in production may lead to a rebound effect on the production side (Berkhout et al. 2000). While these production-side rebound effects clearly have microeconomic foundations, nearly all research on them has been at the macroeconomic level, which often aims to take into account the full set of changes in production and prices across the economy.

Macroeconomic Rebound Effects

The macroeconomic, or sometimes ‘economy-wide’, rebound effects involve several channels by which market responses could influence the energy savings from an energy efficiency improvement. There is a **macroeconomic price effect**, or energy market effect, which describes how a shift inwards in demand for energy in the market due to the energy efficiency improvement will be accompanied by a subsequent re-equilibration as prices and quantities are set so that supply equals demand. This market response will mean that the reduction in demand will be less than the amount that demand is shifted inward. It is governed by the slopes of the supply and demand curves. The macroeconomic price effect is small if demand is highly elastic and supply is inelastic, for then the market will re-equilibrate at nearly the same quantity as what you would have without the re-equilibration process. Similarly, the macroeconomic price effect is large if demand is inelastic and supply is elastic (Borenstein 2015; Gillingham et al. 2013).

This macroeconomic price effect can occur in any market, but is particularly easy to understand when there is an energy efficiency improvement shifting demand inward in a single region (e.g. from a fuel economy standard in the USA) and there is a broader market for fuel (e.g. the global oil market). In the case of oil, the reduced demand for oil in the USA leads to a lower global oil price, and thus induced oil demand elsewhere.

Another category of macroeconomic rebound effects can be called the **macroeconomic growth effect**, for it describes how the amount of economic growth and patterns of economic growth can be influenced by the energy efficiency improvement (Gillingham et al. 2013). Jevons was on to a version of this type of rebound effect: a sectoral reallocation rebound effect. Just like the substitution effect in consumption, there is analogous effect in investment and production in the economy. When the relative rate of return of a sector increases, we would expect to see more investment and economic growth in this sector. Of course, this sectoral general equilibrium effect depends on two factors: (1) the degree to which the energy efficiency improvement increases the rate of return of the sector, and (2) the energy intensity of production in the sector relative to other sectors. The sectoral reallocation rebound could be positive or negative, depending on the cost of the energy efficiency improvement (e.g. is it a mandatory and costly energy efficiency increase?) and the energy intensity of the energy-using sector relative to other sectors (e.g. is the shift in production from more energy-intensive or less energy-intensive sectors?). The sectoral reallocation effect can also be extended to a reallocation of innovative activity and human capital, such that higher returns in a sector can lead to more innovative activity and human capital moving into that sector (Lemoine 2014).

The macroeconomic growth effect may also involve innovation in another way. The process of researching to find new ways to improve energy efficiency may engender spillovers to other processes and sectors. For example, finding ultra-lightweight materials for aircraft may spill over to other manufacturing areas, such as that of electronics or bicycles, spurring economic growth in other sectors. Thus, energy efficiency improvements may change the path of innovation in multiple areas, leading to broader economic growth.

Another possible pathway for a macroeconomic growth effect is through a macroeconomic multiplier. Macroeconomists have posited that income gains (usually from a government programme) may have a multiplier effect in times of high unemployment when there is unused capacity in the economy (Ramey 2011). This multiplier effect would occur if a dollar of additional income is spent in a way that uses some of the underutilised labour and capital in the economy. Thus, additional income would generate further income and economic growth more broadly. Of course, this effect may be dampened by any future expected taxes or debt incurred to provide the income. However, in the context of freed-up income from an energy efficiency improvement, the multiplier would not be associated with any additional taxes or government debt, so the effect might be expected to be different (Borenstein 2015).

Other channels may also influence the macroeconomic rebound. For example, Lecca et al. (2014) and Turner (2009) posit an interaction between the macroeconomic price effect and sectoral reallocation, which they call 'disinvestment effects'. In the short run, the shift away from energy can lead to excess capacity in energy supply, leading to lower energy prices and thus a greater rebound. In the longer run, the returns to capital will drop, so this excess capacity will be divested, which will put upward pressure on energy prices, serving to constrain the macroeconomic rebound. Thus, in contrast to previous theoretical predictions (e.g. Wei (2007) and Saunders (2008)), the macroeconomic rebound may be larger in the short run.

Evidence on Rebound Effects: Historical Background

There is an extensive literature aiming to estimate rebound effects in one form or another. Work on the subject ranges from theoretical models with calibrated simulations to empirical estimations and computable general equilibrium models. Yet the magnitude of the total rebound effect varies by context and remains controversial. While the literature provides strong guidance for some microeconomic rebound effects in many contexts, such as the direct rebound effect, it is clear that the relevant magnitude varies by location and setting. The current literature provides less guidance on macroeconomic rebound effects, with different studies capturing different effects, and magnitudes ranging from limited rebound to significant backfire.

The rebound effect first entered into the modern academic literature in 1979 with Brookes (1979) and Khazzoom (1980), who resurrected the Jevons Paradox in the context of modern energy efficiency policies. In fact, the Jevons Paradox has been referred to as the 'Khazzoom–Brookes Postulate' by later studies (Saunders 1992). Khazzoom (1980) was particularly focused on microeconomic rebounds and Brookes (1979) on macroeconomic rebounds, but both posited that improvements in energy efficiency may lead to backfire.

This view was shortly thereafter critiqued in papers such as Lovins (1988), Henly et al. (1988) and Grubb (1990), which pointed out that energy demand is relatively inelastic and energy typically is a small percentage of the cost of energy services, so rebound effects for most energy services might be expected to be small. This led to a series of papers exploring what functional forms on economy-wide production could lead to backfire when there is an energy efficiency improvement. Saunders (1992) assumes a Cobb–Douglas production function, which allows substitutability between inputs, and finds that backfire is not only possible, but may even be likely. In contrast, Howarth (1997) assumes an alternative (Leontief) production function where energy, labour and capital are complements, rather than substitutes, and finds that backfire is not likely. A take-away from this theoretical literature is that if it is easier to substitute across inputs into production, then backfire becomes more likely.

The first empirical estimates used to describe rebound effects were simply estimates of price elasticities of demand for energy services, which are taken as a proxy for the direct rebound effect defined above. Greening et al. (2000) perform a review of the literature estimating price elasticities of demand for a variety of energy services for both consumers and firms. They find price elasticities of demand in the wide range of 0 to −0.5, with most studies falling in the range of −0.1 to −0.3 (estimates included are both long-run and short-run). This would be interpreted as a direct rebound effect of 0–30%. Greening et al. (2000) also coined the terms ‘direct effect’, ‘indirect effect’, ‘economy-wide effect’ and ‘transformational effect’. The transformational effect had a vague definition relating to changing preferences and has not been continued in the subsequent literature. Schipper and Grubb (2000) make perhaps the first rough estimate of the indirect rebound effect, finding that respending leads to a 5–15% rebound. None of the studies estimate the substitution effect on other goods and services described above.

Since 2000, the literature on rebound effects has grown dramatically and reached beyond economics into engineering fields, such as industrial ecology. There are three key strands of current literature. Most studies on rebound effects estimate a price elasticity of demand for an energy service, call this the direct rebound effect, and stop there. But there are a few studies estimating the indirect rebound effect. In addition, there has been recent work using computable general equilibrium models and econometric simulation models aiming to estimate different macroeconomic rebound effects. The next sections discuss each of these three strands of literature in turn.

Evidence on Rebound Effects: Price Elasticities of Demand

The literature estimating price elasticities of demand for energy services is quite large, with perhaps hundreds of papers. Of course, with a literature so vast, current estimates still range widely, depending on the energy service, time frame (short-run or long-run), years covered, location and estimation methodology. Some of the most recent reviews of the literature that focus on the rebound effect include Sorrell (2007), Azevedo (2014), and Gillingham et al. (2015). Each of these papers includes a table reviewing the estimates in the literature. Broadly, estimates for the direct rebound effect still tend to be in the range of 0–50%, just as in the earlier Greening et al. (2000) review. Gillingham et al. (2015) narrows this set further by looking at only more recent studies for a variety of energy services that the authors believe deal with empirical identification issues in a convincing way; these short-run estimates are in a range of 5–40%, with most studies falling in a range of 5–25%.

Notably, most well-identified estimates of the price elasticity of demand are from developed countries, with the most common relating to transport in the USA. Sorrell (2007) and others have suggested that the direct rebound effect in developing countries may be larger, since the demand for energy services may be far from saturated. Indeed, studies from developing countries show an even greater range of estimates, including some very large direct rebound effect estimates (Sorrell 2007). Gillingham et al. (2015) argue that these should be interpreted cautiously and that most of the developing country estimates tend to fall in the same 0–50% range as estimates from developed countries.

The studies on price elasticities of demand for energy services that contribute to the ranges of estimates above tend to use detailed disaggregated data from short time periods. This is useful for understanding the price responsiveness during that time period, but says little about the responsiveness during other time periods. A few studies take a longer-term economic history perspective. For example, Fouquet and Pearson (2012) use historical time series data on lighting in the UK from 1750 to the present and estimate a time-varying price elasticity of lighting demand. The results indicate a price elasticity in the eighteenth and nineteenth centuries that was indicative of backfire. After 1900, the elasticity was closer to zero, but still indicated a substantial responsiveness to price (e.g. in the range of −0.5 to nearly −1).

Fouquet (2012) performs a similar analysis for transport in the UK and also finds a declining responsiveness to energy service price, with the long-run price elasticity of passenger transport demand changing from −1.5 in 1860 to −0.6 in

2010. While these estimates are indicative of more responsiveness than other recent estimates of the price elasticity of transport demand from the USA (e.g. Small and van Dender (2007) and Gillingham (2014)), it is possible that they are consistent; not only is the setting different (e.g. gasoline and diesel prices are higher in the UK), but the time frame of the estimate is different (e.g. long-term versus short-term or medium-term).

Other long-term estimates include Tsao et al. (2010), who find backfire in lighting over several centuries and Saunders (2013), who estimates backfire or large rebound in many sectors over the past four decades. These backfire results are perhaps understood in light of the substantial assumptions involved in the analyses. For example, the estimates in Tsao et al. (2010) are based on the same Cobb–Douglas functional form from earlier work by Saunders (1992), along with many other assumptions. Saunders (2013) relies on a translog cost function, but makes other assumptions that have been critiqued (Gillingham et al. 2015). Another interpretation is that these studies are estimating something different than the rest of the literature, such as a longer-run effect that implicitly includes other rebound effects, such as macroeconomic rebound effects.

Evidence on Rebound Effects: Estimates from Policies

Rather than estimate price elasticities of demand for energy services, which hold all other attributes of the product constant and assume a costless increase in energy efficiency, a few studies relax these assumptions and analyse the rebound effect from a particular policy or treatment. Gillingham et al. (2015) name this type of rebound a 'Policy-induced improvement' (PII). A perfect example is Davis et al. (2015). This study analyses an experiment in Mexico that provides direct cash payments and subsidised financing to consumers replacing old refrigerators and air conditioners. The switch from an old to new appliance is potentially associated with a very large change in attributes, with the new appliances providing a much better energy service. Moreover, there is an income effect from the transfer. The results indicate an extremely large rebound from this policy; for instance, electricity use increases after replacement of the air conditioner and only drops by 7% after replacement of the refrigerator.

Other examples of studies that estimate this type of rebound are Davis (2008) and Gillingham (2013). Davis (2008) examines a field experiment where households received free energy-efficient clothes washers. Subsequently, they increased washing by 5.6%. These clothes washers were not only more energy-efficient, but also were larger and gentler on more clothes, so this rebound effect estimated is the combined effect of the efficiency and the improved energy service. Gillingham (2013) estimates the effect of a policy that incentivises consumers to purchase more efficient new vehicles in California. The results account for the differing attributes of the vehicles being purchased and imply an elasticity of driving with respect to operating costs of -0.15.

Evidence on Rebound Effects: Other Goods and Services

Given the inconsistent definition of the indirect rebound effect, it can be difficult to compare across studies. Many studies focus on only the income effects on other goods and services, which is more straightforward. Other studies aim to include at least a bound on the substitution effects on other goods and services.

To estimate the income effects on other goods and services, there are a few common approaches. Many of the studies cross over into the industrial ecology literature and rely on input–output analysis. It is also common for the studies in this literature to estimate rebound effects in terms of carbon dioxide or greenhouse gas emissions in addition to or instead of energy.

One approach is to assume proportional re-spending, so that any income available to be re-spent would be spent according to average spending patterns throughout the economy (e.g. see Lenzen and Dey (2002) and Thomas and Azevedo (2013a)). Thus the average energy intensity of economic activity is used to determine the reduction in energy savings due to the income effects on other goods and services. A concern with this methodology is that the spending of an additional marginal dollar may be very different than the average spending overall.

A second approach aims to understand the energy implications of a marginal dollar of spending by comparing the spending patterns of consumers in different income brackets (Thiesen et al. 2008). The underlying assumption of this approach is that as any consumer becomes wealthier, they will begin to emulate consumers in higher income brackets. This is effectively using cross-sectional variation in income to estimate income elasticities. On the margin, this may be plausible, especially if the income brackets are fine enough and we are comparing consumers in the same location. The methodology is more questionable with coarse brackets or comparing consumers across a broad region. A third approach is to use income elasticities across a broad range of sectors estimated by other studies. Druckman et al. (2011), Chitnis et al. (2014) and (2013) use this approach in the UK.

Some studies use a combination of methods and also attempt to estimate both the substitution and income effects on other goods and services. Thomas and Azevedo (2013a) make several alternative assumptions using both income elasticities and the assumption of proportional re-spending. In addition, Thomas and Azevedo (2013a) aim to bound the substitution effects on other goods and services by using existing cross-price elasticity estimates.

Brannlund et al. (2007) and Mizobuchi (2008) take a different approach than all of the above studies and estimate a system of household demand equations to provide results on both the cross-price and income elasticities. This provides estimates of both the substitution and income effects on other goods and services. The results of these studies diverge. Lenzen and Dey (2002) find very large estimates of the indirect rebound that lead to backfire when combined with the direct rebound for Australia. Thomas and Azevedo (2013a) and Druckman et al. (2011) are more recent and comprehensive studies that suggest estimates on the order of 5–15% for the USA and the UK respectively. The results in Brannlund et al. (2007) suggest backfire in Sweden, while the results in Mizobuchi (2008) account for capital costs and suggest much smaller rebound effect. Of course, these last two studies cannot be directly compared to the others, given their very different methodology.

Evidence on Rebound Effects: Macroeconomic

There are only a few studies aiming to estimate macroeconomic rebound effects, but it is an area of rapid growth. We begin with the macroeconomic price effect. While there is no question that the macroeconomic price effect exists, its magnitude depends on the slope of supply and demand curves in the energy market of interest. Borenstein (2015) performs a useful sensitivity analysis for the global market to emphasise how the effect may be quite significant depending on assumptions. Even with a relatively inelastic oil demand elasticity of −0.4 and an elastic oil supply elasticity of 1.0, the macroeconomic price effect is on the order of 30%.

Recent studies estimating macroeconomic growth effects tend to use computable general equilibrium models and econometric simulation models. These models include a variety of different channels, depending on the model. They generally include both microeconomic and macroeconomic rebound effects and model energy efficiency improvements as energy-augmenting technical change that has no impact on other factor inputs (Sorrell and Dimitropoulos 2007). The range in results is wide: from negative rebound to backfire (Turner 2009, 2013; Broberg et al. 2014).

Some notable recent estimates using econometric simulation models show macroeconomic rebound effects of 11% (Barker et al. 2007) and 21% (Barker et al. 2009). The estimates from these two papers include substitution effects on all other goods and services, but do not include the direct rebound effect (the direct rebound is treated separately). Other general equilibrium studies, such as Lecca et al. (2014), include all of the effects together, such that disentangling the different effects is not possible. More broadly, the current modelling efforts include a variety of channels, but do not tend to include macroeconomic rebounds from the macroeconomic multiplier or induced innovation, leaving these as open research topics.

Conclusions and Policy Implications

There is no debate that rebound effects can occur and are important to consider in analysis of energy efficiency policies. Rebound effects can reduce the energy and emissions savings from an energy efficiency improvement, which would reduce the energy savings of the policy. Moreover, rebound effects may have external costs associated with them, such as air pollution and carbon dioxide emissions, reducing the net benefits of the policy. Thus, rebound effects can increase the welfare difference between first-best policies, such as direct pricing of external costs, and second-best energy efficiency policies (Gillingham et al. 2015).

With such policy importance, a debate continues in the academic literature and policy communities about the magnitude of such an effect and whether anything should be done about it (Gillingham et al. 2013). A review of the literature reveals wide disparity in plausible magnitudes, depending on the context, location and time frame of the rebound effects, as well as which of the rebound effects are quantified in the study. From a theory perspective, neither negative rebound nor backfire can be ruled out.

Fortunately, there is empirical evidence that can provide some guidance. Price elasticities of energy service demand suggest that for many energy services and contexts, the direct rebound effect is in the range of 5–60% and may be at the lower end for some important contexts where the empirical evidence is the strongest. Passenger transport in the USA is a notable example. Estimates of the indirect rebounds from the income and substitution effects on other goods and services vary as widely as the definition of the indirect rebound effect. Recent studies suggest estimates in the range of 5–15% in developed countries. Unfortunately, the evidence on both the direct and indirect rebound effects in developing countries is much weaker than in developed countries. Some recent studies are beginning to examine the rebound effects from a policy in both developed and developing countries, with varying findings depending on the context. Macroeconomic rebound effects are the least well understood, and current estimates contain a variety of different channels, leading to a variety of different results ranging from backfire to negative rebound.

Stepping back, it is important to recognise that unless there are large external costs associated with rebound effects, they are generally social welfare improving, for they come about from the choice to use more of a valued energy service or from induced innovation (Chan and Gillingham 2014; Gillingham et al. 2015). However, while they may be beneficial for social welfare, their existence may still reduce the energy savings from energy efficiency policy, tilting us further towards first-best policies to address externalities.

See Also

- ▶ Climate Change, Economics of
- ▶ Energy Economics
- ▶ Exhaustible Resources

Acknowledgments The author would like to acknowledge Roger Fouquet, Karen Turner and an anonymous reviewer for very helpful comments on this draft and Matthew Kotchen, David Rapson and Gernot Wagner for many useful conversations on this topic.

R

Bibliography

Azevedo, I.L. 2014. Consumer end-use energy efficiency and rebound effects. *Annual Review of Environment and Resources* 39: 393–418.

Barker, T., P. Ekins, and T. Foxon. 2007. The macroeconomic rebound effect and the UK economy. *Energy Policy* 35: 4935–4946.

Barker, T., A. Dagoumas, and J. Rubin. 2009. The macroeconomic rebound effect and the world economy. *Energy Efficiency* 2: 411–427.

Berkhout, P., J. Muskens, and J. Velthuijsen. 2000. Defining the rebound effect. *Energy Policy* 28: 425–432.

Borenstein, S. 2015. A microeconomic framework for evaluating energy efficiency rebound and some implications. *Energy Journal* 36: 1–21.

Brannlund, R., T. Ghalwash, and J. Nordstrom. 2007. Increased energy efficiency and the rebound effect: Effects on consumption and emissions. *Energy Economics* 29: 1–17.

Broberg, T., C. Berg, and E. Samakovlis. 2014. The economy-wide rebound effect from improved energy efficiency in Swedish industries. Umea University CERE Working Paper. Umea, Sweden.

Brookes, L. 1979. A low energy strategy for the UK by G. Leach et al., a review and reply. *Atom* 269: 3–8.

Chan, N., and K. Gillingham. 2014. The microeconomic theory of the rebound effect and its welfare implications. *Journal of the Association of Environmental and Resource Economists* 2: 133–159.

Chitnis, M., S. Sorrell, A. Druckman, S. Firth, and T. Jackson. 2013. Turning lights into flights: Estimating direct and indirect rebound effects for UK households. *Energy Policy* 55: 234–250.

Chitnis, M., S. Sorrell, A. Druckman, S. Firth, and T. Jackson. 2014. Who rebounds most? Estimating direct and indirect rebound effects for different UK socioeconomic groups. *Ecological Economics* 106: 12–32.

Davis, L. 2008. Durable goods and residential demand for energy and water: Evidence from a field trial. *RAND Journal of Economics* 39: 530–546.

Davis, L., A. Fuchs, and P. Gertler. 2015. Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy* 6: 207–238.

Druckman, A., M. Chitnis, S. Sorrell, and T. Jackson. 2011. Missing carbon reductions? Exploring rebound and backfire effects in UK households. *Energy Policy* 39: 3575–3581.

Fouquet, R. 2012. Trends in income and price elasticities of transport demand (1850–2010). *Energy Policy* 50: 62–71.

Fouquet, R., and P. Pearson. 2012. The long run demand for lighting: Elasticities and rebound effects in different phases of economic development. *Economics of Energy and Environmental Policy* 1: 83–100.

Gillingham, K. 2013. *Selection on anticipated driving and the consumer response to changing gasoline prices*. Yale University Working Paper.

Gillingham, K. 2014. Identifying the elasticity of driving: Evidence from a gasoline price shock. *Regional Science & Urban Economics* 47: 13–24.

Gillingham, K., and K. Palmer. 2014. Bridging the energy efficiency gap: Policy insights from economic theory and empirical analysis. *Review of Environmental Economics & Policy* 8: 18–38.

Gillingham, K., M. Kotchen, D. Rapson, and G. Wagner. 2013. The rebound effect is over-played. *Nature* 493: 475–476.

Gillingham, K., D. Rapson, and G. Wagner. 2015. The rebound effect and energy efficiency policy. *Review of Environmental Economics & Policy* 10: 68–88.

Greening, L.A., D.L. Greene, and C. Difiglio. 2000. Energy efficiency and consumption – The rebound effect – A survey. *Energy Policy* 28: 389–401.

Grubb, M. 1990. Communication: Energy efficiency and economic fallacies. *Energy Policy* 18: 783–785.

Henly, J., M. Ruderman, and M. Levine. 1988. Energy saving resulting from the adoption of more efficient appliances: A follow-up. *Energy Journal* 9: 163–170.

Howarth, R. 1997. Energy efficiency and economic growth. *Contemporary Economic Policy* 15: 1–9.

Jevons, W.S. 1865. *The coal question; an enquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*. London: Macmillan.

Khazzoom, J.D. 1980. The economic implications of mandated efficiency in standards for household appliances. *Energy Journal* 1: 21–40.

Lecca, P., P. McGregor, J.K. Swales, and K. Turner. 2014. The added value from a general equilibrium analysis of increased efficiency in household energy use. *Ecological Economics* 100: 51–62.

Lemoine, D. 2014. *Long-run backfire from energy policies*. University of Arizona Working Paper.

Lenzen, M., and C. Dey. 2002. Economic, energy and greenhouse gas emissions impacts of some consumer choice, technology and government outlay options. *Energy Economics* 24: 377–403.

Lovins, A. 1988. Energy saving resulting from the adoption of more efficient appliances: Another view. *Energy Journal* 9: 155–162.

Mizobuchi, K. 2008. An empirical study on the rebound effect considering capital costs. *Energy Economics* 30: 2486–2516.

Ramey, V. 2011. Can government purchases stimulate the economy? *Journal of Economic Literature* 49: 673–685.

Saunders, H. 1992. The Khazzoom–Brookes Postulate and neoclassical growth. *Energy Journal* 13: 131–148.

Saunders, H. 2008. Fuel conserving (and using) production functions. *Energy Economics* 30: 2184–2235.

Saunders, H. 2013. Historical evidence for energy consumption rebound in 30 US sectors and a toolkit for rebound analysts. *Technological Forecasting and Social Change* 80: 1317–1330.

Schipper, L., and M. Grubb. 2000. On the rebound? Feedback between energy intensities and energy use in IEA countries. *Energy Policy* 28: 367–388.

Small, K.A., and K. Van Dender. 2007. Fuel efficiency and motor vehicle travel: The declining rebound effect. *Energy Journal* 28: 25–51.

Sorrell, S. 2007. *The rebound effect: An assessment of the evidence for economy-wide energy savings from improved energy efficiency*. Sussex: UK Energy Research Center Report.

Sorrell, S. and J. Dimitropoulos. 2007. *UKERC review of evidence for the rebound effect: Technical report 5: Energy, productivity and economic growth studies*. University of Sussex Working Paper.

Sorrell, S., and J. Dimitropoulos. 2008. The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics* 65: 636–649.

Thiesen, J., T. Christensen, T. Kristensen, R. Andersen, B. Brunoe, T. Gregersen, M. Thrane, and B. Weidema. 2008. Rebound effects of price differences. *International Journal of Life Cycle Assessment* 13: 104–114.

Thomas, B.A., and I.L. Azevedo. 2013a. Estimating direct and indirect rebound effects for U.S. households with input–output analysis. Part 2: Simulation. *Ecological Economics* 86: 188–198.

Thomas, B.A., and I.L. Azevedo. 2013b. Estimating direct and indirect rebound effects for US households with input–output analysis. Part 1: Theoretical framework. *Ecological Economics* 86: 199–210.

Tsao, J., H. Saunders, J.R. Creighton, M. Coltrin, and J. Simmons. 2010. Solidstate lighting: An energy-economics perspective. *Journal of Physics D: Applied Physics* 43: 1–17.

Turner, K. 2009. Negative rebound and disinvestment effects in response to an improvement in the UK economy. *Energy Economics* 31: 648–666.

Turner, K. 2013. 'Rebound' effects from increased energy efficiency: A time to pause and reflect. *Energy Journal* 34: 25–42.

Wei, T. 2007. Impact of energy efficiency gains on output and energy use with Cobb–Douglas production function. *Energy Policy* 35: 2023–2030.

Reciprocity and Collective Action

Rajiv Sethi

Abstract

A collective action problem arises when the private incentives faced by individual members of a group are not properly aligned with their shared goals. Such problems can be overcome if opportunistic behaviour is restrained by explicit sanctions or internalized social norms. In particular, collective action is facilitated by norms of reciprocity that induce individuals to undertake pro-social actions whenever they expect others to do the same. From this perspective, collective action requires coordinated expectations and effective communication. Experimental evidence suggests that reciprocity norms are widespread in human populations, and evolutionary mechanisms that can account for their prevalence have been identified.

Keywords

Assortative matching; Cheap talk; Collective action; Cooperation; Coordination problems and communication; Free-rider problem; Public goods; Reciprocity; Social norms; Subgame perfection; Tragedy of the commons

JEL Classifications

C9

Advancing the common interest of a group sometimes requires its members to sacrifice their private interests. Such situations, in which individual incentives are not properly aligned with shared goals, are called collective action problems. They arise frequently in economic and social life, for instance in the context of political mobilization, electoral turnout, pollution abatement, common property management and the provision of public goods. They can involve relatively small groups such as families, teams, or business partnerships, or very large groups that cut across national boundaries.

In his classic work on collective action, Mancur Olson (1965) conjectured that individuals would be unable to overcome such problems unless their behaviour was constrained by rules that were externally imposed and enforced. Along similar lines, Garret Hardin (1968) argued in an influential paper that, left to their own devices, individuals would face a 'tragedy of the commons' which could be overcome only by 'mutual coercion, mutually agreed upon'. This view continues to have considerable currency in economics in the form of the *free-rider hypothesis,* which maintains that voluntary contributions that are socially beneficial but privately costly will not generally be observed (Bergstrom et al. 1986).

Despite the compelling logic underlying the free-rider hypothesis, there are numerous instances of groups having overcome collective action problems without external pressure, sometimes by designing and abiding by their own set of rules, and sometimes on the basis of less formal arrangements codified in social norms. The success of OPEC in constraining production to maintain price levels is based on a mutually beneficial agreement among member countries that has been sustained despite strong incentives for some producers to free-ride on the restraint practised by others. On a smaller scale, many examples of successful collective action in the management of local fisheries, forests, and other renewable resources have been documented (Bromley 1992; Ostrom 1990). Such resources are often held as common property, and the maintenance of sustainable stocks requires restraint in individual extraction levels. Restraint is typically enforced by formal or informal sanctions, and participation in such punishment mechanisms is itself a form of collective action. There also exist examples of collective action in the absence of any sanctioning mechanism. For instance, voter turnout is often substantial in large elections, contrary to the predictions of the free-rider hypothesis.

It has been argued that many instances of successful collective action arise in small and stable groups whose members interact with each other repeatedly. Under such circumstances, pro-social behaviour can be fully consistent with the standard economic hypotheses of rationality and self-interest. When interactions are repeated, self-interested cooperation can arise if one believes that non-cooperative actions will be punished in future periods. Moreover, such threats of punishment can be credible if abstaining from punishment is itself punished. Formally, cooperative behaviour can be sustained in subgame perfect equilibrium if interactions are infinitely (or indefinitely) repeated (Fudenberg and Maskin 1986). Hence the tension between individual and common interest is less severe and collective action more likely to arise in small and stable groups.

While the threat of future punishment or the promise of future reward might motivate collective action in some instances, there are many situations in which individual actions are unobservable or repetition too infrequent for such considerations to be decisive. Voter turnout, for instance, or private donations to charity are not easily explained as self-interested responses to material incentives. Similarly, sacrifices involving risks to life and limb, as in the case of battlefield heroism or spontaneous collective violence, are unlikely to be driven by a calculated response to future costs and benefits. What, then, could account for such phenomena?

There is now a considerable body of experimental evidence to suggest that many individuals are willing to take actions that further the common interest provided that they are reasonably sure that other group members will also take such actions. Furthermore, they are willing to sanction the opportunistic behaviour of others even at some cost to themselves (Fehr and Gächter 2000). The widespread prevalence of such preferences for reciprocity suggests that collective action can sometimes be viewed as a coordination problem: if the members of a group confidently expect others to further the common good, such expectations can be self-fulfilling. On the other hand, expectations of widespread free-riding can also be self-fulfilling, so building confidence in the behaviour of others is a critical ingredient of successful collective action. Communication among group members can help coordinate expectations, and it is therefore not surprising that allowing for communication among experimental subjects can result in dramatically increased levels of cooperation. This is the case even if communication takes the form of 'cheap talk', with neither threats nor promises being enforceable (Ostrom et al. 1992).

If preferences for reciprocity are indeed part of the explanation for successful collective action, this raises the question of how such preferences have come to be widespread in human populations in the first place. The existence of a willingness to sacrifice one's own material interest for the common good poses an evolutionary puzzle. In order to survive and spread in human populations, the possession of such preferences must confer on an individual some advantage relative to those who are entirely self-interested. One intriguing possibility is that, despite being disadvantageous to individuals within groups, traits that are advantageous for the group itself may survive because of competition among groups:

> There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to give aid to each

> other and to sacrifice themselves for the common good, would be victorious over other tribes; and this would be natural selection. (Darwin 1871, p. 166)

In order to be effective, however, this mechanism requires variability across groups to be sustained while variability within groups is suppressed (Sober and Wilson 1998). Whether or not the conditions for this are empirically plausible remains an open question.

There exist other channels through which a preference for reciprocity can be materially advantageous to individuals. One is assortative interaction: if individuals with preferences for reciprocity are more likely to interact with each other than with opportunists, the former can end up with higher material payoffs than the latter. Such assortation arises naturally in structured populations with local interaction. Even in unstructured populations with random matching, a propensity to reciprocate or to sanction opportunistic behaviour can confer an advantage provided that such preferences are observable to others. The visible possession of such propensities can alter the behaviour of those with whom one is interacting in such a manner as to be materially rewarding. Even opportunistic individuals might be induced to behave cooperatively in interactions with those who have a credible reputation for reciprocity. Such considerations can provide the basis for an evolutionary theory of reciprocity (Sethi and Somanathan 2001).

Reciprocity is a key feature of successful collective action, both in repeated interactions and in more spontaneous settings. The willingness to further the common good even at considerable personal cost is widespread in human populations, but is often contingent on the willingness of others to do the same. This perspective suggests that collective action problems are not insurmountable, but that communication and coordination are critical in overcoming them.

See Also

- ▶ Collective Action
- ▶ Common Property Resources
- ▶ Cooperation
- ▶ Coordination Problems and Communication
- ▶ Public Goods
- ▶ Social Norms
- ▶ Social Preferences

Bibliography

Bergstrom, T.C., L. Blume, and H. Varian. 1986. On the private provision of public goods. *Journal of Public Economics* 29: 25–49.

Bromley, D.W. 1992. *Making the commons work: Theory, practice and policy*. San Francisco: Institute for Contemporary Studies.

Darwin, C. 1871. *The descent of man and selection in relation to sex*. London: Murray.

Fehr, E., and S. Gächter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90: 980–994.

Fudenberg, D., and E. Maskin. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54: 533–554.

Hardin, G. 1968. The tragedy of the commons. *Science* 162: 1243–1248.

Olson, M. 1965. *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ostrom, E. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.

Ostrom, E., J. Walker, and R. Gardner. 1992. Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86: 404–417.

Sethi, R., and E. Somanathan. 2001. Preference evolution and reciprocity. *Journal of Economic Theory* 97: 273–297.

Sober, E., and D.S. Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

R

Recreation

Jack L. Knetsch

Problems of market failure and preferences for a wider distribution of recreational opportunities than might be provided by private market exchanges, have led to a large measure of public provision of parks and other recreational facilities in countries throughout the world. This nonmarket nature of recreation allocation has prompted most of the attention to recreation economics.

Expenditures and Standards

The earliest interest in the economics of recreation centred on the impacts generated by expenditures on equipment, accommodation and travel associated with participation in recreational activities and the use of parks. Such expenditures have been seen to benefit local commercial interests and have been widely asserted to be appropriate measure of the economic value of recreation facilities and activities as well.

Consistent with the continuing interest in the impacts of such expenditure, most economic studies have been concerned with rural or resource-based recreation. Until recently, little attention was given to recreation taking place within urban areas, in spite of the far larger numbers of people and likely larger economic values involved. Most of the planning in cities and towns has been based on various standards of desired numbers of parks or other recreation facilities or the area of open space per unit of resident population, with little note of economic evaluations (Butler 1959).

Demand and Supply

The more contemporary interest in the economics of recreation began in the late 1950s with an awareness of the rapidly increasing participation in recreation activities by large segments of populations, and a growing recognition of the appropriate economic claim of recreational use on scarce resources. The early writings of Clawson (e.g. 1959) and the report of the Outdoor Recreation Resources Review Commission (1962) in the US did much to focus attention on recreation in terms of demand and supply and to demonstrate the important impacts of the locations of recreation areas on people's participation in recreational activities. While the conclusions were initially based on US data, other studies confirmed similar patterns and problems of locational imbalances in other countries.

Largely because of planning and project justification requirements, much of the continuing interest in recreation demand has centred on the use and value of specific individual recreation sites rather than on more aggregate demand parameters. While many early studies considered individual sites in isolation, some dealt more formally with several determinants of site demand and with multiple sites (for example, Burt and Brewer 1971). These studies have provided further insights into the demand for recreation, but their usefulness has been limited by their small number, the lack of much attention to factors other than population proximity, the poor specification of site characteristics, and the little attention given to such things as crosselasticities between sites and types of facilities.

Useful progress on characterizing supply has been even more modest than on specifying demand. The lack of uniformity among recreation sites and facilities and the importance of location have proved to be severe impediments to incorporating supply into meaningful economic analyses. Much of the work has concentrated on classification of landscapes for potential recreation purposes and inventories of recreation resources.

In some studies, participation rates, which are a function of both the demand for facilities and their availability or supply, have often been taken to be due to demand alone – a difficulty somewhat akin to the identification problem. This confusion of use with demand has led to provision of more of what had already been supplied, and has added to spatial imbalances as observed use was taken as demand for similar facilities in the same areas.

Valuation

The increased use of benefit cost analysis, particularly for water resource projects in the US, was largely responsible for the development of evaluation methods. An early, if incomplete, response to the requirement for explicit valuations was the use of an agreed-upon schedule of a specific value per site visit. This practice continues, but is widely regarded as largely arbitrary and to allow little discrimination in values to reflect relative scarcities of different resource attractions and facilities.

By far the most widely discussed improvement for valuing recreation is the so-called travel cost technique suggested by Clawson (1959) (a related technique was proposed by Hotelling 1947). In this method, a demand curve is estimated from the observed relationship between travel costs and visit rates by calculating the decrease in the number of people that would be expected to visit from each origin as a result of varied increments of increased cost. The area under the resulting demand curve is normally taken as the measure of value.

The technique yields a meaningful economic measure of the recreation benefits provided by a site, but estimation problems remain in spite of many suggested improvements (for example, Mansfield 1971; Smith 1971; Vickerman 1974). Multiple destination trips provide further problems as does the poor ability to take account of the value of travel time and recreation occurring over broad areas rather than at single sites.

An alternative evaluation technique is the contingent valuation method, which derives values based on survey respondents' expressed preference between paying to have access to a recreation facility and not having it (Davis 1964). Alternative question formats, such as openand closed-ended questions and sequential bids, are used to overcome problems of bias and the reluctance of respondents to nominate a price.

The contingent valuation technique can be used in a much wider range of circumstances of non-pecuniary evaluations than can other techniques. For example, it can be used to evaluate recreation not associated with specific sites, and to assess the gain or loss of general environmental amenities. However, tests for biases and the effect of varying format remain incomplete.

Contingent valuation studies provided early evidence, confirmed later by experiments and studies involving real money exchanges, of an unexpected large disparity between alternative economic measures of a loss (Knetsch and Sinden 1984). Contrary to the conventional assumption of near equivalence, differing only by a usually trivial amount due to an income or wealth effect, the empirical evidence from a wide range of studies indicates that the compensation demanded to give up an entitlement is very often far larger than the willingness to pay to keep the same entitlement. These findings suggest that many losses may be seriously understated when the willingness-to-pay measure is used instead of the more appropriate compensation-demanded measure. The inconsistency between the evidence and the traditional expectations remains unresolved, although data from a large body of psychological tests suggesting that people typically make evaluations on the basis of changes from a neutral reference point and value losses more than gains, offer at least a partial explanation (Kahneman and Tversky 1979).

Impact of Economic Studies

The direct impact of the considerable attention given to the development of techniques of economic analyses of recreation in terms of adoption and widespread use, has been very modest. Little direct use is yet being routinely made of demand and supply models or of the evaluation techniques. Evidence from economic studies is sometimes used by advocacy groups in support of particular interests, but such decisions typically still turn more on expenditure and employment claims than on more appropriate economic analyses.

Indirectly, the impact of recreation economics studies has no doubt been considerable and beneficial (Pigram 1983). While seldom used to resolve specific issues, economic analyses have clearly demonstrated that recreation facilities may well have large economic value in many cases, and that recreation demands can have a commensurate economic claim on resources (Clawson and Knetsch 1966).

See Also

► Environmental Economics

Bibliography

Burt, O.R., and D. Brewer. 1971. Evaluation of net social benefits from outdoor recreation. *Econometrica* 39: 813–827.

Butler, G.D. 1959. *Introduction to community recreation*, 3rd ed. New York: McGraw-Hill.

Clawson, M. 1959. *Methods of measuring the demand for and value of outdoor recreation*, Reprint No. 10. Washington, DC: Resources for the Future.

Clawson, M., and J.L. Knetsch. 1966. *Economics of outdoor recreation*. Baltimore: Johns Hopkins Press.

Davis, R.K. 1964. The value of big game hunting in a private forest. In *Transactions of the 29th North America Wildlife and Natural Resources Conference*, 393–402. Washington, DC: Wildlife Management Institute.

Hotelling, H. 1947. Letter. In *The economics of public recreation: An economic study of the monetary evaluation of recreation in the national parks*, ed. R.A. Prewitt. Washington, DC: US National Park Service.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.

Knetsch, J.L., and J.A. Sinden. 1984. Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *Quarterly Journal of Economics* 99: 507–521.

Mansfield, N.W. 1971. The estimation of benefits from recreation sites and the provision of a new recreation facility. *Regional Studies* 5: 55–69.

Outdoor Recreation Resources Review Commission. 1962. *Outdoor recreation for America*. Washington, DC: Government Printing Office.

Pigram, J.J. 1983. *Outdoor recreation and resource management*. London: Macmillan.

Smith, R.J. 1971. The evaluation of recreation benefits: The Clawson method in practice. *Urban Studies* 8: 89–102.

Vickerman, R.W. 1974. The evaluation of benefits from recreation projects. *Urban Studies* 11: 277–288.

Recursive Competitive Equilibrium

Rajnish Mehra

Abstract

In this article we define a recursive competitive equilibrium, provide an example and review the related literature.

Keywords

Arrow–Debreu model of general equilibrium; Bellman's principle of optimality; Business cycles; Dynamic decentralization; Markov processes; Neoclassical stochastic growth model; Optimality; Recursive competitive equilibrium; Time-invariant equilibrium decision rules; Total factor productivity; Valuation equilibrium

JEL Classifications

D4; D10

The underlying structure of most dynamic business-cycle and consumption-based asset-pricing models is a variant of the neoclassical stochastic growth model. Such models have been analysed by, among others, Cass (1965), Brock and Mirman (1972), and Donaldson and Mehra (1983). They focus on how an omniscient central planner seeking to maximize the present value of expected utility of a representative agent optimally allocates resources over the infinite time horizon.

Production is limited by an aggregate production function subject to technological (total factor productivity) shocks. The solution to the planning problem is characterized by time-invariant decision rules, which determine optimal consumption and investment each period. These decision rules have as arguments the economy's period aggregate capital stock and the shock to technology.

Business cycles, however, are not predicated on the actions of a central planner, but arise from interactions among economic agents in competitive markets. Given the desirable features of the stochastic growth paradigm – the solution methods are well known and the model generates well-defined proxies for all the major macro aggregates: consumption, investment, output, and so on – it is natural to ask if the allocations arising in that model can be viewed as competitive equilibria. That is, do price sequences exist such that economic agents, optimizing at these prices and interacting through competitive markets, achieve the allocations in question as competitive equilibria? This is the essential question of dynamic-decentralization theory.

Alternative Approaches to Dynamic Decentralization: Valuation Equilibrium

One way of modelling uncertain dynamic economic phenomena is to use Arrow–Debreu general equilibrium structures and to search for optimal actions conditional on the sequence of realizations of all past and present random variables or shocks. The commodities traded are contingent claim contracts. These contracts deliver goods (for example, consumption and capital goods) at a future date, contingent on a particular sequential realization of uncertainty. Markets are assumed to be complete, so that, for any possible future realization of uncertainty (sequence of technology shocks) up to and including some future period, a market exists for contracts that will deliver each good at that date contingent on that realization (event). This requires a very rich set of markets. All trading occurs in the first period: consumers contract to receive consumption and investment goods and to deliver capital goods in all future periods contingent on future states so as to maximize the expected present value of their utility of consumption over their infinite lifetimes. Firms choose their production plans so as to maximize the present value of discounted profits. Given current prices, they contract to deliver consumption and investment goods to, and to receive capital goods from, the consumer-investors. Under standard preference structures, these contingent choices never need to be revised. That is, if markets reopen, no new trades will occur.

In its most general formulation, a valuation equilibrium is characterized simply as a continuous linear functional that assigns a value to each bundle of contingent commodities. Only under more restrictive assumptions can this function be represented as a price sequence (Bewley 1972; Prescott and Lucas 1972; Mehra 1988). The basic result is that for any solution to the planner's problem – that is, sequences of consumption, investment and capital goods – a set of state-contingent prices exists such that these sequences coincide with the contracted quantities in the valuation equilibrium.

This decentralization concept is quite broad and applies to central-planning formulations much more general than the neoclassical growth paradigm. It reminds us that the financial structure underlying the stochastic growth paradigm is fundamentally one of complete contingent commodity markets. Nevertheless, it is a somewhat unnatural perspective for macroeconomists (all macro policies must be announced at time zero), and it presumes a set of markets much richer than any observed. These shortcomings led to the development of the concept of a recursive competitive equilibrium.

Recursive Competitive Theory

An alternative approach that has proved very useful in developing testable theories is to replace the attempt to locate equilibrium sequences of contingent functions with the search for time-invariant equilibrium decision rules. These decision rules specify current actions as a function of a limited number of 'state variables' which fully summarize the effects of past decisions and current information. Knowledge of these state variables provides the economic agents with a full description of the economy's current state. Their actions, together with the realization of the exogenous uncertainty, determines the values of the state variables in the next sequential time period. This is what is meant by a recursive structure. In order to apply standard time-series methods to any testable implications, these equilibrium decision rules must be time-invariant.

Recursive competitive theory was first developed by Mehra and Prescott (1977) and further refined in Prescott and Mehra (1980). These papers also establish the existence of a recursive competitive equilibrium and the supportability of the Pareto optimal through the recursive price functions. Excellent textbook treatments are contained in Harris (1987), Stokey et al. (1989) and Ljungqvist and Sargent (2004). Since its introduction, it has been widely used in exploring a wide variety of economic issues including business-cycle fluctuations, monetary and fiscal policy, trade-related phenomena, and regularities in asset price co-movements. (See, for example, Kydland and Prescott 1982; Long and Plosser 1983; Mehra and Prescott 1985.)

The recursive equilibrium abstraction postulates a continuum of identical economic agents indexed on the unit interval (again with

R

preferences identical to those of the representative agent in the planning formulation), and a finite number of firms. As in the valuation equilibrium approach, consumers undertake all consumption and saving decisions. Firms, which have equal access to a single constant-returns-to-scale technology, maximize their profits each period, and are assumed to produce two goods: a consumption good and a capital good. Unlike in the valuation equilibrium approach, trading between agents and firms occurs every period. (This is in contrast to markets in an Arrow–Debreu setting where, as mentioned earlier, no trade would occur if markets were to reopen.) At the start of each period, firms observe the technological shock to productivity and purchase capital and labour services, which are supplied inelastically at competitive prices. The capital and labour are used to produce the capital and consumption goods. At the close of the period, individuals, acting competitively, use their wages and the proceeds from the sale of capital to buy the consumption and capital goods produced by the firms. Consumers then retain the capital good into the next period, when it again becomes available to firms and the process repeats itself. Note that firms are liquidated at the end of each period (retaining no capital assets while technology is freely available), and that no trades between firms and consumer-investors extend over more than one time period. Capital goods carried over from one period to the next are the only link between periods, and period prices depend only on the state variables in that period.

Formally, a *recursive competitive equilibrium* (RCE) is characterized by *time invariant* functions of a limited number of 'state variables', which summarize the effects of past decisions and current information. These functions (decision rules) include (*a*) a pricing function, (*b*) a value function, (*c*) a period allocation policy specifying the individual's decision, (*d*) a period allocation policy specifying the decision of each firm and (*e*) a function specifying the law of motion of the capital stock.

While the restrictive structure of markets and trades makes this concept less general than the valuation equilibrium approach, it provides an interpretation of decentralization that is better suited to macro-analysis. More recently, the recursive equilibrium concept has been generalized to admit an infinitely lived firm which maximizes its value. When an RCE is Pareto optimal, its allocation coincides with that of the associated planning problem. The solution to the central-planning stochastic-growth problem may then be regarded as the aggregate investment and consumption functions that would arise from a decentralized, recursive homogeneous consumer economy. We illustrate this with the help of an example below, which considers an economy with a single capital good. The reader is referred to Prescott and Mehra (1980) for the more general case with multiple capital types.

An Example

Consider the simplest central planning stochastic growth paradigm

$$w(k_0, \lambda_0) = \max\ E\left\{\sum_{t=0}^{\infty} \beta^t u(c_t)\right\} \qquad \text{(P1)}$$

subject to

$$c_t + k_{t+1} \leq \lambda_t f(k_t, l_t), \quad \lambda_0, k_0 \quad \text{given}, \quad l_t = 1\ \forall t.$$

In this formulation, u(·) is the period utility function of a representative consumer defined over his period t consumption c_t; k_t denotes capital available for production in period t and l_t denotes period t labour supply which is inelastically supplied by the consumer-investor at $l_t = 1$, for all t. The expression $f(k_t, l_t)$ represents the period technology (production function) which is shocked by the bounded stationary stochastic factor λ_t. (It is assumed that λ_t is subject to a stationary Markov process with a bounded ergodic set.) E denotes the expectations operator and the central planner is assumed to have rational expectations; that is, he uses all available information to rationally anticipate future variables. In particular he knows the conditional distribution of future technology shocks $F(\lambda_{t+1}; \lambda_t)$. For the purposes of this example we restrict preferences to be logarithmic and assume a Cobb–Douglas technology (to the best of my knowledge, this parameterization is the simplest example known to result in closed form solutions): $u(c_t) = \ln c_t$ and $f(k_t, l_t) = k_t^a l_t^{1-a}$. We also assume that α, $\beta < 1$ and that capital fully depreciates each period.

These conditions are sufficient to guarantee a closed form solution to the planning problem:

$$c_t = (1 - \alpha\beta)k_t^\alpha \lambda_t, \quad \text{and}$$
$$k_{t+1} = i_t = \alpha\beta k_t^\alpha \lambda_t$$

where we identify as investment, i_t, the capital stock held over for production in period $t + 1$. These allocations are Pareto optimal.

We will show that the investment and consumption policy functions arising as a solution to this problem may be regarded as the aggregate investment and consumption functions arising from a *decentralized* homogenous consumer economy.

We first qualitatively describe the RCE underlying this model, and then demonstrate the relevant equilibrium price and quantity functions explicitly. The one capital good is assumed to produce two goods – a consumer good and an investment (capital) good. At the beginning of each period, firms observe the shock to productivity (λ_t) and purchase capital and labour from individuals at competitively determined rates. Both capital and labour are used to produce the two output goods. Individuals use their proceeds from the sale of capital and labour services to buy the consumption good (c_t) and the investment good (i_t) at the end of the period. This investment good is used as capital (k_t+1) available for sale to the firm next period and the process continues recursively.

To cast this problem formally as a recursive competitive equilibrium, we introduce some additional notation. Let k_t denote the capital holdings of a particular (measure zero) individual at time t, and $\underline{k}_t$ the distribution of capital amongst other individuals in the economy. This latter distinction allows us to make formal the competitive assumption: all the economic participants will assume that $\underline{k}_t$ is exogenous to them and that the price functions depend solely on this aggregate (in addition to the technology shock). Clearly, in equilibrium, $k_t = \underline{k}_t$ for our homogeneous consumer economy. In addition, let p_i, p_c and p_k be the price of the investment, consumption and capital goods respectively and p_l be the wage rate. These prices are presumed to be functions of the economy-wide state variables exclusively and all participants take these prices as given for their own decision making purposes. The 'state variables' characterizing the economy are ($\underline{k}$, λ) and the individual are (k, $\underline{k}$, λ).

We use the symbols (c, i, k, l) to denote points in the 'commodity space' for the firm and the consumer. The c in the commodity point of the firm is a function specifying the consumption good supplied by the firm and is written as $c^s(\underline{k}_t, \lambda_t)$. Similarly, the c in the commodity point of the individual is the amount of the consumption good demanded by the individual and is written as $c^d\ (k_t, \underline{k}_t, \lambda_t)$. In equilibrium (as mentioned earlier, in equilibrium $k_t = \underline{k}_t$), since the market clears, of course $c^s = c^d$. The same comments apply to the other elements of the commodity point.

In the decentralized version of this economy, the problem facing a typical household is

$$v(k_0, \underline{k}_0, \lambda_0) = \max\ E\left\{\sum_{t=0}^{\infty} \beta^t \ln\ c^d(k_t, \underline{k}_t, \lambda_t)\right\} \tag{P2}$$

subject to

$$\begin{aligned}&p_c(\underline{k}_t, \lambda_t)c^d(k_t, \underline{k}_t, \lambda_t)\\&\quad + p_i(\underline{k}_t, \lambda_t)i^d(k_t, \underline{k}_t, \lambda_t) \le p_k(\underline{k}_t, \lambda_t)\\&\quad \times k^s(k_t, \underline{k}_t, \lambda_t) + p_l(\underline{k}_t, \lambda_t)l^s(k_t, \underline{k}_t, \lambda_t)\end{aligned}$$

$$\begin{aligned}k_{t+1} &\equiv k^s\left(k_{t+1}, \underline{k}_{t+1}, \lambda_{t+1}\right)\\&= i^d(k_t, \underline{k}_t, \lambda_t), l^s(k_t, \underline{k}_t, \lambda_t) \le 1 \ \text{ and}\end{aligned}$$

$$k_{t+1} = \psi(\underline{k}_t, \lambda_t)$$

is the law of motion of the aggregate capital stock.

With capital and labour priced competitively each period, the firm's objective function is especially simple – maximize period profits. The firm's problem then is

$$\max\{p_c(\underline{k}_t,\lambda t)c^s(\underline{k}_t,\lambda_t)+p_i(\underline{k}_t,\lambda_t)i^s(\underline{k}_t,\lambda_t)$$
$$-p_k(\underline{k}_t,\lambda_t)k^d(\underline{k}_t,\lambda_t)-p_l(\underline{k}_t,\lambda_t)\,l^d(\underline{k}_t,\lambda_t)\}$$

subject to

$$c_t^s+i_t^s\le\lambda_t\left(k_t^d\right)^\alpha\left(l_t^d\right)^{1-\alpha}.$$

Via Bellman's principle of optimality, the recursive representation of the individual's problem P2 is

$$v(k_t,\underline{k}_t,\lambda_t)=\max_{\{c^d,i^d,l^s,k^d\}}\left\{\ln\left(c^d(k_t,\underline{k}_t,\lambda_t)\right)\right.$$
$$\left.+\beta\int v\left(i^d(k_t,\underline{k}_t,\lambda_t),\psi(\underline{k}_t,\lambda_t),\lambda_{t+1}\right)dF(\lambda_{t+1}|\,\lambda_t)\right\}$$

subject to

$$p_c(\underline{k}_t,\lambda_t)c^d(k_t,\underline{k}_t,\lambda_t)+p_i(\underline{k}_t,\lambda_t)i^d(\underline{k}_t,\underline{k}_t,\lambda_t)$$
$$\le p_k(\underline{k}_t,\lambda_t)\times k^s(k_t,\underline{k}_t,\lambda_t)+p_l(\underline{k}_t,\lambda_t)l^s(k_t,\underline{k}_t,\lambda_t)$$

$\times\, k_{t+1}\equiv k^s\left(k_{t+1},\underline{k}_{t+1},\lambda_{t+1}\right)=i^d(k_t,\underline{k}_t,\lambda_t),l^s$ $(k_t,\underline{k}_t,\lambda_t)\le 1$ and $\underline{k}_{t+1}=\psi(\underline{k}_t,\lambda_t)$

is the law of motion of the aggregate capital stock.

The firm of course, simply maximizes its period profits and hence does not have a multi-period problem.

The following functions that are a solution to the individual and firm maximization problem above satisfy the definition of recursive competitive equilibrium:

1. A value function $v(k_0,\underline{k}_0,\lambda_0)=E\left\{\sum_{t=0}^{\infty}\beta^{\mathrm{t}}\ln\left[(1-\alpha\beta)\,\lambda_t\underline{k}_t^{\alpha-1}\{a(k_t,-\underline{k}_t)+\underline{k}_t\}\right]\right\}$. It can be shown that $v(k_0,\underline{k}_0,\lambda_0)=A+Bln\underline{k}_0+Cln\lambda_0$ where A, B and C are constants which are functions of the preference and technology parameters.
2. A continuous pricing function $p(\underline{k}_t,\lambda t)=\{p_c(\underline{k}_t,\lambda_t),\ p_i(\underline{k}_t,\lambda_t),\ p_k(\underline{k}_t,\lambda_t),$ $p_l(\underline{k}_t,\lambda_t)\}$ that has the same dimensionality as the commodity point, where

$$pc(\underline{k}_t,\lambda_t)=p_i(\underline{k}_t,\lambda_t)=1$$

(We have chosen the consumption good to be the numeraire.)

$$p_k(\underline{k}_t,\lambda_t)=\alpha\lambda_t\underline{k}_t^{\alpha-1}$$
$$p_l(k_t,\lambda_t)=(1-\alpha)\lambda_t\underline{k}_t^{\alpha-1}.$$

3. Consumption and investment functions for the individual that are a function of the current state of the individual $(k,\underline{k},\lambda)$

$$c^d\left(k_t,\underline{k}_t,\lambda_t\right)=(1-\alpha\beta)\lambda_t\underline{k}_t^{\alpha-1}\{\alpha(k_t-k_t)+\underline{k}_t\}$$
$$l^s(k_t,\underline{k}_t,\lambda_t)=1$$
$$i^d(k_t,\underline{k}_t,\lambda_t)=\alpha\beta\lambda_t\underline{k}_t^{\alpha-1}\{a(k_t-\underline{k}_t)+\underline{k}_t\}$$
$$k^s\left(k_{t+1},\underline{k}_{t+1},\lambda_{t+1}\right)=i^d(k_t,\underline{k}_t,\lambda_t).$$

4. Decision rules for the firm that are contingent on the state of the economy $(\underline{k},\lambda)$

$$c^s(\underline{k}_t,\lambda_t)=(1-\alpha\beta)\lambda_t\underline{k}_t^\alpha,$$
$$l^d(\underline{k}_t,\lambda_t)=1,$$
$$i^s(\underline{k}_t,\lambda_t)=\alpha\beta\lambda_t\underline{k}_t^\alpha,$$
$$k^d\left(\underline{k}_{t+1},\lambda_{t+1}\right)=i^s(\underline{k}_t,\lambda_t).$$

5. The law of motion for the capital stock specifying the next period capital stock as a function of the current state of the economy $(\underline{k}_t,\lambda_t)$
$\underline{k}_{t+1}=\psi(\underline{k}_t,\lambda_t)=\alpha\beta\lambda_t\underline{k}_t^\alpha.$
6. The consumption and investment decisions of the individual $c^s(k,\underline{k},\lambda),l^s(k,\underline{k},\lambda)$ and $i^s(k,\underline{k},\lambda)$ maximize the expected utility subject to the budget constraint. So that

$$v(k_t,\underline{k}_t,\lambda_t)=\ln\left((1-\alpha\beta)\lambda_t\underline{k}_t^{\alpha-1}(\alpha(k_t-\underline{k}_t)+\underline{k}_t)\right)$$
$$+\beta\int v(\alpha\beta\lambda_t\underline{k}_t^{\alpha-1}(\alpha(k_t-\underline{k}_t)+\underline{k}_t),$$
$$\alpha\beta\lambda_t\underline{k}_t^\alpha)dF(\lambda_{t+1}|\,\lambda_t).$$

7. The decision rules of the firm $c^d(\underline{k}_t,\lambda_t),l^d(\underline{k}_t,\lambda_t)$ $,i^d(\underline{k}_t,\lambda_t)$ maximize firm profit.

Demand equals supply

$$c^d\left(k_{t+1},\underline{k}_{t+1},\lambda_{t+1}\right)=c^s(\underline{k}_t,\lambda_t),$$
$$l^s\left(k_{t+1},\underline{k}_{t+1},\lambda_{t+1}\right)$$
$$=l^d(\underline{k}_t,\lambda_t)\ \text{and}\ i^s\left(k_{t+1},\underline{k}_{t+1},\lambda_{t+1}\right)$$
$$=i^d(\underline{k}_t,\lambda_t).$$

The law of motion of the representative consumers capital stock is consistent with the maximizing behaviour of agents $\psi(\underline{k}_t,\lambda_t)=i^d(\underline{k}_t,\underline{k}_t,\lambda_t)$. It is readily demonstrated that since v

$(\underline{k}_0, \underline{k}_0, \lambda_0) = w(\underline{k}_0, \lambda_0)$, the competitive allocation is Pareto optimal. See Eqs. (P1) and (P2).

Having formulated expressions for the prices of the various assets and their laws of motion, it is a relatively simple matter to calculate rates of return (price ratios) and study their dynamics. For an application to risk premia, see Donaldson and Mehra (1984).

Some researchers have formulated models that can be cast in this same recursive setting, yet whose equilibria are not Pareto-optimal. As a consequence, the model's equilibrium can no longer be obtained as the solution to a central-planning-optimum formulation. These models incorporate various features of monetary phenomena, distortionary taxes, non-competitive labour market arrangements, externalities, or borrowing-lending constraints. Besides increasing general model realism, such features enable the models not only to better replicate the stylized facts of the business cycle, but also to provide a rationale for interventionist government policies. Monetary models of this class include those of Lucas and Stokey (1987, a monetary exchange model) and Coleman (1996, a monetary production model). Bizer and Judd (1989) and Coleman (1991) present models in which non-optimality is induced by tax distortions, while Danthine and Donaldson (1990) present a model in which non-optimality results from efficiency-wage considerations. In these models, equilibrium is characterized as an aggregate-consumption and an aggregate-investment function which jointly solves a system of first-order optimality equations on which market-clearing conditions have been imposed. Coleman (1991) provides a widely applicable set of conditions under which these suboptimal equilibrium functions exist. As already noted, however, these optimality conditions cannot, in general, characterize the solution to an optimum problem.

See Also

- ▶ Arrow–Debreu Model of General Equilibrium
- ▶ Decentralization
- ▶ Neoclassical Growth Theory
- ▶ Real Business Cycles

Bibliography

Bewley, T. 1972. Existence of equilibria in economies with infinitely many commodities. *Journal of Economic Theory* 4: 514–540.

Bizer, D., and K. Judd. 1989. Taxation and uncertainty. *American Economic Review Papers and Proceedings* 19: 331–336.

Brock, W.A., and L.J. Mirman. 1972. Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory* 4: 497–513.

Cass, D. 1965. Optimal growth in an aggregative model of capital accumulation. *Review of Economic Studies* 32: 233–240.

Coleman, W.J. 1991. Equilibrium in a production economy with an income tax. *Econometrica* 59: 1091–1104.

Coleman, W.J. 1996. Money and output: A test of reverse causation. *American Economic Review* 86: 90–111.

Danthine, J.P., and J.B. Donaldson. 1990. Efficiency wages and the business cycle puzzle. *European Economic Review* 34: 1275–1301.

Donaldson, J.B., and R. Mehra. 1983. Stochastic growth with correlated production shock. *Journal of Economic Theory* 29: 282–312.

Donaldson, J.B., and R. Mehra. 1984. Comparative dynamics of an equilibrium intertemporal asset pricing model. *Review of Economic Studies* 51: 491–508.

Harris, M. 1987. *Dynamic economic analysis*. New York: Oxford University Press.

Kydland, F.E., and E.C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50: 1345–1371.

Ljungqvist, L., and T.J. Sargent. 2004. *Recursive macroeconomic theory.* 2nd ed. Cambridge, MA: MIT Press.

Long, J.B. Jr., and C.I. Plosser. 1983. Real business cycles. *Journal of Political Economy* 91: 39–69.

Lucas, R.E. Jr., and N. Stokey. 1987. Money and interest in a cash advance economy. *Econometrica* 55: 491–513.

Mehra, R. 1988. On the existence and representation of equilibrium in an economy with growth and non-stationary consumption. *International Economic Review* 29: 131–135.

Mehra, R., and E.C. Prescott. 1977. Recursive competitive equilibria and capital asset pricing. In *Essays in financial economics*, ed. R. Mehra. Doctoral dissertation, Carnegie Mellon University.

Mehra, R., and E.C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15: 145–162.

Prescott, E.C., and R.E. Lucas Jr. 1972. A note on price systems in infinite dimensional space. *International Economic Review* 13: 416–422.

Prescott, E.C., and R. Mehra. 1980. Recursive competitive equilibria: The case of homogeneous households. *Econometrica* 48: 1365–1379.

Stokey, N., R.E. Lucas, and E.C. Prescott. 1989. *Recursive methods in economic dynamics*. Cambridge, MA: Harvard University Press.

Recursive Contracts

Albert Marcet

Abstract

A number of dynamic models in economics are formulated with forward-looking elements in the constraints – for example, models of risk-sharing with participation constraints and models of optimal policy. Here, standard dynamic programming does not apply. Recent contributions show how to reformulate these models by either rewriting the forward-looking constraints (promised utility approach) or by using a Lagrangean formulation (recursive Lagrangean). Both make it possible to obtain a recursive formulation that allows for easier computation and analytical results. A number of applications can be found to optimal fiscal or monetary policy, risk sharing or investment with various financial constraints, and employment decisions.

Keywords

Bellman equation; Commitment; Contract theory; Debt constraints; Dynamic programming; Incentive constraints; International capital flows; Lagrange multipliers; Optimal fiscal policy; Optimal monetary policy; Optimal taxation; Participation constraints; Principal and agent; Private information; Ramsey equilibria; Recursive contracts; Risk sharing; Saddle point functional equations; Time consistency; Unemployment insurance

JEL Classifications

D4; D10

In contract theory it is standard to introduce a participation constraint (PC) insuring that the contract offered to the agent delivers a utility higher than the best outside option. In a dynamic set-up agents may abandon the contract at any point in time, even after the contract has been in place for a while. For example, workers can leave a labour contract at almost no cost, or a borrower can stop repaying the loan if he or she declares bankruptcy. The possibility that the agent does not continue with the plan of the contract is usually called ‘default’. Hence, in a dynamic context, it is natural to require that the PC is satisfied in all periods, in order to avoid default.

It turns out that, if a PC in all periods and realizations is introduced in the design of the optimal contract, standard dynamic programming does not apply, the Bellman equation does not hold, and the solution is not guaranteed to be a time-invariant function of the usual state variables. This complicates enormously the solution of these models.

To discuss this in a simple risk-sharing model, consider two agents $i = 1,2$ with utility function $E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t^i\right)$, where $\beta \in (0, 1)$ is the discount factor and u the instantaneous utility. Each agent receives a stochastic endowment w_t^i and the realization of endowments is known both to the agents and the principal. The principal has full commitment, and will stick to his announced plan. Endowments provide the only supply of consumption good so that the following feasibility condition holds

$$c_t^1 + c_t^2 = w_t^1 + w_t^2 \tag{1}$$

A Pareto-optimal risk-sharing contract (implemented by a competitive equilibrium under complete markets) would set $\frac{u'\left(c_t^1\right)}{u'\left(c_t^2\right)}$ constant for all periods, so that agents would share all idiosyncratic risks. This allocation would be chosen as the optimal contract if agents would commit to never leave the risk-sharing arrangement. We refer to this allocation as the first best. The optimum satisfies the usual recursive structure in dynamic models, namely, that $c_t = F(w_t)$ where F is a time-invariant function and $w_t = \left(w_t^1, w_t^2\right)$.

Assume now agents cannot commit to staying in the contract for ever. An agent can leave the contract and consume for ever his individual endowment, so that a contract can only be implement if it satisfies

$$E_t \sum_{j=0}^{\infty} \beta^j u\left(c_{t+j}^i\right) \geq V_i^a(w_t)$$

at all periods and realizations, where $V_i^a(w_t) \equiv E_t \sum_{t=0}^{\infty} \beta^j u\left(w_{t+j}^i\right)$ is the utility of consuming in autarchy for ever after t.

It is clear that the above PC is likely to be violated by the first best allocation. In periods when w_t^i is high, the right side of the PC is high, but the agent has to surrender a large part of his endowment in the first best and the left side of the PC is too low. Therefore, PCs are often binding and they make the first best unfeasible.

A Pareto-optimal risk-sharing contract with PCs can now be found by maximizing the weighted utility of the two agents $E_0 \sum_{t=0}^{\infty} \beta^t \left[\lambda u\left(c_t^1 c\right) + (1-\lambda) u\left(c_t^2\right)\right]$ subject to the above PC for all periods and realizations and for both agents. The parameter λ indexes all such Pareto-optimal allocations. The result is an optimal contract under full commitment by the principal and partial commitment by the agents.

The Bellman equation does not give the solution to this problem. A key feature of standard dynamic programming is that the set of feasible actions must depend only on variables that were determined last period and the current shock. But it is possible to evaluate if a certain consumption level $\overline{c}_t^i$ satisfies the PC at time t only if future plans for consumption are known.

Intuitively, a promise of higher consumption in the future makes a lower consumption today compatible with the PC. But in order to implement this plan the principal has to 'remember' all the promises for higher consumption that were made in the past. Therefore, the optimal solution is unlikely to be a function of only today's endowment, the principal also needs to recall if, say, ten periods ago, the PC of one of the agents was binding.

As argued by Kydland and Prescott (1977), the same problem arises in models of optimal policy. The future restricts today's actions through the first order conditions of optimality of the agents, this causes the Bellman equation to fail and, in their language, the solution was time inconsistent. We find the same difficulty in contracting models of private information with incentive constraints, where some relevant piece of information is hidden from the principal, and more generally, in game theoretical models where an agent optimizes subject to the plans for the future of another agent.

If the Bellman equation fails, the solution could depend on all past shocks, and solving for the variables as a function of all past shocks would be very difficult. Too many variables would appear as arguments of the decision function. To overcome this difficulty the 'recursive contracts' literature provides several alternatives. The general idea is to recover a recursive formulation by adding a co-state variable.

One approach builds on the paper of Abreu et al. (1990; hereafter APS). To show how this can be applied in the above risk-sharing model with PCs, consider the case where w_t is i.i.d. and has two possible realizations $\overline{w}$ and $\overline{\overline{w}}$ with probabilities π and $(1-\pi)$. Denote the utility of agent i for the whole future at t if $w_t = \overline{w}$ by $\overline{V}_t^i \equiv \left(E_t \sum_{t=0}^{\infty} \beta^j u\left(c_{t+j}^i\right) | w_t = \overline{w}\right)$, and let $\overline{\overline{V}}_t^i$ be the analogue for realization $\overline{\overline{w}}$. The above PC can be reformulated as

$$V_t^i = u\left(c_t^i\right) + \beta\left(\pi \overline{V}_{t+1}^i + (1-\pi)\overline{\overline{V}}_{t+1}^i\right) \quad (2)$$

$$\overline{V}_{t+1}^i \geq \overline{V}_i^a(\overline{w}),\ \overline{\overline{V}}_{t+1}^i \geq \overline{V}_i^a\left(\overline{\overline{w}}\right) \text{ for all } t > 0,$$

where V_t^i is the actual realized utility. The first equation insures that V_t^i is the expected discounted utility, the second guarantees that the PC holds.

We can view the planner's choice at t as choosing the promised utilities $\overline{V}_{t+1}^i, \overline{\overline{V}}_{t+1}^i$ and consumption c_t^i,while V_t^i is given by past choices. It is clear that, in the APS approach, today's choice variables $x_t = \left(\overline{V}_{t+1}^i, \overline{\overline{V}}_{t+1}^i,\ c_t^i\right)$ are restricted by yesterday's promised utilities only, and the Bellman equation delivers the optimal contract after the realized V_t^i is included in the list of state variables. The promised utility V_t^i plays the same role as capital in a standard growth model, and (2) plays the role of the transition equation. Therefore, the optimal solution for the choices can be described recursively by a time-invariant function $x_t = F(w_t, V_t)$ for all $t > 0$.

A crucial caveat is that (2) is not sufficient to insure that the PCs are satisfied. The principal could choose arbitrarily high consumption and have ever higher Vs to satisfy (2), in a sort of Ponzi scheme for utility. The promised utilities have to be further restricted to belong to a feasible set. Let us call $\overline{\boldsymbol{S}} \subset R$ the feasible set of utilities such that, for each element $v \in \overline{\boldsymbol{S}}$, there is a sequence of consumptions $\left\{c^i_{t+j}\right\}$ that satisfy (1) and the PCs such that $v = \left(E_t \sum_{t=0}^{\infty} \beta^j u\left(c^i_{t+j}\right) | w_t = \overline{w}\right)$. Results in APS insure that this set is convex. Since in this case $\overline{\boldsymbol{S}} \subset R$, this set is an interval and there exist bounds $\overline{V}^i_L$ and $\overline{V}^i_U$ such that adding the constraints

$$\overline{V}^i_L \leq \overline{V}^i_{t+1} \leq \overline{V}_U$$

(and similarly for $\overline{\overline{V}}^i_t$) to (2) is enough to insure feasibility. These bounds can be easily introduced in the Bellman equation and this guarantees that the chosen consumption sequences satisfies the PC. The only complication is that upper bound $\overline{V}_U$ needs to be computed separately, as it is not a datum of the problem ($\overline{V}^i_L$ is trivially equal to V^a_i $(\overline{w})$).

Another difference with standard dynamic programming is that the initial utility V^1_0 is an outcome of the solution and it is not fixed beforehand. This feature shows how time inconsistency arises in this model, since the choice for V in period zero is not given, but in future periods it is given from the past.

Promised utilities as co-states have been used extensively in models with incentive or participation constraints. Among others, Phelan and Townsend (1991) studied a model of risk-sharing with incentive constraints, Kocherlakota (1996) analysed the risk-sharing model with the PC described above, Hopenhayn and Nicolini (1997) a model of unemployment insurance and Alvarez and Jermann (2000) a decentralized version of the above risk-sharing model with debt constraints. In models of Ramsey equilibria it has been used by Golosov et al. (2003) to study optimal taxation under private information and Chang (1998) in a model of optimal monetary policy.

The main problem with this approach is the computation of the set of feasible utilities. In the specific model described above this is not too costly, because it involves finding only two numbers, namely, the upper bounds $\overline{V}_U, \overline{\overline{V}}_U$. But the difficulties multiply when more than one co-state variable is needed. For example, if a third agent is included in the above risk-sharing model, the co-state variables would be (V^1_t, V^2_t). Results in APS guarantee that the set of feasible utilities $\overline{\boldsymbol{S}}$ $\subset R^2$ is convex, but now it is a generic set, not an interval. Computing a set is much harder than computing two numbers. Some papers overcome these difficulties; for example, Abraham and Pavoni (2005), who show how to find such a set in a model of saving under private information, or the paper of Judd et al. (2003). But the difficulties increase very fast with the dimensionality of the promised utilities.

Furthermore, in some models, the set of feasible promised utilities changes every period. If a 'traditional' state variable (say, capital stock) appears in the problem, the set of feasible utilities is different depending on the level of capital, so that the feasible set is now given by a correspondence $\overline{\boldsymbol{S}}(k)$. The researcher now needs to solve for a mapping from capital stock to sets. Phelan and Stacchetti (2001) compute in this way the optimal fiscal policy in a model with capital.

An alternative to APS is the Lagrangean approach described in Marcet and Marimon (1998). The Lagrangean for the optimal risk-sharing problem with PC is

$$L = E_0 \sum_{t=0}^{\infty} \beta^t \left[\lambda u\left(c^1_t\right) + (1-\lambda) u\left(c^2_t\right) + \sum_{i=1,2} \gamma^i_t \left(E_t \sum_{j=0}^{\infty} \beta^j u\left(c^i_{t+j}\right) - V^a_i(w_t) \right) \right]$$

where $\gamma^i_t \geq 0$ is the Lagrange multiplier of the PC. This can be rewritten as

$$L = E_0 \sum_{t=0}^{\infty} \beta^t \left[\left(\lambda + \mu^1_t\right) u\left(c^1_t\right) + \left(1 - \lambda + \mu^2_t\right) u\left(c^2_t\right) \right]$$
$$\text{s.t. } \mu^i_t = \mu^i_{t-1} + \gamma^i_t, \gamma^i_t \geq 0 \; \mu^i_{-1} = 0$$

In this formulation, only current and past variables enter in the objective and in the constraints of this Lagrangean, and a proper initial condition for μ is given. In this approach, μ_t plays the role of the co-state variable instead of the promised utility in the APS approach. A saddle point functional equation (analogous but not equal to the Bellman equation) is satisfied, insuring that the optimal solution satisfies $(c_t, \gamma_t) = G(\mu_{t-1}, w_t)$ with $\mu^i_{-1} = 0$ for a time invariant function G.

The equilibrium satisfies $\frac{u'(c_t^1)}{u'(c_t^2)} = \frac{1-\lambda+\mu_t^2}{\lambda+\mu_t^1}$. If the PC for agent i is binding, the corresponding γ_t^i is strictly positive, the weight μ_t^i goes up and so does c_t^i. The increase in μ is permanent (at least until another PC is binding). In this way the principal avoids default by spreading the reward over time in order to enhance smoothing of consumption.

Note that the initial value of μ is given and equal to zero, while in future periods μ_{t-1} needs to be set according to past Lagrange multipliers. Therefore, the initial value of the co-state does not need to be found separately as in APS. It is clear that, if the principal could re-optimize ignoring past commitments at sometime t, he or she would ignore the past co-state and reset $\mu = 0$. This is how time inconsistency is reflected in this formulation.

In the Lagrangean approach there is no need to find the set of feasible utilities. The only constraint on the co-states is the non-negativity constraint on γs. Application to models with capital accumulation and several co-states is much easier; for example, Marcet and Marimon (1992) solve a risk-sharing growth model with PC as described above and capital accumulation, Aiyagari et al. (2002) in a Ramsey equilibrium for fiscal policy under incomplete markets, where debt is a state variable, Attanasio and Ríos-Rull (2000) risk-sharing in small villages, Scott (2007) a model of optimal taxes with capital, Kehoe and Perri (2002) international capital flows with capital accumulation under PC, King, Kahn and Wolman (2003) optimal monetary policy, Cooley et al. (2004) a model of investment under private information, Abraham and Carceles-Poveda (2006) discuss how to decentralize a model with participation constraints, and Ferrero and Marcet (2004) and Scholl (2004) a model of temporary exclusion in the case of default. The drawback of the Lagrangean approach is that, at this writing, the theory for the non-convex case and for the private information case is still incomplete.

See Also

- Agency Problems
- Bellman Equation
- Dynamic Programming
- Income Taxation and Optimal Policies
- Optimal Fiscal and Monetary Policy (with Commitment)
- Optimal Taxation
- Risk Sharing
- Time Consistency of Monetary and Fiscal Policy

Bibliography

Abraham, A., and E. Carceles-Poveda. 2006. Endogenous incomplete markets, enforcement constraints, and intermediation. *Theoretical Economics* 1: 439–459.

Abraham, A., and N. Pavoni. 2005. The efficient allocation of consumption under moral hazard and hidden access to the credit market. *Journal of the European Economic Association* 3: 370–381.

Abreu, D., D. Pierce, and E. Stachetti. 1990. Towards a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58: 1041–1063.

Aiyagari, R., A. Marcet, T.J. Sargent, and J. Seppälä. 2002. Optimal Taxation without state-contingent debt. *Journal of Political Economy* 110: 1220–1254.

Alvarez, F., and U.J. Jermann. 2000. Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica* 68: 775–798.

Attanasio, O., and J.V. Ríos-Rull. 2000. Consumption smoothing in island economies: Can public insurance reduce welfare? *European Economic Review* 44: 1225–1258.

Chang, R. 1998. Credible monetary policy with long-lived agents: Recursive approaches. *Journal of Economic Theory* 81: 431–461.

Cooley, T.F., R. Marimon, and V. Quadrini. 2004. Aggregate consequences of limited contract enforceability. *Journal of Political Economy* 112: 817–847.

Ferrero, G., and A. Marcet. 2004. Limited commitment and temporary exclusion. Mimeo, Institut d'Analisi Economica, CSIC.

Golosov, M., N.R. Kocherlakota, and A. Tsyvinski. 2003. Optimal indirect and capital taxation. *Review of Economic Studies* 70: 569–587.

Hopenhayn, H.A., and J.P. Nicolini. 1997. Optimal unemployment insurance. *Journal of Political Economy* 105: 412–438.

R

Judd, K.L., S. Yeltekin, and J. Conklin. 2003. Computing supergame equilibria. *Econometrica* 71: 1239–1254.
Kehoe, P.J., and F. Perri. 2002. International business cycles with endogenous incomplete markets. *Econometrica* 70: 907–928.
King, R.G., A. Kahn, and A.L. Wolman. 2003. Optimal monetary policy. *Review of Economic Studies* 70: 825–860.
Kocherlakota, N.R. 1996. Implications of efficient risk sharing without commitment. *Review of Economic Studies* 63: 595–609.
Kydland, F.E., and E.C. Prescott. 1977. Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy* 85: 473–492.
Marcet, A., and R. Marimon. 1992. Communication, commitment and growth. *Journal of Economic Theory* 58: 219–249.
Marcet, A., and R. Marimon. 1998. *Recursive contracts*. Working paper, Universitat Pompeu Fabra.
Phelan, C., and E. Stacchetti. 2001. Sequential equilibria in a Ramsey tax model. *Econometrica* 69: 1491–1518.
Phelan, C., and R.M. Townsend. 1991. Computing multiperiod, information-constrained optima. *Review of Economic Studies* 58: 853–881.
Scholl, A. 2004. *Do endogenous incomplete markets explain cross-country consumption correlations and the dynamics of the terms of trade?* Berlin: Mimeo, Humboldt University.
Scott, A. 2007. Optimal taxation and OECD labor taxes. *Journal of Monetary Economics* 54: 925–944.

Recursive Preferences

David K. Backus, Bryan R. Routledge and Stanley E. Zin

Abstract

Recursive preferences characterize the trade-offs between current and future consumption by summarizing the future with a single index, the certainty equivalent of next period's utility. Recursive utility functions are built from two components. A risk aggregator encodes trade-offs across the outcomes of a static gamble and, hence, defines the certainty equivalent of future utility. A time aggregator encodes trade-offs between current consumption and the certainty equivalent of future utility. We suggest functional forms for time and risk aggregators with desirable properties for applications in economics and finance, such as the standard intertemporal consumption/portfolio problem, which we solve using dynamic programming.

Keywords

Bellman equation; Certainty equivalent; Disappointment aversion; Dynamic optimization; Elasticity of intertemporal substitution; Expected utility; Impatience; Infinite horizons; Preferences; Rational expectations; Recursive preferences; Risk aggregator; Risk aversion; Stochastic dynamic models; Time aggregator; Time preference; Utility functions; Weighted utility

JEL Classifications

D4; D10

Introduction

Recursive methods have become a standard tool for studying economic behaviour in dynamic stochastic environments. In this chapter, we characterize the class of preferences that is the natural complement to this framework, namely *recursive preferences*.

Why model preferences rather than behaviour? Preferences play two critical roles in economic models. First, preferences provide, in principle, an unchanging feature of a model in which agents can be confronted with a wide range of different environments, institutions, or policies. For each environment, we derive behaviour (decision rules) from the same preferences. If we modelled behaviour directly, we would also have to model how it adjusted to changing circumstances. The second role played by preferences is to allow us to evaluate the welfare effects of changing policies or circumstances. Without the ranking of opportunities that a model of preferences provides, it's not clear how we should distinguish good policies from bad.

Why recursive preferences? Recursive preferences focus on the trade-off between current-period utility and the utility to be derived from all future periods. Since an agent's actions today can affect the evolution of opportunities in the future, summarizing the future consequences of these actions with a single index, that is, future utility, allows multi-period decision problems to be reduced to a series of two-period problems, and in the case of a stationary infinite-horizon problem, a single, time-invariant two-period decision problem. As we will see, this logic applies equally well to environments in which current actions affect the values of random events for all future periods. In this case, the two-period trade-off is between current utility and a *certainty equivalent* of random future utility. This recursive approach not only allows complicated dynamic optimization problems to be characterized as much simpler and more intuitive two-period problems, it also lends itself to straightforward computational methods. Since many computational algorithms for solving stochastic dynamic models themselves rely on recursive methods, numerical versions of recursive utility models can be solved and simulated using standard algorithms.

The Stationary Recursive Utility Function

Assume time is discrete, with dates $t = 0,1,2,\ldots$. At each $t > 0$, an event z_t is drawn from a finite set $\mathcal{Z}$, following an initial event z_0. The t-period history of events is denoted by $z^t = (z_0, z_1, \ldots, z_t)$ and the set of possible t-histories by $\mathcal{Z}^t$.

Environments like this, involving time and uncertainty, are the starting point for much of modern economics. A typical agent in such a setting has preferences over payoffs $c(z^t)$ for each possible history. A general set of preferences might be represented by a utility function $U(\{c(z^t)\})$. In what follows, we will think of consumption as a scalar. This is purely for exposition since the extension to a vector of consumption at each point in time is straightforward.

Consider the structure of preferences in this dynamic stochastic environment.

We define the class of stationary recursive preferences by

$$U_t = V[c_t\ , \mu_t\ (U_{t+1})], \tag{1}$$

where U_t is short-hand for utility starting at some date-t history z^t, U_{t+1} refers to utilities for histories $z^{t+1} = (z^t, z_{t+1})$ stemming from z^t, V is a time aggregator, and μ_t is a certainty-equivalent function based on the conditional probabilities $p(z_{t+1}|z^t)$. As with other utility functions, increasing functions of U, with suitable adjustment of μ, imply the same preferences. This structure of preferences leads naturally to recursive solutions of economic problems, with (1) providing the core of a Bellman equation.

In general, the properties of U_t depend on both the properties of the time aggregator and the certainty equivalent. Since the certainty equivalent will be scaled such that $\mu(x) = x$ when x is a perfect certainty, the time aggregator V is all that matters in deterministic settings. Similarly, for a purely static problem with uncertainty, the certainty-equivalent function μ is all that matters. We consider the specification of each of these components in turn.

It is important to note that the utility functions presented in this article are not ad hoc but rather have clear axiomatic foundations, and can be derived from more primitive assumptions on preference orderings. Since utility functions are the typical starting point for applied research, we skip this step and refer the interested reader to the axiomatic characterizations of recursive preferences in the papers cited at the end of this article.

R

The Time Aggregator

Time preference is a natural starting point. Suppose there is no risk and c_t is one-dimensional. Preferences might then be characterized by a general utility function U ($\{c_t\}$). A common measure of time preference in this setting is the marginal rate of substitution between consumption at two consecutive dates (c_t and c_{t+1}, say) along a constant consumption path ($c_t = c$ for all t). If the marginal rate of substitution is

$$\mathrm{MRS}_{t,t+1} = \frac{\partial U/\partial c_{t+1}}{\partial U = \partial c_t}$$

then time preference is captured by the discount factor

$$\beta(c) \equiv \mathrm{MRS}_{t,t+1}(c).$$

(Picture the slope, $-1/\beta$, of an indifference curve along the '45-degree line'.) If $\beta(c)$ is less than one, the agent is said to be impatient: along a constant consumption path (that is, in the absence of diminishing marginal utility considerations), the agent requires more than one unit of consumption at $t+1$ to induce a sacrifice of one unit at t.

For the traditional time-additive utility function,

$$U(\{c_t\}) = \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{2}$$

$\beta(c) = \beta < 1$ regardless of the value of c, so impatience is built in and constant. A popular and useful special case of this utility function implies a constant elasticity of intertemporal substitution by assuming $u(c) = c^{\rho}/\rho$ for $\rho < 1$. Note that we can define the utility function in (2) recursively:

$$U_t = u(c_t) + \beta U_{t+1}, \tag{3}$$

for $t = 1,2,\ldots$. The constant elasticity version can be expressed

$$U_t = \left[(1-\beta)c_t^{\rho} + \beta U_{t+1}^{\rho}\right]^{1/\rho}, \tag{4}$$

where $\rho < 1$ and $\sigma = 1/(1-\rho)$ is the intertemporal elasticity of substitution. (To put this in additive form, use the transformation $\hat{U} = U^{\rho}/\rho$.) Note that U_t is homothetic and that the scaling we have chosen measures utility on the same scale as consumption:

$$U(c,c,c,\ldots) = c.$$

More generally, impatience summarized by the discount factor, $\beta(c)$, could vary with the level of consumption. Koopmans (1960) derives a class of stationary recursive preferences by imposing conditions on a general utility function U for a multi-dimensional consumption vector c. In the Koopmans class of preferences, time preference is a property of the time aggregator V. Consider our measure of time preference:

$$U_t = V(c_t, U_{t+1}) = V[c_t, V(c_{t+1}, U_{t+2})].$$

The marginal rate of substitution between c_t and c_{t+1} is therefore

$$\mathrm{MRS}_{t,t+1} = \frac{V_2(c_t, U_{t+1})V_1(c_{t+1}, U_{t+2})}{V_1(c_t, U_{t+1})}$$

A constant consumption path at c is defined by $U = V(c, U)$, implying $U = g(c) = V[c, g(c)]$ for some function g.

In modern applications, we typically work in reverse order: we specify a time aggregator V and use it to characterize the overall utility function U. Any U constructed this way defines preferences that are stationary and dynamically consistent. In contrast to time-additive preferences, discounting depends on the level of consumption c.

The most common example of Koopmans's structure in applications is a generalization of Eq. (3):

$$V(c, U) = u(c) + \beta(c)U,$$

where there is no particular relationship between the functions u and β. For this example, the intertemporal trade-off is given by

$$\mathrm{MRS}_{t,t+1} = \beta(c_t)\left[\frac{u'(c_{t+1}) + \beta'(c_{t+1})U_{t+2}}{u'(c_t) + \beta'(c_t)U_{t+1}}\right].$$

When $\beta'(c) \neq 0$, optimal consumption plans will depend on the level of future utility. And along a constant consumption path, discounting is decreasing (increasing) in consumption when $\beta'(c) < 0$ ($\beta'(c) > 0$). Also note that U_t in this example is not homothetic.

The Risk Aggregator

Turn now to the specification of risk preferences, which we consider initially in a static setting. Choices have risky consequences or payoffs, and agents have preferences defined over those consequences and their probabilities. To be specific, let us say that the state z is drawn with probability $p(z)$ from the finite set $\mathcal{Z} = \{1, 2,..., Z\}$. Consequences ($c$, say) depend on the state and the agent's preferences are represented by a utility function of state-contingent consequences ('consumption'):

$$U(\{c(z)\}) = U[c(1), c(2),..., c(Z)].$$

At this level of generality there is no mention of probabilities, although we can well imagine that the probabilities of the various states will show up somehow in U. We regard the probabilities as known, which you might think of as an assumption of 'rational expectations'.

We prefer to work with a different (but equivalent) representation of preferences. Suppose, for the time being, that c is a scalar; very little of the theory depends on this, but it streamlines the presentation. We define the *certainty equivalent* of a set of consequences as a certain consequence μ that gives the same level of utility:

$$U(\mu, \mu,..., \mu) = U[c(1), c(2),..., c(Z)].$$

If U is increasing in all its arguments, we can solve this for the certainty-equivalent function $\mu(\{c(z)\})$. Clearly μ represents the same preferences as U, but we find its form particularly useful. For one thing, it expresses utility in payoff ('consumption') units. For another, it summarizes behaviour towards risk directly: since the certainty equivalent of a sure thing is itself, the impact of risk is simply the difference between the certainty equivalent and expected consumption.

The traditional approach to preferences in this setting is expected utility, which takes the form

$$U(\{c(z)\}) = \sum_z p(s)u[c(z)] = Eu(c),$$

or

$$\mu(\{c(z)\}) = u^{-1}\left(\sum_z p(z)u[c(z)]\right)$$
$$= u^{-1}[Eu(c)].$$

Preferences of this form have been used in virtually all economic theory. The utility function of Kreps and Porteus employs a general time aggregator and an expected utility certainty equivalent. Following Epstein and Zin, many recent applications, particularly in dynamic asset pricing models, use the homothetic version of this utility function which combines the constant elasticity time aggregator in (4) with a linear homogeneous (constant relative risk aversion) expected utility certainty equivalent.

Empirical research both in the laboratory and in the field has documented a variety of difficulties with the predictions of expected utility models. In particular, people seem more averse to bad outcomes than implied by expected utility. In response to this evidence, there is a growing body of work that looks at decision making under uncertainty outside of the traditional expected utility framework. Without surveying all of these extensions, we demonstrate the basic mechanics of recursive utility with non-expected utility certainty equivalents by studying one particular analytically convenient class of preferences in detail, the Chew–Dekel class. Notable among the alternatives to this class are recursive and dynamic extensions of the Gilboa and Schmeidler 'max-min' preferences.

The Chew–Dekel certainty equivalent function μ for a set of payoffs and probabilities $\{c(z), p(z)\}$ is defined implicitly by a *risk aggregator* M satisfying

$$\mu = \sum_z p(z)M[c(z), \mu]. \tag{5}$$

Such preferences satisfy a weaker condition than the notorious independence axiom that underlies expected utility, yet like expected utility, they lead to first-order conditions in decision problems that are linear in probabilities, hence easily solved and amenable to econometric analysis. We assume M has the following properties: (i) $M(m, m) = m$ (sure things are their own certainty equivalents), (ii) M is increasing in its first argument (first-order stochastic dominance), (iii) M is concave in its first argument (risk aversion), and (iv) $M(kc, km) = kM(c, m)$ for $k > 0$ (linear

R

homogeneity). Most of the analytical convenience of the Chew–Dekel class follows from the linearity of Eq. (5) in probabilities. (Note that this implies that indifference curves on the probability simplex are linear, but not necessarily parallel.)

Examples of tractable members of the Chew–Dekel class include the following:

1. *Expected utility.* A version with constant relative risk aversion (that is, linear homogeneity) is implied by

$$M(c,m) = c^{\alpha} m^{1-\alpha}/\alpha + m(1 - 1/\alpha).$$

If $\alpha \leq 1$, M satisfies the conditions outlined above. Applying (5), we find

$$\mu = \left(\sum_z p(z) c(z)^{\alpha}\right)^{1=\alpha},$$

the usual expected utility with a power utility function.

2. *Weighted utility.* A relatively easy way to generalize expected utility given (5): weight the probabilities by a function of outcomes. A constant-elasticity version follows from

$$M(c,m) = (c/m)^{\gamma} c^{\alpha} m^{1-\alpha}/\alpha \\ + m[1 - (c/m)^{\gamma} = \alpha].$$

For M to be increasing and concave in c in a neighbourhood of m, the parameters must satisfy either (a) $0 < \gamma < 1$ and $\alpha + \gamma < 0$ or (b) $\gamma < 0$ and $0 < \alpha + \gamma < 1$. Note that (a) implies $\alpha < 0$, (b) implies $\alpha > 0$, and both imply $\alpha + 2\gamma < 1$. The associated certainty equivalent function is

$$\mu^{\alpha} = \frac{\sum_z p(z) c(z)^{\gamma+\alpha}}{\sum_x p(x) c(x)^{\gamma}} = \sum_z \hat{p}(z) c(z)^{\alpha},$$

where

$$\hat{p}(z) = \frac{p(z) c(z)^{\gamma}}{\sum_x p(x) c(x)^{\gamma}}.$$

This version highlights the impact of bad outcomes: they get greater weight than with expected utility if $\gamma < 0$, less weight otherwise.

3. *Disappointment aversion*. Another model that increases sensitivity to bad events ‘disappointments’) is defined by the risk aggregator

$$M(c,m) = \begin{cases} c^{\alpha} m^{1-\alpha}/\alpha + m(1 - 1/\alpha) & c \geq m \\ c^{\alpha} m^{1-\alpha}/\alpha + m(1 - 1/\alpha) + \delta(c^{\alpha} m^{1-\alpha} - m)/\alpha & c < m \end{cases}$$

with $\delta \geq 0$. When $\delta = 0$ this reduces to expected utility. Otherwise, disappointment aversion places additional weight on outcomes worse than the certainty equivalent. The certainty equivalent function satisfies

$$\begin{aligned} \mu^{\alpha} &= \sum_z p(z) c(z)^{\alpha} \\ &\quad + \delta \sum_z p(z) I[c(z) < \mu][c(z)^{\alpha} - \mu^{\alpha}] \\ &= \sum_z \hat{p}(z) c(z)^{\alpha}, \end{aligned}$$

where $I(x)$ is an indicator function that equals one if x is true and zero otherwise, and

$$\hat{p}(z) = \left(\frac{1 + \delta I[c(z) < \mu]}{1 + \delta \sum_x p(x) I[c(x) < \mu]}\right) p(z).$$

It differs from weighted utility in scaling up the probabilities of all bad events by the same factor, and scaling down the probabilities of good events by a complementary factor, with good and bad defined as better and worse than the certainty equivalent. (This implies a ‘kink’ in state-space indifference curves at certainty, which is referred to as 'first-order' risk aversion.) All three expressions highlight the *recursive* nature of the risk aggregator M: we need to know the certainty equivalent to know which states are bad so that we can compute the certainty equivalent (and so on).

Optimization and the Bellman Equation

For an illustrative application of recursive utility, we turn to the classic Merton–Samuelson consumption/portfolio-choice problem. Consider a stationary Markov environment with states z and conditional probabilities $p(z'|z)$. Preferences are represented by a constant-discounting/constant-elasticity aggregator and a general linear homogeneous certainty equivalent. A dynamic consumption/portfolio problem for this environment is characterized by the Bellman equation which implicitly defines the value function:

$$J(a,z) = \max_{c,w} \left\{(1-\beta)c^{\rho} + \beta\mu\left[J(a',z')\right]^{\rho}\right\}^{1/\rho},$$

subject to the wealth constraint, $a' = (a-c)\sum_i w_i r_i(z,z') = (a-c)\sum_i w_i r'_i = (a-c) r'_p$, where a denotes wealth, r_p is the return on the portfolio $\left(w_1, w_2, \ldots, w_{N-1}, 1-\sum_{i=1}^{N-1} w_i\right)$, of assets with risky returns $(r_1, r_2, \ldots, r_N)$. The budget constraint and linear homogeneity of the time and risk aggregators imply linear homogeneity of the value function: $J(a, z) = aL(z)$ for some scaled value function L. The scaled Bellman equation is

$$L(z) = \max_{b,w} \left\{(1-\beta)b^{\rho} + \beta(1-b)^{\rho}\, \mu\left[L(z')r_p(z,z')\right]^{\rho\ 1/\rho},\right.$$

where $b \equiv c/a$. Note that $L(z)$ is the marginal utility of wealth in state z.

This problem divides into separate portfolio and consumption decisions. The portfolio decision solves: choose $\{w_i\}$ to maximize $\mu[L(z')r_p(z, z')]$. The portfolio first-order conditions are

$$\begin{aligned}\textstyle\sum_{z'} p(z'\,|\,z) M_1\left[L(z')r_p(z,z'),\mu\right] \\ L(z')\left[r_i(z,z') - r_j(z,z')\right] = 0\end{aligned} \qquad (6)$$

for any two assets i and j.

Given a maximized μ, the consumption decision solves: choose b to maximize L. The intertemporal first-order condition is

$$(1-\beta)b^{\rho-1} = \beta(1-b)^{\rho-1}\mu^{\rho}. \qquad (7)$$

If we solve for μ and substitute into the (scaled) Bellman equation, we find

$$\begin{aligned}\mu &= [(1-\beta)/\beta]^{1/\rho}[b/(1-b)]^{(\rho-1)/\rho} \\ L &= (1-\beta)^{1/\rho} b^{(\rho-1)/\rho}.\end{aligned} \qquad (8)$$

The first-order condition (7) and value function (8) allow us to express the relation between consumption and returns in a familiar form. Since μ is linear homogeneous, the first-order condition implies $\mu(x'r'_p) = 1$ for

$$x' = L'/\mu\left[\beta(c'/c)^{\rho-1}\left(r'_p\right)^{1-\rho}\right]^{1/\rho}.$$

The last equality follows from $(c'/c) = (b'/b)(1-b)r'_p$, a consequence of the budget constraint and the definition of b. The intertemporal first-order condition can therefore be expressed

$$\mu\left(x'r'_p\right) = \mu\left(\left[\beta(c'/c)^{\rho-1} r'_p\right]^{1/\rho}\right) = 1, \qquad (9)$$

a generalization of the tangency condition for an optimum (set the marginal rate of substitution equal to the price ratio). Similar logic leads us to express the portfolio first-order conditions (6) as

$$E\left[M_1\left(x'r'_p, 1\right)x'\left(r'_i - r'_j\right)\right] = 0.$$

If we multiply by the portfolio weight w_j and sum over j we find

$$E\left[M_1\left(x'r'_p, 1\right)x'r'_i\right] = E\left[M_1\left(x'r'_p, 1\right)x'r'_p\right]. \qquad (10)$$

R

Euler's theorem for homogeneous functions allows us to express the right side as

$$E\left[M_1\left(x'r'_p,1\right)x'r'_p\right] = 1 - EM_2\left(x'r'_p,1\right).$$

Whether this expression is helpful depends on the precise form of M. For example, with disappointment aversion, (10) is

$$E\left[z^{\alpha-1}(1+\delta I[z<1])\frac{r'_i}{r'_p}\right] = 1 + \delta EI[z<1],$$

where $z = \left[\beta(c'/c)^{\rho-1}r'_p\right]^{1/\rho}$. This reduces to the Kreps–Porteus model when $\delta = 0$, and to the time-additive expected utility model when, in addition, $\rho = \alpha$.

Conclusion

A recursive utility function can be constructed from two components: (*a*) a time aggregator that completely characterizes preferences in the absence of uncertainty and (*b*) a risk aggregator that defines the certainty equivalent function that characterizes preferences over static gambles and is used to aggregate the risk associated with future utility. We looked at natural candidates for each of these components and gave an example of how Bellman's equation can be used to characterize optimal plans in a dynamic stochastic environment when agents have recursive preferences.

Further Reading

For more on this subject, see Backus et al. (2004) and the references cited there. Much of the material in this chapter builds from Epstein and Zin (1989), who extend the preferences in Kreps and Porteus (1978) to allow for a stationary infinite-horizon model and for non-expected utility certainty equivalents. They also derive the consumption/portfolio-choice results of section "Optimization and the Bellman Equation". For more on time aggregators, see Koopmans (1960), Uzawa (1968), Epstein and Hynes (1983), Lucas and Stokey (1984), and Shi (1994). Common departures from expected utility are documented in Kreps (1988, ch. 14) and Starmer (2000). Epstein and Schneider (2003) and Hansen and Sargent (2004) propose different dynamic and recursive extensions of the max-min risk preference of Gilboa and Schmeidler (1993). The Chew–Dekel risk aggregator was proposed by Chew (1983, 1989) and Dekel (1986). Examples within this class: weighted utility (Chew 1983), disappointment aversion (Gul 1991), semi-weighted utility (Epstein and Zin 2001), and generalized disappointment aversion (Routledge and Zin 2003).

See Also

▶ Bellman Equation
▶ Time Preference

Bibliography

Backus, D.K., Routledge, B.R. and Zin, S.E. 2004. Exotic preferences for macroeconomists. *NBER macroeconomics annual 2004 , vol. 19* M. Gertler K. Rogoff. Cambridge, MA: MIT Press.

Chew, S.H. 1983. A generalization of the quasi-linear mean with applications to the measurement of inequality and decision theory resolving the Allais paradox. *Econometrica* 51: 1065–1092.

Chew, S.H. 1989. Axiomatic utility theories with the betweenness property. *Annals of Operations Research* 19: 273–298.

Dekel, E. 1986. An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory* 40: 304–318.

Epstein, L.G., and J.A. Hynes. 1983. The rate of time preference and dynamic economic analysis. *Journal of Political Economy* 91: 611–635.

Epstein, L.G., and M. Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113: 1–31.

Epstein, L.G., and S.E. Zin. 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57: 937–969.

Epstein, L.G., and S.E. Zin. 2001. The independence axiom and asset returns. *Journal of Empirical Finance* 8: 537–572.

Gilboa, I., and D. Schmeidler. 1993. Updating ambiguous beliefs. *Journal of Economic Theory* 59: 33–49.

Gul, F. 1991. A theory of disappointment aversion. *Econometrica* 59: 667–686.

Hansen, L.P., and T.J. Sargent. 2004. *Misspecification in recursive macroeconomic theory*. Manuscript: University of Chicago, January.
Koopmans, T.C. 1960. Stationary ordinal utility and impatience. *Econometrica* 28: 287–309.
Kreps, D.M. 1988. *Notes on the theory of choice*. Boulder: Westview Press.
Kreps, D.M., and E.L. Porteus. 1978. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46: 185–200.
Lucas, R.E., and N.L. Stokey. 1984. Optimal growth with many consumers. *Journal of Economic Theory* 32: 139–171.
Routledge, B.R. and Zin, S.E. 2003. Generalized disappointment aversion and asset prices. Working Paper No. 10107. Cambridge, MA: NBER.
Shi, S. 1994. Weakly nonseparable preferences and distortionary taxes in a small open economy. *International Economic Review* 35: 411–428.
Starmer, C. 2000. Developments in non-expected utility theory. *Journal of Economic Literature* 38: 332–382.
Uzawa, H. 1968. Time preference, the consumption function, and optimum asset holdings. In *Value, capital, and growth: Papers in honour of Sir John Hicks*, ed. J.N. Wolfe. Chicago: Aldine.

Reddaway, William Brian (born 1913)

G. C. Harcourt

Brian Reddaway was born in 1913 in Cambridge, England. He read economics at King's College, Cambridge (1932–4). Keynes supervised him at the time when he was writing the *General Theory*. Reddaway absorbed its message so well that he wrote one of the most perceptive reviews (1936) of the book. Reddaway has in recent years played a prominent role in defending Keynesian theory and policy against monetarist critics. Nevertheless, he is an openminded eclectic, accepting ideas from any approach provided that they have an empirical foundation. In Australia, working with Giblin as a Research Fellow in Economics at the University of Melbourne, he so distinguished himself by his evidence (1937) on the Basic Wage to the Arbitration Court that the Wage itself in the year of his evidence (1937) became known as 'The Reddawage'.

In 1938 Reddaway returned to the United Kingdom to a Fellowship at Clare and also to a teaching post in the Faculty of Economics and Politics at Cambridge in 1939. He worked for the Board of Trade (1940–47). After the war, he 'settled down' to the life of a Cambridge don, first, as a Lecturer in the Faculty (1939–55), then as Director of the Department of Applied Economics (1955–70) and, finally, until his 'retirement' in 1980, as Professor of Political Economy (Marshall's chair) (1969–80).

Reddaway is a fine example of an applied economist in the tradition of Marshall and Keynes. He has one of the finest critical minds in the profession; he remorselessly reveals flaws in logic and ignorance of the nature and use of data alike. He is severely practical – the philosophical and speculative aspects of the discipline have little appeal for him and he has no use for theory for its own sake. Reddaway likes to be given questions – the effects of overseas investment on the UK economy (1967, 1968), the true incidence of SET (1970, 1973) – and he produces reports noted for their innovative approach and feel for orders of magnitude. He has also written a number of pioneering works, for example, his first book, on the Russian financial system (1935), his study of the economics of a declining population (1939), and his paper (1959) showing that job opportunities rather than relative wage movements were the main reason for the distribution of labour.

Reddaway's study of the Russian financial system includes an account of a tax system which will provide the appropriate price level to ensure the purchase of the residual production of consumption goods once the level of accumulation has been decided and given a socialist commitment to full employment of labour. While his analysis of the longer-run effects of a declining population is orthodoxly neoclassical – a higher standard of living emerges because of higher capital per person than otherwise would have been the case – his discussion of the relationship between population growth and the level of employment is an astute application of the then, very new Keynesian theory of employment. In addition the policy proposals of the concluding chapter read in a thoroughly modern manner.

Reddaway's work on the problems of developing nations is built up from first principles which are themselves founded in keen common-sense observations and empirical generalizations. The appendix to *The Development of the Indian Economy* (1962) on the importance of lags in the investment decision is one of his most significant insights. His policy recommendations are directed straight-forwardly to the problems in hand, always relevant if sometimes lacking a little in political nous.

Finally, no account of Reddaway's contributions would be complete without mention, first, of his regular column as 'academic investor' in the *Investors Chronicle* in which he reveals how both his college's and his own portfolios have fared; and, secondly, of the remarkable five years (1971–6) as joint editor of the *Economic Journal* with his lifelong friend from King's, David Champernowne.

Selected Works

1935. *The Russian financial system*. London: Macmillan.

1936. General theory of employment, interest and money (review). *Economic Record* 12(22): 28–36.

1939. *The economics of a declining population*. London: George Allen & Unwin.

1959. Wage flexibility and the distribution of labour. *Lloyds Bank Review* 13(54): 32–48.

1962. *The development of the indian economy*. London: George Allen & Unwin.

1967, 1986. *Effects of UK direct investment overseas*, 2 vols. (Vol. 1 with J.O.N. Perkins, S.J. Potter and C.T. Taylor; Vol. 2. with D.R. Glynn, J.D. Sugden, P.M. Croxford, C.H. Fletcher and J.S. O'Donnell.). Cambridge: Cambridge University Press.

1970, 1973. Effects of selective employment tax, 2 vols. Cambridge/London: Cambridge University Press/HMSO.

Redistribution of Income and Wealth

F. A. Cowell

Keywords

Education; Efficiency; Equity; Inequality; Multiple equilibria; Pensions; Price control; Progressive and regressive taxation; Redistribution of income and wealth; Social insurance

JEL Classifications

D3

The topic of redistribution is sometimes interpreted narrowly in rather dry terms: as the description and quantification of the simple fact of change in an income or wealth distribution. This can apply both to an actual change that takes place through time and also to the apparent alteration of the distribution at a point in time by taxes and transfers, and principally involves problems of measurement that are common to other fields of applied economics. However, redistribution can also be seen as a specific goal for economic policymakers: as such it is a subject of special interest in its own right. Sections "The Reason for Wanting to Redistribute", "The Objectives of Redistribution", "What Should Be Redistributed?", "The Available Instruments" below concentrate primarily on this second interpretation; some issues arising under the first interpretation are considered in section "The Effectiveness of the Policy".

The Reason for Wanting to Redistribute

Perhaps the simplest and most direct reason for wishing to see a redistribution of income, consumption or wealth in the community is simple fellow feeling on the part of the citizens of the community. This can be incorporated into the utilitarian approach to welfare judgements within the tradition of Bentham and Mill, in two ways.

One might suppose that judgements about distribution are made in a state of primordial ignorance about one's own position in the distribution: social aversion to inequality is thus rationalized as individual aversion to risk (Harsanyi 1955). Secondly, it might be supposed that the poor are made to feel worse off in their plight by the very knowledge that the well-to-do are well-to-do, and the rich are made to feel uncomfortable by the low living standards of the poor – see Hochman and Rodgers (1969). Thus the problems of inequality are rationalized within individual utilities as 'externalities' in a manner similar to health hazards from pollution. A weakness of this approach is that it puts a heavy burden on the particular configuration of individual preferences that happen to be present within a given community at a given moment: should one *really* only redistribute if enough citizens happen to feel upset by it? And what if some citizens *like* knowing that the very poor are very poor?

An alternative approach is to take the motivation for redistribution as a direct moral imperative – see Tawney (1965), Rawls (1971); improvement in the wellbeing of the disadvantaged is perceived as a social objective in its own right, along with other apparently desirable goals such as civil liberties and growth in national income.

The Objectives of Redistribution

Whatever the precise reasons for wishing to redistribute income or wealth may be, in broad terms the principal goals of redistribution policy can be stated very simply: the primary objectives are usually some goal of greater equity and of 'social insurance'; and as a secondary, though important, desideratum, one is usually also concerned with economic efficiency.

In order to examine these objectives in more detail two concepts need to be carefully distinguished: redistribution '*ex ante*' – the rearrangement of the structure of *income-earning opportunities* – and redistribution '*ex post*' – the reallocation of income or wealth that results from the economic processes of production and exchange, whatever individual opportunities may have been. In practice the two concepts may be difficult to disentangle since a policy measure that apparently just rearranges the prizes (such as an income-tax scheme) may also have repercussions on some people's *ex ante* opportunities (by, for example, affecting market wages); but both are relevant to a discussion of the relationship between equity and other goals.

In a very simplified model of the distribution of income, 'equity' can be expressed fairly easily: if one considers that the cake has been cut very unequally, then one sets about trying to even up the slices. But in a dynamic view of the economy where people make economic choices which affect their future incomes, the slices-of-a-fixed-size-cake analogy can be misleading, and the position may be further complicated when those choices have to be made in the face of uncertainty. Obviously the size of the national cake to be 'shared out' is not, in practice, fixed: individual incomes (and hence the total income in the community) are determined by the choices people make as to how much they work, save, and take entrepreneurial risks, and again the total stock of wealth obviously also depends on the rate at which people save. So the elementary equity question of who ought to get what cannot be divorced in practice from the issue of how individual incomes and wealth holdings are generated: efficiency considerations have to be taken into account in the pursuit of greater equity. There is a second, more subtle, difficulty: because of incompetence, ignorance or plain 'bad luck' people who may have looked alike in terms of their original economic opportunities turn out to be very dissimilar in terms of outcomes once a few rounds have been played of the great economic game that determines how much everybody actually gets. Hence there is a good case for a government concerned with distributional equity to pay attention to both the *ex ante* and the *ex post* concepts of redistribution (Hammond 1981).

For this reason an interest in social insurance is often taken to be a natural counterpart of a concern for equity. The public provision of protection against the slings and arrows of outrageous fortune is particularly important for those events for which conventional insurance markets are likely to give inadequate coverage, such as unemployment or ill health, for example – see Atkinson (1987). By filling such gaps social insurance may actually improve the efficient working of the economy. Besides this, social insurance can also apply to *ex post* redistribution that is intended to circumvent the otherwise unsatisfactory workings of some markets. For example, the markets for private insurance and savings might, under ideal circumstances, allow people to look after themselves effectively; but in practice problems such as imperfect information and the consequent rationing of insurance of credit to those people who are perceived to be good risks will mean that coverage is far from complete (Arrow 1985). Hence the provision of state pensions as a means of cushioning the possibly unfortunate effects of restrictions on savings by people of modest means.

What Should Be Redistributed?

Whether it is *income* (the flow of spending power during a given period) or *wealth* (the command over resources that a person may possess at a given point in time) that is to be redistributed depends to some extent on the precise definition of these terms (in particular the relevant period over which income is measured and the range of assets to be counted in as personal wealth) and also on the degree of importance that one attaches to *ex ante* or *ex post* concepts of redistribution. For example, some components of wealth (land, financial assets) may be regarded as part of the range of economic opportunities which results in the flow of spendable income. Again weekly money income might be more relevant than broader concepts of wealth or long-term income if one's primary concern is for redistribution to alleviate short-term need rather than to alter the structure of economic opportunities (Atkinson 1983, ch. 3).

However, the issue of what one ought to use in order to achieve the objectives of redistribution cited above raises further questions. One of the most important of these is whether one ought to redistribute income itself (which yields purchasing power over consumption goods) or rather the consumption goods directly. The standard answer provided by economists is that cash is unquestionably more effective, since it allows individuals to be the judges of what is best for their welfare and to make substitutions between different goods under varying market conditions in pursuit of that welfare: money to buy soup is supposedly more effective than the provision of soup kitchens. However, this conclusion is strictly relevant only if one imposes a number of stringent conditions, for example, the assumption that everyone has access to perfect market opportunities and accurate information on which to base his judgement in the market. It is invalid in the presence of multiple market equilibria (Foldes 1967). It ignores pressing requirements of crises such as war and famine: extreme circumstances may require direct intervention to act more swiftly and reliably to maintain living standards than the often capricious and sluggish movements of the 'invisible hand'.

The Available Instruments

Among the more obvious instruments available for *ex post* redistribution are taxes on income, wealth and the transfer of wealth via gifts and bequests, and transfer payments such as pensions and social-security benefits. However, it is not easy to draw a hard-and-fast line around the range of instruments that might be taken to be redistributive tools, particularly if one is concerned with description rather than prescription. There appears to be a good case in practice for including 'indirect' taxes (such as value added tax), subsidies and also those benefits 'in kind' which are bestowed on *particular* households or persons, since the impact of these items on personal spending power is usually fairly clear. This may, for example, be extended to include such goods as state-provided education. However the precise distributional impact of publicly provided goods that are really consumed *jointly* by the community (in which category we might include items such as public sanitation, the police services, or even national defence) is less easy to determine, but should not be assumed to be negligible.

As an alternative to raising taxes and the public provision of goods and services, a government wishing to redistribute real spending power may choose to intervene directly in the market mechanism. The most obvious example of this policy is price control. This term applies not only to rationing and the regulation of prices paid by consumers for goods – which can be an effective method of intervention to achieve redistribution – in emergencies such as wartime, but also to the control of prices that individuals receive for services that they may supply (for example, minimum wage legislation) or assets that they possess (control of house rents).

The instruments available for the purposes of *ex ante* redistribution (that is, the means of reorganizing the *opportunities* for creating income and accumulating wealth) are more disparate. One has the immediate problem that the range of policies considered to be available is strongly influenced by the economic philosophy which one considers to be relevant to the analysis and by the political and social system within the community. Take a prime example of this: education. There are many opinions on the potential for using this as a redistributional tool, some of which may be crudely summarized by the following three views: (a) it is a passport to higher positions on a ladder of economic opportunity whose rungs are pretty rigidly fixed, so that greater equality can be achieved simply by changing the method of issuing the passports; (b) it forms part of a complex of personal or family investment decisions, whereby intervention in the provision of education might upset the efficient allocation of the market mechanism without doing anything to alleviate the inequality of economic opportunity; (c) even if effective redistribution *could* be achieved in principle, substantial reorganization of educational opportunities is bound to be limited by what are seen as fundamental freedoms of choice. Note that the divergence of view concerns both economic role of education and the extent to which one is free to use it as an instrument of public policy (Le Grand 1982, ch. 4).

The Effectiveness of the Policy

Any attempt to quantify the effectiveness of redistribution policy has to surmount a number of extremely troublesome obstacles.

In the first place one has to confront the problem of ‘unequal inequalities’, which essentially arises from an attempt to compare intrinsically complex social states. Even if one puts this in elementary terms, whereby every person’s welfare is accurately measured by his or her income, a fundamental difficulty remains: apart from special circumstances – for example, a comparison involving a hypothetical state of perfect equality – the question of which of two distributions is the more unequal does not generally have a clear-cut answer. In practice, even a very successful redistribution policy will have diminished rather than completely eliminated real income differences, so that an assessment of the policy’s impact necessarily involves a comparison of the apparent change in inequality that has been achieved relative to the degree of inequality that would have obtained otherwise.

There is no single method for measuring such inequality changes that commands universal support, and hence no generally agreed measuring rod to ascertain the extent of redistribution under all circumstances (Cowell 1977; Foster 1985). One of the practical difficulties to which this gives rise is that it is difficult to be dogmatic about labelling policy instruments in terms of degrees of ‘progressivity’ (Lambert 1985). Moreover, in many cases redistribution may involve not just a narrowing (or indeed expansion) of income differentials, but also a *re-ranking* of income receivers within the pecking order so that, to quantify redistribution effectively, more is required than a simple measurement of the change in overall dispersion (Cowell 1985).

The second problem follows directly from this: who is to say what *would* have happened otherwise and, therefore, what change in inequality has actually been achieved? If one is merely concerned with the documentation of trends in the perceived inequities of income distribution through time, this may not be too difficult. But if at any moment of history one attempts to draw

R

the inference that 'according to our chosen inequality index, the inequality of disposable income would have been 20 per cent higher than it is now but for the high marginal tax rates on upper income groups', then one is making a much bolder assertion about how the underlying economic mechanisms are supposed to work. For the very presence of the instruments of redistribution policy will have influenced the choices people make about their jobs, business enterprises and savings, which in turn, can be expected to affect the resulting income distribution. The 'distribution before tax' – obtainable from a statistical office's published figures – cannot automatically be taken to be the same thing as the distribution *without* the tax – the income distribution that one might expect to see if the relevant redistribution instrument were to be abolished.

Some allowance for this problem is usually possible in the case of *ex post* redistribution instruments – for example, it is possible to estimate the likely repercussion on the supply of different types of labour that will arise because of the supplementation of some people's incomes by public transfers and the reduction in other people's incomes through taxation (Hausman 1985; Killingsworth 1983), or the impact on private savings of the presence of state-provided pensions and social insurance schemes (Danziger et al. 1981; Kotlikoff 1984). However, the allowance to be made for these feedback effects is usually quite sensitive to the particular model of household behaviour that is applied.

Despite these reservations, some broad conclusions are possible. Very narrowly based measures run the danger of the 'demarcation trap': for example, subsidizing particular commodities or taxing only certain forms of income or wealth may present some people with an incentive to change their behaviour, or even misrepresent their true status, so as to profit by the artificial distinctions drawn by the selective tax or subsidy scheme. The effectiveness of the measure may thereby be reduced and, even if this does not happen, the discrimination of the scheme may itself create substantial inequities by treating essentially similar people in different ways. On the other hand, very broadly based measures may scatter their shot so widely that much of it misses the target: blanket allowances or exceptions within income- or wealth-tax laws, and some broadly defined educational subsidies are often found to be regressive in their actual *ex post* impact on income and wealth. Finally it is usually the case that *taxes*, taken as a whole, turn out not to be very progressive in terms of their *ex post* impact whereas *transfers* usually are.

See Also

► Progressive and Regressive Taxation
► Social Insurance

Bibliography

Arrow, K.J. 1985. The economics of agency. In *Principals and agents: The structure of business*, ed. J. Pratt and R. Zeckhauser. Cambridge, MA: Harvard Business School Press.

Atkinson, A.B.. 1983. *The economics of inequality*. 2nd ed. Oxford: Oxford University Press.

Atkinson, A.B.. 1987. Income maintenance and social insurance: A survey. In *Handbook of public economics*, ed. A.J. Auerbach and M.S. Feldstein, vol. 2. Amsterdam: North-Holland.

Cowell, F.A. 1977. *Measuring inequality*. Oxford: Philip Allan.

Cowell, F.A. 1985. Measures of distributional change: An axiomatic approach. *Review of Economic Studies* 52: 135–151.

Danziger, S., R.H. Haveman, and R. Plotnick. 1981. How income transfer programs affect work, savings and the income distribution: A critical review. *Journal of Economic Literature* 19: 975–1028.

Foldes, L.P. 1967. Income redistribution in money and in kind. *Economica* 34: 30–41.

Foster, J. 1985. Inequality measurement. In *Fair allocation*, ed. H.P. Young. Providence: American Mathematical Society.

Hammond, P.J. 1981. *Ex-ante* and *ex-post* welfare economics. *Economica* 48: 235–250.

Harsanyi, J.C. 1955. Cardinal welfare, individualist ethics and interpersonal comparisons of utility. *Journal of Political Economy* 63: 309–321.

Hausman, J.A. 1985. Taxation and labour supply. In *Handbook of public economics*, ed. A.J. Auerbach and M.S. Feldstein, vol. 1. Amsterdam: North-Holland.

Hochman, J.M., and J.D. Rodgers. 1969. Pareto optimal redistribution. *American Economic Review* 59: 542–557.
Killingsworth, M.R. 1983. *Labour supply*. Cambridge: Cambridge University Press.
Kotlikoff, L.J. 1984. Taxation and savings: A neoclassical perspective. *Journal of Economic Literature* 22: 1576–1629.
Lambert, P.J. 1985. *Tax-progressivity: A survey of the literature*, Working Paper. London: Institute for Fiscal Studies.
Le Grand, J. 1982. *The strategy of equality*. London: Allen & Unwin.
Rawls, J. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.
Tawney, R.H. 1965. *Equality*. London: Allen & Unwin.

Redlining

Ethan Cohen-Cole

Keywords
Credit; Discrimination; Lending; Race; Redlining; Scoring

History

Redlining is the practice of restricting or denying access to services in a spatially defined area. Typically, redlining refers to the practice of restricting access to financial service products, such as mortgages, to residents of minority areas. It is widely assumed that racially discriminatory practices had long existed in financial services – the first written evidence appeared as a result of the Home Owners Act of 1933. It created the Home Owners' Loan Corporation (HOLC), which famously created colour-coded maps that designated risks associated with mortgage lending; among the various criteria for risk coding was the racial composition of the neighbourhood. The methods used by the HOLC quickly spilled over into commercial banks and the Federal Housing Administration (FHA). The red lines that marked black (and to some degree poor white) neighbourhoods led to the term *redlining*.

As a result, only a tiny fraction of mortgages offered by HOLC went to black people. To understand the significance of this discrimination, HOLC refinanced distressed loans, and refinanced nearly 20% of the housing stock at terms that were much better than available commercially. For example, HOLC would require only 20% down payments versus 50% or 60%. It similarly offered term mortgages of up to 18 years rather than 3–5 years offered by banks.

A variety of research (examples include Jackson (1985), Squires et al. (1987), Squires et al. (1987)), Squires et al. (1991), Taggart and Smith (1981), Immergluck (2002) and Schill and Wachter (1993)) has credited redlining with the decline of minority urban areas. Decades of allocation of private and government credit until the 1960s away from redlined areas meant that historically Black and urban areas around the country fell into disrepair.

In the late 1960 and early 1970s, Congress acted to undo these policies. The Fair Housing Act of 1968 prohibited housing discrimination; then, to ensure compliance with the Fair Housing Act, Congress passed the Home Mortgage Disclosure Act of 1975, requiring the release of loan underwriting information that included the location of the mortgage and the race of the borrower. Reluctance by lenders to lend in these areas led to pressure by urban activist groups. By 1977, Congress had responded to this pressure by passing the Community Reinvestment Act of 1977 (CRA), meant to ensure that banks lend in areas where they accept deposits.

Nonetheless, research suggests that in spite of the regulations, redlining continued to occur. In the 1990s, the growing availability of data led to a new round of research that found that mortgage redlining persisted. Tootell (1996), Munnell et al. (1996) and Ross and Tootell (2004) all find the continued presence of discrimination in lending. Ross and Yinger (2002), Hillier (2003) and Cohen-Cole (2010) provide overviews of this line of research.

R

Redlining, Credit Scores and Inference

The principal method used in the literature to infer the presence of redlining is to estimate the coefficient of a race variable in a regression of an individual credit approval decisions. The goal of the studies is to control for the variables used by lenders in their decisions and then infer whether race explaining any remaining variation in lending decisions. While most studies of these have concentrated on mortgages, the methodological issues faced are instructive for all forms of credit. In principle, a lender's approval decision should be some function of a borrower's ability to repay the loan. Lenders' internal models are typically binary ones, and as a result, most studies have used discrete choice models:

$$\Pr(\text{Approval} = 1) = f(\text{X}1_i) \tag{1}$$

where $X1_i$ is a vector of individual credit characteristics, where i indexes individuals. These variables should include any credit characteristic that is reasonably correlated with a borrower's ability to repay the loan. This could include prior history of defaults, amount of other credit available, utilisation patterns etc. Most lenders will create a summary statistic, called a credit score, based on these characteristics. Specifically, credit scores are an inverse ordinal ranking of the probability of default.

We can now rewrite (1) above as:

$$\Pr(\text{Approval} = 1) = f(\text{Credit score}) \tag{2}$$

Notice that some information in (1) may be lost in the use of (2) alone. The credit score is typically created using a linear probability model in which the components of $X1_{\{i\}}$ are additively separable. This simple summary statistic can capture most, but not all, of the variation in default events. Because lenders do not have immediate access to a potential borrower's entire credit history, they may prefer to use the credit score alone in lending decisions. Notice that this can lead to inference problems in studies of redlining. If the lender uses the full information set on the borrower, $X1_{\{i\}}$ but the econometrician uses only Credit score$_i$, the residual variation may be correlated with borrower race. The econometrician could then incorrectly attribute the positive coefficient on the race variables as evidence of discrimination, when it simply reflects the inability of the credit score to capture racial differences in credit behaviour. As a result, most studies have used the full set of covariates.

Of course, individual characteristics such as income and job status may also be relevant to the ability to repay (income and job information is not included in most credit scores). With these included, a typical redlining study may specify that approval is a linear function of variables that predict whether a person will default and variables designed to measure discrimination.

$$\begin{aligned}&\Pr(\text{Approval} = 1)\\&= f\Big(\langle b\rangle^0 + \langle b\rangle^1 \text{black}_{\{i\}} + \langle b\rangle^2 X1_{\{i\}} + \langle b\rangle^3 \text{X}2_{\{i\}}\\&+ \langle b\rangle^4 \text{percentblack}_{\{j\}} + \langle b\rangle^5 Y_{\{j\}}\\&+ \langle e\rangle_{\{j\}}\end{aligned} \tag{3}$$

Where $X1_{\{i\}}$ are again the set of credit history and characteristics, $X2_{\{i\}}$ is a vector of non-credit related individual characteristics, and $Y_{\{j\}}$ is a set of regional or local characteristics, with j an index of some geographic area. Local variables are included to control for a lender's non-decisions to allocate lending according to local characteristics other than race. Finally, the variables black$_{\{i\}}$ and percentblack$_{\{j\}}$ refer to a variable indicating that the applicant is black and a variable measuring the percentage of black individuals in neighbourhood j, respectively. Then one typically evaluates the significance of the $\langle \text{b}\rangle^1$ or $\langle \text{b}\rangle^4$ coefficients.

A negative coefficient on $\langle \text{b}\rangle^1$ indicates that a black applicant is less likely to receive a loan. Probably the most prominent of these analyses (Munnell et al. 1996), later dubbed the 'Boston Fed Study', found a negative coefficient on $\langle \text{b}\rangle^1$ that was robust to a myriad of specifications. This paper (as well as Tootell 1996) used individual-level transaction data from the Home Mortgage Disclosure Act (HMDA) along with census tract information and credit histories to show evidence of disparities in access to mortgages. The study finds that, conditional on applying for a mortgage, the probability of receiving credit is lower for blacks than for whites.

A negative coefficient on $\langle b \rangle^4$ indicates that an applicant who lives in a black neighbourhood is less likely to receive a loan. Cohen-Cole (2010) finds evidence of this phenomenon in credit card lending. Because many credit card applications are taken by mail or by phone, they are typically issued based on less information than a mortgage and without direct knowledge of the borrower's race. Negative coefficients on either $\langle b \rangle^1$ or $\langle b \rangle^4$ have been labelled as redlining. The latter is a direct restriction of credit based on a spatially defined area. The former is based on the applicant's race; however, because redlining historically targeted black areas, and because it also prevented black borrowers from buying houses in white neighbourhoods, redlining has become synonymous with discrimination in lending based on race.

The negative coefficient can be viewed in a couple of ways. The first is based on Becker's (1971) argument that some individuals have a 'taste' for discrimination. In Becker's formulation, this is costly to the individual and is minimised by competition. The second is the argument that equilibrium phenomena (such as supply differences by group or location) may occur even with *ex ante* identical groups. Asymmetries can arise based on very minor differences in preferences (Schelling 1972), based on incentives to specialise (Moro and Norman 2004; Coate and Loury 1993), or based on differences in information precision related to collateral valuation (Lang and Nakamura 1993). One can explain this type of phenomenon in the mortgage context as follows: if applicant choices (e.g. whether to apply) are correlated with their own credit quality and with race, then this can lead to correlations in the lender's applicant pool between race and creditworthiness. Applicant actions serve as an informative signal to lenders that can then be used for credit decisions. As a result, one could observe disparities in approval rates across races even if each lending decision is unbiased with respect to race. Notice that this can occur even in the absence of an omitted variables problem.

These phenomena reflect the presence of profit-seeking-based statistical lending or marketing criteria that lead, *ex post*, to differences in access by race.

Other Forms of Credit

Once the pervasive impacts of mortgage redlining were realised, researchers looked into lender practices in other areas. In a series of papers and books, Squires (Squires et al. 1987, 1991; Squires 1997) has found evidence of redlining in consumer insurance products. Similar evidence is found in auto lending (Charles et al. 2008), and small business credit (Bostic and Lampini 1999; Blanchard eet al. 2008). While these literatures are less developed, the finding of redlining in these areas is indicative of the pervasiveness of the issue.

Credit Cards and Building Credit

So why are other forms of credit important? As alluded to above, other types of credit, in particular credit cards, are an integral part of the consumer finance experience and form the building blocks of a consumer's ability to access credit in the future. Because credit scores are based on the average probability of default of others, individuals with little or no credit record of repayment will have low scores and thus a more difficult time obtaining credit.

For every credit card, auto loan, mortgage, or other product that a consumer has, the amount of credit and whether or not it has been paid on time will be reported to a credit bureau. There are three large such bureaus to which all large and the vast majority of small credit issuers provide data. This matters to consumers because any new credit (and some old credit) is evaluated based on the information held at these agencies. Pay credit card balances on time and your credit 'score' will be higher than if you had not. With a higher score, a consumer will have more access to additional credit. As such, a consumer who wishes to buy a house, but has never had a credit card, or has failed to pay cards on time, will be unlikely to be approved for a mortgage. Or, if they are approved, will likely be charged a higher interest rate on the loan (see Edelberg (2007) for a recent study that found disparities in the terms of loans by race).

Because of the need to 'build' credit, the ability to obtain credit cards or other entry products is a crucial step in determining one's ability to secure a mortgage or car loan in the future. This phenomenon is particularly important if redlining exists for entry products; if minorities are discriminated against in obtaining credit cards, the ones that obtain them will be a disproportionately better credit risk. Cohen-Cole (2010) finds redlining in credit card lending. Thus the mortgage regressions above will be biased *against* a finding of discrimination; when the race coefficient is significant, the discrimination is particularly large.

Zipcode Redlining and Consumer Behaviour

In spite of legal challenges and public condemnation, redlining has persisted even to the present. Notably, the nature of the practice has changed over time. In part because of legal challenges, the historical practice of drawing red lines around minority neighbourhoods is no longer common. In its place has emerged a range of new techniques. These techniques serve one of two purposes. One, for institutions or individuals that wish to live by the letter of the law, but not its intent, new statistical techniques can avoid using the presence of minorities as a lending criterion, but capture similar credit risks. For example, a lender may observe that individuals who live in locations with high vacancy rates tend to default more often. The fact that more minorities live in this area is not considered.

Two, lenders that wish to continue using race itself as a criterion, can search for variables that are correlated with minority presence. For example, if minorities live in zip codes that also have high vacancy rates, a lender can condition lending on the regional vacancy rates. This avoids using race directly in the lending decision, but achieves the same outcome.

Notice that *ex post*, from a statistical perspective, the two are identical. Cohen-Cole (2010) as well as the public press (see Harney (2008) and others) have found evidence that lenders use zip-code characteristics to determine credit supply. Nonetheless, no lenders would claim that race played a role in the zip-code risk ratings.

A second method is the use of consumer behaviour such as shopping patterns to adjust credit lines. In a widely reported case (Lieber 2009), American Express admitted that they had adjusted the credit line of a borrower based on his shopping patterns. Lenders notice that the use of various stores is associated with higher default probabilities. The press noted that the stores were frequented principally by minorities. If an individual shops at these locations, they may have their credit lines reduced or eliminated. Again, there is an inference problem. Either American Express searched for locations that were in locations with minority customers *or* they found stores that had customers with high default probabilities and these coincidentally had minority customers.

Regardless of the intent, the outcome of these lending techniques is the same as overt redlining.

See Also

▶ Anti-discrimination Law
▶ Credit Card Industry

Bibliography

Becker, G.S. 1971. *The economics of discrimination*. Chicago: University of Chicago Press.

Blanchard, L., B. Zhao, and J. Yinger. 2008. Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics* 63: 467–497.

Charles, K.K., E. Hurst, and M. Stephens. 2008. Rates for vehicle loans: Race and loan source. *American Economic Review: Papers & Proceedings* 98: 315–320.

Coate, S., and G.C. Loury. 1993. Will affirmative action policies eliminate negative stereotypes? *American Economic Review* 83(5): 1220–1240.

Cohen-Cole, E. 2010. Credit card redlining. *Review of Economics and Statistics* (forthcoming).

Edelberg, W. M. 2007. Racial dispersion in consumer credit interest rates. *Finance and Economics Discussion Series* 2007–28.
Harney, K. R. 2008. Zip Code 'redlining': A sweeping view of risk. Washington Post, 2 February. http://www.washingtonpost.com/wp-dyn/content/article/2008/02/01/AR2008020101680.html.
Hillier, A.E. 2003. Spatial analysis of historical redlining: A methodological explanation. *Journal of Housing Research* 14: 137–168.
Immergluck, D. 2002. Redlining redux: Black neighborhoods, black-owned firms, and the regulatory cold shoulder. *Urban Affairs Review*, 22–41.
Jackson, K.T. 1985. *Crabgrass frontier: The suburbanization of the United States*. New York: Oxford University Press.
Lang, W., and L. Nakamura. 1993. A model of redlining. *Journal of Urban Economics* 33(2): 223–224.
Lieber, R. 2009. American Express kept a (very) watchful eye on charges. New York Times, 30 January. http://www.nytimes.com/2009/01/31/your-money/credit-and-debit-cards/31money.html?_r%E2%80%89=%E2%80%892%26ref%E2%80%89=%E2%80%89business%26pagewanted%E2%80%89=%E2%80%89all.
Moro, A., and P. Norman. 2004. A general equilibrium model of statistical discrimination. *Journal of Economic Theory* 114(1): 1–30.
Munnell, A.H., G.M.B. Tootell, L.E. Browne, and J. McEneaney. 1996. Mortgage lending in Boston: Interpreting the HMDA data. *American Economic Review* 86: 25–53.
Ross, S., and G.M.B. Tootell. 2004. Redlining, the community reinvestment act, and private mortgage insurance. *Journal of Urban Economics* 55: 278–297.
Ross, S.L., and J. Yinger. 2002. *The color of credit: Mortgage discrimination, research methodology, and fair-lending enforcement*. Cambridge, MA: MIT Press.
Schelling, T.C. 1972. A process of residential segregation: Neighborhood tipping. In *Racial discrimination in economic life*, ed. A. Pascal. Lexington: Lexington Books.
Schill, M., and S.M. Wachter. 1993. A tale of two cities: Racial and ethnic geographic disparities in home mortgage lending in Boston and Philadephia. *Journal of Housing Research* 4(2): 245–275.
Squires, G.D. (ed.). 1997. *Insurance redlining: Disinvestment, reinvestment, and the evolving role of financial institutions*. Washington, DC: Urban Institute Press.
Squires, G.D., L. Bennett, K. McCourt, and P. Nyden. 1987. *Chicago: Race, class, and the response to urban decline*. Philadelphia: Temple University Press.
Squires, G., W. Velez, and K.E. Taeuber. 1991. Insurance redlining, agency location, and the process of urban disinvestment. *Urban Affairs Review* 26(4): 567–588.
Taggart, H., and K.W. Smith. 1981. Redlining. *Urban Affairs Review* 17(1): 91–107.
Tootell, G.M.B. 1996. Redlining in Boston: Do mortgage lenders discriminate against neighborhoods? *Quarterly Journal of Economics* 111: 1049–1079.

Reduced Rank Regression

Søren Johansen

Abstract

The reduced rank regression model is a multivariate regression model with a coefficient matrix with reduced rank. The reduced rank regression algorithm is an estimation procedure which estimates the reduced rank regression model. It is related to canonical correlations and involves calculating eigenvalues and eigenvectors. We give a number of different applications to regression and time series analysis, and show how the reduced rank regression estimator can be derived as a Gaussian maximum likelihood estimator. We briefly mention asymptotic results.

Keywords

Instrumental variable estimation; Limited information maximum likelihood; Maximum likelihood; Reduced rank regression

JEL Classifications

C10; C13

Reduced rank regression is an explicit estimation method in multivariate regression that takes into account the reduced rank restriction on the coefficient matrix.

Reduced rank regression model: We consider the multivariate regression of Y on X and Z of dimension p, q, and k, respectively: $Y_t = \Pi X_t + \Gamma Z_t + \varepsilon_t$, $t = 1, \ldots , T$. The hypothesis that Π has reduced rank less than or equal to r is expressed as $\Pi = \alpha\beta'$, where α is $p \times r$, and β is $q \times r$, where $r < \min(p,q)$, and gives the reduced rank model

$$Y_t = \alpha\beta' X_t + \Gamma Z_t + \varepsilon_t, t = 1, \ldots, T. \quad (1)$$

Reduced rank regression algorithm: In order to describe the algorithm, which we call $RRR(Y, X|Z)$,

R

we introduce the notation for product moments $S_{yx} = T^{-1}\sum_{t=1}^{T} Y_t X_t'$, $S_{yx.z} = S_{yx} - S_{yz}S_{zz}^{-1}S_{zx}$, and so on. The algorithm consists of the following steps:

1. First, regress Y and X on Z and form residuals $(Y|Z)_t = Y_t - S_{yz}S_{zz}^{-1}Z_t, (X|Z)_t = X_t - S_{xz}S_{zz}^{-1}Z_t$ and product moments and so on.

$$\begin{aligned} S_{yx.z} &= T^{-1}\sum_{t=1}^{T}(Y|Z)_t\,(X|Z)_t' \\ &= S_{yx} \; - S_{yz}S_{zz}^{-1}S_{zx}, \end{aligned}$$

2. Next, solve the eigenvalue problem

$$|\lambda S_{xx.z} - S_{xy.z}S_{yy.z}^{-1}S_{yx.z}| = 0, \qquad (2)$$

where $|.|$ denotes determinant. The ordered eigenvalues are $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_q)$ and the eigenvectors are $V = (v_1,\ldots,v_q)$, so that $S_{xx.z}V\Lambda = S_{xy.z}$ $S_{xx.z}^{-1}S_{xy.z}V$, and V is normalized so that $V'\,S_{xx.z}V = I_p$ and $V'S_{yx.z}S_{xx.z}^{-1}S_{xy.z}V = \Lambda$. The singular value decomposition provides an efficient way of implementing this procedure; see Doornik and O'Brien (2002).

3. Finally, define the estimators

$$\hat{\beta} = (v_1,\ldots,v_r)$$

together with $\hat{\alpha} = S_{yx.z}\hat{\beta}$, and $\hat{\Omega} = S_{yy.z} - S_{yx.z}\hat{\beta}\left(\hat{\beta}'S_{xx.z}\hat{\beta}\right)^{-1}\hat{\beta}'S_{xy.z}$. Equivalently, once $\hat{\beta}$ has been determined, $\hat{\alpha}$ and $\hat{\Gamma}$ are determined by regression.

The technique of reduced rank regression was introduced by Anderson and Rubin (1949) in connection with the analysis of limited information maximum likelihood and generalized to the reduced rank regression model (1) by Anderson (1951). An excellent source of information is the monograph by Reinsel and Velu (1998), which contains a comprehensive survey of the theory and history of reduced rank regression and its many applications.

Note the difference between the unrestricted estimate $\hat{\prod}_{OLS} = S_{yx.z}S_{xx.z}^{-1}$ and the reduced rank regression estimate $\hat{\prod}_{RRR} = S_{yx.z}\hat{\beta}\left(\hat{\beta}'S_{xx.z}\hat{\beta}\right)^{-1}\hat{\beta}'$ of the coefficient matrix to X.

Applications of the Reduced Rank Model and Algorithm

The reduced rank model (1) has many interpretations depending on the context. It is obviously a way of achieving fewer parameters in the possibly large $p \times q$ coefficient matrix Π. Another interpretation is that, although X is needed to explain the variation of Y, in practice only a few, r, factors are needed as given by the linear combinations $\beta'X$ in (1).

Restrictions on Π: Anderson (1951) formulated the problem of estimating Π under $p - r$ unknown restrictions '$\ell'\Pi = 0$. In (1) these are given by the matrix $\ell = \alpha_\perp$, that is, a $p \times (p - r)$ matrix of full rank for which $\alpha_\perp'\alpha = 0$. The matrix $\alpha_\perp$ is estimated by solving the dual eigenvalue problem $|\lambda S_{yy.z} - S_{yx.z}S_{xx.z}^{-1}S_{xy.z}| = 0$, which has eigenvalues Λ and eigenvectors W, and the estimate is $\hat{\alpha}_\perp = \left(w_{r+1},\ldots,w_p\right)$. If $p = q$, we can choose $W = S_{yy.z}^{-1}S_{yx:z}V\Lambda^{-1/2}$.

Canonical correlations: Reduced rank regression is related to canonical correlations (Hotelling 1936). This is most easily expressed if $p = q$, where we find

$$\begin{pmatrix} W & 0 \\ 0 & V \end{pmatrix}'\begin{pmatrix} S_{yy.z} & S_{yx.z} \\ S_{xy.z} & S_{xx.z} \end{pmatrix}\begin{pmatrix} W & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} I_p & \Lambda^{1/2} \\ \Lambda^{1/2} & I_q \end{pmatrix}.$$

This shows that the variables $W'Y$ and $V'X$ are the empirical canonical variates.

Instrumental variable estimation: Let the variables U, V, and X be of dimension p, q, and k respectively with $k \geq q$. Assume that they are jointly Gaussian with mean zero and variance Σ, and that $E(U - \gamma'V)X' = 0$, so that X is an instrument for estimating γ. This means that $\Sigma_{ux} = \gamma'\Sigma_{vx}$, so that

$$E\left(\begin{pmatrix} U \\ V \end{pmatrix} | X\right) - \begin{pmatrix} \gamma' \sum_{vx} \\ \sum_{vx} \end{pmatrix} \sum_{xx}^{-1} X = \begin{pmatrix} \gamma' \\ I_q \end{pmatrix} \sum_{vx} \sum_{xx}^{-1} X = \alpha\beta'$$

It follows that the $(p + q) \times k$ coefficient matrix in a regression of $Y = (U', V')'$ on X has rank q. Thus a reduced rank regression of Y on X is an algorithm for estimating the parameter of interest γ using the instruments X. This is the idea in Anderson and Rubin (1949) for the limited information maximum likelihood estimation.

Non-stationary time series: The model

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Gamma \Delta Y_{t-1} + \varepsilon_t, t = 1, \ldots, T \quad (3)$$

determines a multivariate times series Y_t, and the reduced rank of $\alpha\beta'$ implies non-stationarity of the time series. Under suitable conditions (see Johansen 1996) Y_t is non-stationary and ΔY_t and $\beta' Y_t$ are stationary. Thus Y_t is a cointegrated time series; see Engle and Granger (1987).

Common features: Engle and Kozicki (1993) used model (3) and assumed reduced rank of the matrix $(\alpha, \Gamma) = \xi\eta'$, so that $\Delta Y_t = \xi\eta'\left(Y'_{t-1}\beta, \Delta Y'_{t-1}\right)' + \varepsilon_t$. In this case $\xi'_\perp \Delta Y_t = \xi'_\perp \varepsilon_t$ determines a random walk, where the common cyclic features have been eliminated.

Prediction: Box and Tiao (1977) analysed the model $Y_t = \Pi Y_{t-1} + \Gamma Y_{t-2} + \varepsilon_t$, and asked which linear combinations of the current values, $v' Y_t$, are best predicted by a linear combination of the past (Y_{t-1}, Y_{t-2}), and hence introduced the analysis of canonical variates in the context of prediction of times series.

The Gaussian Likelihood Analysis

If the errors ε_t in model (1) are i.i.d. Gaussian $N_p(0, \Omega)$, and independent of $(X_s, Z_s, s \leq t\}$, the (conditional or partial) Gaussian likelihood is

$$-\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - \alpha\beta' X_t - \Gamma Z_t)' \Omega^{-1}(Y_t - \alpha\beta' X_t - \Gamma Z_t).$$

Anderson (1951) introduced the *RRR* algorithm as a calculation of the maximum likelihood estimator of $\alpha\beta'$. The Frisch–Waugh theorem shows that one can partial out the parameter Γ by regression, as in the first step of the algorithm. We next regress $(Y|Z)$ on $(\beta' X|Z)$ and find estimates of α and Ω, and the maximized likelihood function as functions of β:

$$\begin{aligned} \hat{\alpha}(\beta) &= S_{yx.z}\beta(\beta' S_{xx.z}\beta)^{-1}, \\ \hat{\Omega}(\beta) &= S_{yy.z} - S_{yx.z}\beta(\beta' S_{xx.z}\beta)^{-1}\beta' S_{xy.z}, \\ L_{\max}^{-2}(\beta) &= |\hat{\Omega}(\beta)|. \end{aligned} \quad (4)$$

The identity $\left|\begin{pmatrix} S_{yy.z} & S_{yx.z}\beta \\ \beta' S_{yx.z} & \beta' S_{xx.z}\beta \end{pmatrix}\right| = |S_{yy.z}||\beta' S_{xx.z}\beta - \beta' S_{xy.z} S_{yy.z}^{-1} S_{yx.z}\beta| = |\beta' S_{xx:z}\beta||S_{yy.z} - S_{yx.z}\beta (\beta' S_{xx.z}\beta)^{-1}\beta' S_{xy.z}| = |\beta' S_{xx.z}\beta||\hat{\Omega}(\beta)|$ shows that $L_{\max}^{-2/T}(\beta) = |S_{yx.z}||\beta'\left(S_{xx.z} - S_{yx.z}S_{yy.z}^{-1}S_{yx.z}\right)\beta|/|\beta' S_{xx.z}\beta|$ so that β has to be chosen to minimize this. Differentiating with respect to β we find that β has to satisfy the relation $S_{xx.z}\beta = S_{xy.z} S_{yy.z}^{-1} S_{yx.z}\beta\xi$ for some $r \times r$ matrix ξ. This shows (see Johansen 1996) that the space spanned by the columns of β, is spanned by r of the eigenvectors of (2), and hence, that choosing the largest λ_i gives the smallest value $L_{\max}^{-1/T}(\beta)$, so that

$$\hat{\beta} = (v_1, \ldots, v_r), L_{\max}^{-2/T}\left(\hat{\beta}\right) = |S_{yy.z}|\prod_{i=1}^{r}(1 - \lambda_i).$$

Hypothesis testing: The test statistic for rank of Π can be calculated from the eigenvalues because the eigenvalue problem solves the maximization of the likelihood for all values of r simultaneously. The likelihood ratio test statistic for the hypothesis rank$(\Pi) \leq r$, as derived by Anderson (1951), is

$$-2\log LR(\text{rank}(\Pi) \leq r) = T\sum_{i=r+1}^{\min(p,q)}\log(1 - \lambda_i). \quad (5)$$

Bartlett (1938) suggested using this statistic to test that r canonical correlations between Y and X were zero and hence that Π had reduced rank.

The simplest hypothesis to test on β is $\beta = H\varphi$. We can estimate β under this restriction by $RRR(Y, H'X|Z)$ and therefore calculate the likelihood ratio statistic using reduced rank regression. If, on the other hand, we have restrictions on the individual vectors $\beta = (H_1\varphi_1,\ldots, H_r\varphi_r)$, then reduced rank regression does not provide a maximum likelihood estimator, but we can switch between reduced rank regressions as follows. For fixed $\beta_1,\ldots, \beta_{i-1}, \beta_{i+1},\ldots, \beta_r$, we can find an estimator for ϕ_i and hence $\beta_i = H_i\varphi_i$ by

$$RRR\left(Y, H_i'X|\,Z, \left(\beta_1, \ldots\beta_{i-1}, \beta_{i+1},\ldots, \beta_r\right)'X\right).$$

By switching between the vectors in β, we have an algorithm which is useful in practice and which maximizes the likelihood in each step.

A switching algorithm: Another algorithm, that is useful for this model, is to consider the first order condition for β, when Γ has been eliminated, which has solution

$$\hat{\beta}(\alpha, \Omega) = S_{xx.z}^{-1}S_{xy.z}\Omega^{-1}\alpha\left(\alpha', \Omega^{-1}\alpha\right)^{-1}.$$

Combining this with (4) suggests a switching algorithm, as follows.

First choose some initial estimator $\hat{\beta}_0$, then switch between estimating α and Ω for fixed β by least squares, and estimating β for fixed α and Ω by generalized least squares.

This switching algorithm maximizes the likelihood function in each step and any limit point will be a stationary point. It seems to work well in practice. There are natural hypotheses one can test in the reduced rank model, like general linear restrictions on β, which are not solved by the reduced rank regression algorithm, whereas the above algorithm can be modified to give a solution.

Asymptotic distributions in the stationary case: The asymptotic distributions of the estimators and test statistics can be described under the assumption that the process (Y_t, X_t, Z_t) is stationary with finite second moments. It can be shown that estimators are asymptotically Gaussian and test statistics for hypotheses both for rank and for β are asymptotically χ^2, see Robinson (1973).

Asymptotic distributions in the non-stationary case: If the processes are non-stationary a different type of asymptotics is needed. As an example consider model (3) for $I(1)$ variables. When discussing the asymptotic distribution of the estimators, the normalization by $\hat{\beta}'S_{xx.z}\hat{\beta} = I_r$ is not convenient, and it is necessary to identify the vectors differently.

One can then prove (see Johansen 1996), that the estimates of α, Γ and Ω are asymptotically Gaussian and have the same limit distribution as if β were known: that is, the asymptotic distribution they have in the regression of ΔY_t on the stationary variables $\beta' Y_{t-1}$ and ΔY_{t-1}.

The asymptotic distribution of $\hat{\beta}$ is mixed Gaussian, where the mixing parameter is the (random) limit of the observed information. Therefore, by normalizing on the observed information, we obtain asymptotic χ^2 inference for hypotheses on β.

The limit distribution of the likelihood ratio test statistic for rank, see (5), is given by a generalization of the Dickey–Fuller distribution:

$$DF_{p-r} = \mathrm{tr}\left\{\int_0^1 (dW)W'\left(\int_0^1 WW'\mathrm{d}u\right)^{-1}\int_0^1 W(dW)'\right\},$$

where W is a standard Brownian motion in $p - r$ dimensions. The quantiles of this distribution can at present only be calculated by simulation if $p - r > 1$. The limit distribution has to be modified if deterministic terms are included in the model.

See Also

- Instrumental Variables
- Maximum Likelihood

Bibliography

Anderson, T.W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22: 327–351.

Anderson, T.W., and H. Rubin. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20: 46–63.
Bartlett, M.S. 1938. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society* 34: 33–40.
Box, G.E.P., and G.C. Tiao. 1977. A canonical analysis of multiple time series. *Biometrika* 64: 355–365.
Doornik, J.A., and R.J. O'Brien. 2002. Numerically stable cointegration analysis. *Computational Statistics and Data Analysis* 41: 185–193.
Engle, R.F., and C.W.J. Granger. 1987. Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55: 251–276.
Engle, R.F., and S. Kozicki. 1993. Testing for common factors (with comments). *Journal of Business Economics and Statistics* 11: 369–378.
Hotelling, H. 1936. Relations between two sets of variables. *Biometrika* 28: 321–377.
Johansen, S. 1996. *Likelihood based inference on cointegration in the vector autoregressive model*. Oxford: Oxford University Press.
Reinsel, G.C., and R.P. Velu. 1998. *Multivariate reduced rank regression. Lecture notes in statistics*. New York: Springer.
Robinson, P.M. 1973. Generalized canonical analysis for time series. *Journal of Multivariate Analysis* 3: 141–160.

Regime Switching Models

James D. Hamilton

Abstract

If the parameters of a time-series process are subject to change over time, then a full description of the data-generating process must include a specification of the probability law governing these changes, for example, postulating that the parameters evolve according to the realization of an unobserved Markov chain. This article describes classical and Bayesian algorithms for estimation and inference in such models and discusses some of the issues that arise in particular cases such as GARCH and state-space models.

Keywords

ARMA models; Asset prices; Econometrics; GARCH models; Gaussian densities; Gibbs sampler; Kalman filter; Markov chain Monte Carlo methods; Markov processes; Maximum likelihood; Numerical optimization methods in economics; Regime-switching models; State-space models; Vector autoregressions

JEL Classifications

C1

Many economic time series occasionally exhibit dramatic breaks in their behaviour, associated with events such as financial crises (Jeanne and Masson 2000; Cerra and Saxena 2005; Hamilton 2005) or abrupt changes in government policy (Hamilton 1988; Sims and Zha 2006; Davig 2004). Of particular interest to economists is the apparent tendency of many economic variables to behave quite differently during economic downturns, when underutilization of factors of production rather than their long-run tendency to grow governs economic dynamics (Hamilton 1989; Chauvet and Hamilton 2006). Abrupt changes are also a prevalent feature of financial data, and the approach described below is quite amenable to theoretical calculations for how such abrupt changes in fundamentals should show up in asset prices (Ang and Bekaert 2002a, b; Garcia et al. 2003; Dai et al. 2003).

Consider how we might describe the consequences of a dramatic change in the behaviour of a single variable y_t. Suppose that the typical historical behaviour could be described with a first-order autoregression,

$$y_t = c_1 + \varphi y_{t-1} + \varepsilon_t, \tag{1}$$

with which seemed to adequately describe the observed data for $t = 1; 2;\ldots, t_0$. Suppose that at date t_0 there was a significant change in the average level of the series, so that we would instead wish to describe the data according to

$$y_t = c_2 + \varphi y_{t-1} + \varepsilon_t \tag{2}$$

for $t = t_0 + 1; t_0 + 2,\ldots$This fix of changing the value of the intercept from c_1 to c_2 might help

R

the model to get back on track with better forecasts, but it is $\varepsilon_t \sim N(0, \sigma^2)$, rather unsatisfactory as a probability law that could have generated the data. We surely would not want to maintain that the change from c_1 to c_2 at date t_0 was a deterministic event that anyone would have been able to predict with certainty looking ahead from date $t = 1$. Instead, there must have been some imperfectly predictable forces that produced the change. Hence, rather than claim that expression (1) governed the data up to date t_0 and (2) after that date, what we must have in mind is that there is some larger model encompassing them both,

$$y_t = c_{st} + \varphi_{y_{t-1}} + \varepsilon_t, \tag{3}$$

where s_t is a random variable that, as a result of institutional changes, happened in our sample to assume the value $s_t = 1$ for $t = 1; 2; \ldots; t_0$ and $s_t = 2$ for $t = t_0 + 1; t_0 + 2; \ldots$ A complete description of the probability law governing the observed data would then require a probabilistic model of what caused the change from $s_t = 1$ to $s_t = 2$. The simplest such specification is that s_t is the realization of a two-state Markov chain with

$$\begin{aligned}\Pr(s_t = j|\, s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) \\ = \Pr(s_t = j|\, s_{t-1} = i) = p_{ij}.\end{aligned} \tag{4}$$

On the assumption that we do not observe s_t directly, but only infer its operation through the observed behavior of y_t, the parameters necessary to fully describe the probability law governing y_t are then the variance of the Gaussian innovation σ^2, the autoregressive coefficient φ , the two intercepts c_1 and c_2, and the two state transition probabilities, p_{11} and p_{22}.

The specification in (4) assumes that the probability of a change in regime depends on the past only through the value of the most recent regime, though, as noted below, nothing in the approach described below precludes looking at more general probabilistic specifications. But the simple time-invariant Markov chain (4) seems the natural starting point and is clearly preferable to acting as if the shift from c_1 to c_2 was a deterministic event. Permanence of the shift would be represented by $p_{22} = 1$, though the Markov formulation invites the more general possibility that $p_{22} < 1$. Certainly in the case of business cycles or financial crises, we know that the situation, though dramatic, is not permanent. Furthermore, if the regime change reflects a fundamental change in monetary or fiscal policy, the prudent assumption would seem to be to allow the possibility of it changing back again, suggesting that $p_{22} < 1$ is often a more natural formulation for thinking about changes in regime than $p_{22} = 1$.

A model of the form of (3)–(4) with no autoregressive elements ($\varphi = 0$) appears to have been first analysed by Lindgren (1978) and Baum et al. (1980). Specifications that incorporate autoregressive elements date back in the speech recognition literature to Poritz (1982), Juang and Rabiner (1985), and Rabiner (1989), who described such processes as 'hidden Markov models'. Markov-switching regressions were introduced in econometrics by Goldfeld and Quandt (1973), the likelihood function for which was first correctly calculated by Cosslett and Lee (1985). The formulation of the problem described here, in which all objects of interest are calculated as a by-product of an iterative algorithm similar in spirit to a Kalman filter, is due to Hamilton (1989, 1994). General characterizations of moment and stationarity conditions for such processes can be found in Tjøstheim (1986), Yang (2000), Timmermann (2000), and Francq and Zakoïan (2001).

Econometric Inference

Suppose that the econometrician observes y_t directly but can only make an inference about the value of s_t based on what we see happening with y_t. This inference will take the form of two probabilities

$$\xi_{jt} = \Pr(st = j|\, \Omega_t;\, \boldsymbol{\theta}) \tag{5}$$

for $j = 1, 2$, where these two probabilities sum to unity by construction. Here $\Omega_t = \{y_t, y_{t-1}, \ldots, y_1, y_0\}$ denotes the set of observations obtained as of date t, and $\boldsymbol{\theta}$ is a vector of population parameters, which for the above example would be $\boldsymbol{\theta} = (\sigma, \varphi, c_1, c_2, p_{11}, p_{22})'$,and which for now we presume to be known with certainty. The inference is performed iteratively for $t = 1; 2;. . ; T$, with step t accepting as input the values

$$\xi_{i,t-1} = \Pr(s_{t-1} = i | \Omega_{t-1}; \boldsymbol{\theta}) \tag{6}$$

for $i = 1, 2$ and producing as output (5). The key magnitudes one needs in order to perform this iteration are the densities under the two regimes,

$$\begin{aligned} \eta_{jt} &= f(y_t | s_t = j, \Omega_{t-1}; \boldsymbol{\theta}) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_t - c_j - \varphi y_{t-1})^2}{2\sigma^2}\right], \end{aligned} \tag{7}$$

for $j = 1, 2$. Specifically, given the input (6) we can calculate the conditional density of the tth observation from

$$f(y_t | \Omega_{t-1}; \boldsymbol{\theta}) = \sum_{i=1}^{2} \sum_{j=1}^{2} p_{ij} \xi_{i,t-1} \eta_{jt} \tag{8}$$

and the desired output is then

$$\xi_{jt} = \frac{\sum_{i=1}^{2} p_{ij} \xi_{i,t-1} \eta_{jt}}{f(y_t | \Omega_{t-1}; \boldsymbol{\theta})} \tag{9}$$

As a result of executing this iteration, we will have succeeded in evaluating the sample conditional log likelihood of the observed data

$$\begin{aligned} &\log f(y_1, y_2, \ldots, y_T | y_0; \boldsymbol{\theta}) \\ &= \sum_{t=1}^{T} \log f(y_t | \Omega_{t-1}; \boldsymbol{\theta}) \end{aligned} \tag{10}$$

for the specified value of $\boldsymbol{\theta}$. An estimate of the value of $\boldsymbol{\theta}$ can then be obtained by maximizing (10) by numerical optimization.

Several options are available for the value $\xi i0$ to use to start these iterations. If the Markov chain is presumed to be ergodic, one can use the unconditional probabilities

$$\xi_{i0} = \Pr(s_0 = i) = \frac{1 - p_{jj}}{2 - p_{ii} - p_{jj}}.$$

Other alternatives are simply to set $\xi_{i0} = 1/2$ or estimate ξ_{i0} itself by maximum likelihood.

The calculations do not increase in complexity if we consider an $(r \times 1)$ vector of observations $\mathbf{y}_t$ whose density depends on N separate regimes. Let $\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$ be the observations through date t, $\mathbf{P}$ be an $(N \times N)$ matrix whose row j, column i element is the transition probability p_{ij}, $\boldsymbol{\eta}_t$ be an $(N \times 1)$ vector whose jth element $f(\mathbf{y}_t | s_t = j, \Omega_{t-1}; \boldsymbol{\theta})$ is the density in regime j, and $\hat{\boldsymbol{\xi}}_{t|t}$ an $(N \times 1)$ vector whose jth element is $\Pr = (s_t = j, \Omega_t, \boldsymbol{\theta})$ Then (8) and (9) generalize to

$$f(\mathbf{y}_t | \Omega_{t-1}; \boldsymbol{\theta}) = \mathbf{1}'\left(\mathbf{P}\hat{\boldsymbol{\xi}}_{t-t|t-1} \odot \boldsymbol{\eta}_t\right) \tag{11}$$

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\mathbf{P}\hat{\boldsymbol{\xi}}_{t-t|t-1} \odot \boldsymbol{\eta}_t}{f(\mathbf{y}_t | \Omega_{t-1}; \boldsymbol{\theta})} \tag{12}$$

where $\mathbf{1}$ denotes an $(N \times 1)$ vector all of whose elements are unity and $\odot$ denotes element-by-element multiplication. Markov-switching vector autoregressions are discussed in detail in Krolzig (1997). Vector applications include describing the comovements between stock prices and economic output (Hamilton and Lin 1996) and the tendency for some series to move into recession before others (Hamilton and Perez-Quiros 1996). There further is no requirement that the elements of $\boldsymbol{\eta} t$ be Gaussian densities or even from the same family of densities. For example, Dueker (1997) studied a model in which the degrees of freedom of a Student t distribution change depending on the economic regime.

One is also often interested in forming an inference about what regime the economy was in at date t based on observations obtained through a later date T, denoted $\hat{\boldsymbol{\xi}}_{t|T}$. These are referred to as 'smoothed' probabilities, an efficient algorithm for whose calculation was developed by Kim (1994).

Extensions

The calculations in (11) and (12) remain valid when the probabilities in $\mathbf{P}$ depend on lagged values of $\mathbf{y}_t$ or strictly exogenous explanatory variables, as in Diebold et al. (1994), Filardo (1994) and Peria (2002). However, often there are relatively few transitions among regimes, making it difficult to estimate such parameters accurately, and most applications have assumed a time-invariant Markov chain. For the same reason, most applications assume only $N = 2$ or 3 different regimes, though there is considerable promise in models with a much larger number of regimes, either by tightly parameterizing the relation between the regimes (Calvet and Fisher 2004), or with prior Bayesian information (Sims and Zha 2006).

In the Bayesian approach, both the parameters $\boldsymbol{\theta}$ and the values of the states $\mathbf{s} = (s_1, s_2, \ldots, s_T)'$ are viewed as random variables. Bayesian inference turns out to be greatly facilitated by Monte Carlo Markov chain methods, specifically, the Gibbs sampler. This is achieved by sequentially (for $k = 1, 2, \ldots$) generating a realization $\boldsymbol{\theta}^{(k)}$ from the distribution of $\boldsymbol{\theta}|\mathbf{s}^{(k-1)}, \Omega_T$, followed by a realization of $\mathbf{s}^{(k)}$ from the distribution of $\mathbf{s}|\boldsymbol{\theta}^{(k)}, \Omega_T$. The first distribution, $\boldsymbol{\theta}|\mathbf{s}^{(k-1)}, \Omega_T$, treats the historical regimes generated at the previous iteration $s_1^{(k-1)}, s_2^{(k-1)}, \ldots, s_T^{(k-1)}$, as if fixed known numbers. Often this conditional distribution takes the form of a standard Bayesian inference problem whose solution is known analytically using natural conjugate priors. For example, the posterior distribution of φ given other parameters is a known function of easily calculated OLS coefficients. An algorithm for generating a draw from the second distribution, $\mathbf{s}|\boldsymbol{\theta}^{(k)}$, Ω_T, was developed by Albert and Chib (1993). The Gibbs sampler turns out also to be a natural device for handling transition probabilities that are functions of observable variables, as in Filardo and Gordon (1998).

It is natural to want to test the null hypothesis that there are N regimes against the alternative of $N + 1$, for example when $N = 1$, to test whether there are any changes in regime at all. Unfortunately, the likelihood ratio test of this hypothesis fails to satisfy the usual regularity conditions because, under the null hypothesis, some of the parameters of the model would be unidentified. For example, if there is really only one regime, the maximum likelihood estimate $\hat{p}_{11}$ does not converge to a well-defined population magnitude, meaning that the likelihood ratio test does not have the usual χ^2 limiting distribution. To interpret a likelihood ratio statistic, one instead needs to appeal to the methods of Hansen (1992) or Garcia (1998). An alternative is to rely on generic tests of the hypothesis that an N-regime model accurately describes the data (Hamilton 1996), though these tests are not designed for optimal power against the specific alternative hypothesis of $N + 1$ regimes. A test recently proposed by Carrasco et al. (2004) that is easy to compute but not based on the likelihood ratio statistic seems particularly promising. Other alternatives are to use Bayesian methods to calculate the value of N implying the largest value for the marginal likelihood (Chib 1998) or the highest Bayes factor (Koop and Potter 1999), or to compare models on the basis of their ability to forecast (Hamilton and Susmel 1994).

A specification where the density depends on a finite number of previous regimes, $f(\mathbf{y}_t| s_t, s_{t-1}, \ldots, s_{t-m}, \Omega_{t-1}; \boldsymbol{\theta})$ can be recast in the above form by a suitable redefinition of regime. For example, if s_t follows a 2-state Markov chain with transition probabilities $\Pr(s_t = j| s_{t-1} = i)$ and $m = 1$, one can define a new regime variable s_t^* such that $f\left(\mathbf{y}_t| s_t^*, \Omega_{t-1}; \boldsymbol{\theta}\right) = f\left(y_t| s_t, s_{t-1}, \ldots, s_{t-m}, \Omega_{t-1}; \boldsymbol{\theta}\right)$ as follows:

$$s_t^* = \begin{cases} 1 & \text{when} \quad s_t = 1 \quad \text{and} \quad s_{t-1} = 1 \\ 2 & \text{when} \quad s_t = 2 \quad \text{and} \quad s_{t-1} = 1 \\ 3 & \text{when} \quad s_t = 1 \quad \text{and} \quad s_{t-1} = 2 \\ 4 & \text{when} \quad s_t = 2 \quad \text{and} \quad s_{t-1} = 2 \end{cases}.$$

Then s_t^* itself follows a 4-state Markov chain with transition matrix.

$$\mathbf{P}^* = \begin{bmatrix} p_{11} & 0 & p_{11} & 0 \\ p_{12} & 0 & p_{12} & 0 \\ 0 & p_{21} & 0 & p_{21} \\ 0 & p_{22} & 0 & p_{22} \end{bmatrix}$$

More problematic are cases in which the order of dependence m grows with the date of the observation t. Such a situation often arises in models whose recursive structure causes the density of y_t given Ω_{t-1} to depend on the entire history y_{t-1}, $y_{t-2}, \ldots, y_1$ as is the case in ARMA, GARCH or state-space models. Consider for illustration a GARCH(1,1) specification in which the coefficients are subject to changes in regime, $y_t = h_t v_t$, where $v_t \sim N(0,\ 1)$ and

$$h_t^2 = \gamma_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} h_{t-1}^2. \qquad (13)$$

Solving (13) recursively reveals that the conditional standard deviation $h_{t.}$ depends on the full history $\{y_{t-1}, y_{t-2}, \ \ldots, y_0, s_t, s_{t-1}, \ \ldots, s_1\}$. One way to avoid this problem was proposed by Gray (1996), who postulated that, instead of being generated by (13), the conditional variance is characterized by

$$h_t^2 = \gamma_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{st} \tilde{h}_{t-1}^2 \qquad (14)$$

where

$$\tilde{h}_{t-1}^2 = \sum_{i=1}^{N} \tilde{\xi}_{i,t-1|t-2} \left(\gamma_i + \alpha_i y_{t-2}^2 + \beta_i \tilde{h}_{t-2}^2 \right)$$

In Gray's model, $h_{t.}$ in (14) depends only on s_t since $\tilde{h}_{t-1}^2$ is a function of data Ω_{t-1} only. An alternative solution, due to Haas et al. (2004), is to hypothesize N separate GARCH processes whose values h_{it} all exist as latent variables at date t,

$$h_{it}^2 = \gamma_i + \alpha_i y_{t-1}^2 + \beta_i h_{i,t-1}^2 \qquad (15)$$

and then simply pose the model as $y_t = h_{st} v_t$. Again, the feature that makes this work is the fact that h_{it} in (15) is a function solely of the data Ω_{t-1} rather than the states $\{s_{t-1}, s_{t-2}, \ \ldots, s_1\}$.

A related problem arises in Markov-switching state-space models, which posit an unobserved state vector $\mathbf{z}_t$ characterized by

$$\mathbf{z}_t = \mathbf{F}_{s_t} \mathbf{z}_{t-1} + \mathbf{Q}_{s_t} \mathbf{v}_t$$

with $\mathbf{v}_t \sim N(0,\ \mathbf{I}_n)$, with observed vectors $\mathbf{y}_t$ and $\mathbf{x}_t$ governed by

$$\mathbf{y}_t = \mathbf{H}_{s_t}' \mathbf{z}_t + \mathbf{A}_{s_t}' \mathbf{x}_t + \mathbf{R}_{s_t} \mathbf{w}_t$$

for $\mathbf{w}_t \sim N(0,\ \mathbf{I}_r)$. Again, the model as formulated implies that the density of $\mathbf{y}_t$ depends on the full history $\{s_t, s_{t-1}, \ \ldots, s_1\}$. Kim (1994) proposed a modification of the Kalman filter equations similar in spirit to the modification in (14) that can be used to approximate the log likelihood. A more common practice recently has been to estimate such models with numerical Bayesian methods, as in Kim and Nelson (1999).

See Also

- Markov Chain Monte Carlo Methods
- Markov Processes
- Maximum Likelihood
- Mixture Models
- Non-linear Time Series Analysis
- Numerical Optimization Methods in Economics
- Structural Change

Bibliography

Albert, J., and S. Chib. 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* 11: 1–15.

Ang, A., and G. Bekaert. 2002a. International asset allocation with regime shifts. *Review of Financial Studies* 15: 1137–1187.

Ang, A., and G. Bekaert. 2002b. Regime switches in interest rates. *Journal of Business and Economic Statistics* 20: 163–182.

Baum, L., E. Petrie, G. Soules, and N. Weiss. 1980. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41: 164–171.

Calvet, L., and A. Fisher. 2004. How to forecast long-run volatility: Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics* 2: 49–83.

Carrasco, M., L. Hu, and W. Ploberger. 2004. Optimal test for Markov switching. Working paper. University of Rochester.

Cerra, V., and S. Saxena. 2005. Did output recover from the Asian crisis? *IMF Staff Papers* 52: 1–23.

Chauvet, M., and J. Hamilton. 2006. Dating business cycle turning points. In *Nonlinear time series analysis of business cycles*, ed. C. Milas, P. Rothman, and D. van Dijk. Amsterdam: Elsevier.

Chib, S. 1998. Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86: 221–241.

Cosslett, S., and L.-F. Lee. 1985. Serial correlation in discrete variable models. *Journal of Econometrics* 27: 79–97.

Dai, Q., K. Singleton, and W. Yang. 2003. Regime shifts in a dynamic term structure model of U.S. Treasury bonds. Working paper, Stanford University.

Davig, T. 2004. Regime-switching debt and taxation. *Journal of Monetary Economics* 51: 837–859.

Diebold, F., J.-H. Lee, and G. Weinbach. 1994. Regime switching with time-varying transition probabilities. In *Nonstationary time series analysis and cointegration*, ed. C. Hargreaves. Oxford: Oxford University Press.

Dueker, M. 1997. Markov switching in GARCH processes and mean-reverting stockmarket volatility. *Journal of Business and Economic Statistics* 15: 26–34.

Filardo, A. 1994. Business cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12: 299–308.

Filardo, A., and S. Gordon. 1998. Business cycle durations. *Journal of Econometrics* 85: 99–123.

Francq, C., and J.-M. Zakoïan. 2001. Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* 102: 339–364.

Garcia, R. 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39: 763–788.

Garcia, R., R. Luger, and E. Renault. 2003. Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics* 116: 49–83.

Goldfeld, S., and R. Quandt. 1973. A Markov model for switching regressions. *Journal of Econometrics* 1: 3–16.

Gray, S. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42: 27–62.

Haas, M., S. Mittnik, and M. Paolella. 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2: 493–530.

Hamilton, J. 1988. Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12: 385–423.

Hamilton, J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–384.

Hamilton, J. 1994. *Time series analysis*. Princeton: Princeton University Press.

Hamilton, J. 1996. Specification testing in Markov-switching time-series models. *Journal of Econometrics* 70: 127–157.

Hamilton, J. 2005. What's real about the business cycle? *Federal Reserve Bank of St. Louis Review* 87: 435–452.

Hamilton, J., and G. Lin. 1996. Stock market volatility and the business cycle. *Journal of Applied Econometrics* 11: 573–593.

Hamilton, J., and G. Perez-Quiros. 1996. What do the leading indicators lead? *Journal of Business* 69: 27–49.

Hamilton, J., and R. Susmel. 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64: 307–333.

Hansen, B. 1992. The likelihood ratio test under non-standard conditions. *Journal of Applied Econometrics* 7: S61–S82. Erratum, 11(1996), 195–198.

Jeanne, O., and P. Masson. 2000. Currency crises, sunspots, and Markov-switching regimes. *Journal of International Economics* 50: 327–350.

Juang, B.-H., and L. Rabiner. 1985. Mixture autoregressive hidden Markov models for speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 30: 1404–1413.

Kim, C. 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60: 1–22.

Kim, C., and C. Nelson. 1999. *State-space models with regime switching*. Cambridge, MA: MIT Press.

Koop, G., and S. Potter. 1999. Bayes factors and nonlinearity: Evidence from economic time series. *Journal of Econometrics* 88: 251–281.

Krolzig, H.-M. 1997. *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*. Berlin: Springer.

Lindgren, G. 1978. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics* 5: 81–91.

Peria, M. 2002. A regime-switching approach to the study of speculative attacks: A focus on EMS crises. In *Advances in Markov-switching models*, ed. J. Hamilton and B. Raj. Heidelberg: Physica Verlag.

Poritz, A. 1982. Linear predictive hidden Markov models and the speech signal. *Acoustics, Speech and Signal Processing, IEEE Conference on ICASSP '82* 7: 1291–1294.

Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77: 257–286.

Sims, C., and T. Zha. 2006. Were there switches in U.S. monetary policy? *American Economic Review* 96: 54–81.

Timmermann, A. 2000. Moments of Markov switching models. *Journal of Econometrics* 96: 75–111.

Tjøstheim, D. 1986. Some doubly stochastic time series models. *Journal of Time Series Analysis* 7: 51–72.

Yang, M. 2000. Some properties of vector autoregressive processes with Markov-switching coefficients. *Econometric Theory* 16: 23–43.

Regional and Preferential Trade Agreements

Pravin Krishna

Abstract

This article discusses analytical developments in the literature on the economics and politics of preferential trade agreements (PTAs). It describes results obtained in the traditional theory that demonstrate the ambiguous welfare outcomes of preferential trade liberalization. Theoretical approaches to designing necessarily welfare- improving PTAs are also discussed. Finally, this article sets out recent analyses in the literature concerning the dynamic expansion of trade blocs, the endogenous determination of policy (relating to preferences within a PTA and to extra-union trade), and the effects of preferential agreements on the multilateral trade system.

Keywords

ASEAN free trade area (AFTA); Bargaining; Coalitions; Comparative advantage; Customs unions; European economic community (EEC); European free trade association (EFTA); Free trade areas; General agreement on tariffs and trade (GATT); Great depression; Income distribution; International trade; Meade, J. E.; Mercosur; Most-favoured-nation (MFN) principle; Multilateral free trade; Mundell, R.; Natural trading partners hypothesis; North American free trade agreement (NAFTA); Preferential trade agreements; Redistribution of income; Regional and preferential trade agreements; Sunk costs; Tariffs; Terms of trade; Trade creation; Trade diversion; Viner, J

JEL Classifications

F1; F15

Strongly influenced by the perception that restricted commerce and preferences in trade relations had contributed to the Great Depression of the 1930s and the subsequent outbreak of war, the discussions leading to the General Agreement on Tariffs and Trade (GATT) in 1947 were driven by the desire to create an international economic order based on a liberal and non-discriminatory multilateral trade system. Enshrined in Article I of the GATT, the principle of non-discrimination (commonly referred to as the most-favoured-nation or MFN clause) precludes member countries from discriminating against imports based upon the country of origin. However, in an important exception this central prescript, the GATT, through its Article XXIV, permits its members to enter into preferential trade agreements (PTAs), provided these preferences are complete. In so doing, it sanctions the formation of free trade areas (FTAs), whose members are obligated to eliminate internal import barriers, and customs unions (CUs), whose members additionally agree on a common external tariff against imports from non-members. Additional derogations to the principle of non-discrimination now include the Enabling Clause, which allows tariff preferences to be granted to developing countries (in accordance with the Generalized System of Preferences) and permits preferential trade agreements among developing countries in goods trade. Among the more prominent existing PTAs are the North American Free Trade Agreement (NAFTA), the European Economic Community (EEC) and the European Free Trade Association (EFTA), all formed under Article XXIV, and the Mercosur (the CU between the Argentine Republic, Brazil, Paraguay, and Uruguay) and the ASEAN (Association of South East Asian Nations) Free Trade Area (AFTA), both formed under the Enabling Clause.

Static Welfare Analysis

Motivated by ongoing discussions concerning optimal trade arrangements in the post-war period, especially over the possibility of a

European customs union, Jacob Viner (1950, pp. 41–50) developed a seminal analysis of the economics of preferential trade. Viner's analysis undermines the presumption that cutting tariffs is necessarily welfare improving. On the one hand, because of discriminatory liberalization, there will be commodities that a member country may 'newly import from the other but which it formerly did not import at all because the price of the protected domestic good was lower than the price of any foreign source plus the duty'. Viner calls this shift from a high-cost to a lower-cost point 'trade creation' and associates it with welfare-improvement for the importing country. He also argues that, on the other hand, 'there may be other commodities, which one of the members will now newly import from the other', whereas before the PTA it 'imported them from a third country, because that was the cheapest possible source of supply even after the payment of duty'. He calls this shift in imports from a low-cost third country to a higher-cost member country 'trade diversion,' associating it with an increase in the cost of imports and, thus, welfare losses for the importing country.

The demonstration that preferential trade liberalization may be welfare decreasing stimulated a substantial theoretical literature on the 'static' welfare effects of PTAs. Post-Vinerian analysis of the welfare effects of preferential trade include Meade's (1955) more explicit and comprehensive formulation of the problem in a three-country three-good setting. Meade argues that not only the magnitudes of trade creation and trade diversion but also the extent of cost reductions (in the former) and increases (in the latter) were necessary to arrive at a welfare evaluation. Subsequent analysis also developed examples of both welfare improving trade diversion and welfare-decreasing trade creation in general equilibrium contexts broader than those considered by Viner (see, for instance, Gehrels 1956–1957, Lipsey 1957, and Bhagwati 1971). However, the intuitive appeal of the concepts of trade creation and trade diversion has ensured their continued use in the economic analysis of preferential trade agreements, especially in policy analysis (see Panagariya 2000, for a comprehensive survey).

The effects of preferences on intra-union and extra-union terms-of-trade are analysed by Mundell (1964), who argues that a country granting tariff preferences moves intra-union terms of trade against it and in favour of its partner by increasing its demand for imports from its partner. Extra-union terms of trade are improved for the partner, but change ambiguously for the preference-granting country. Thus, tariff preferences have asymmetric effects on the preference-granting and preference-receiving country. More sharply, Panagariya (1997a) shows how, even with fixed extra-union terms of trade, if a preference-granting country continues to import from the rest of the world, intra-union terms-of-trade losses (manifesting themselves as intra-union tariff revenue transfers as also seen in Berglas 1979, and Riezman 1979) unambiguously worsen its welfare, while its preference-receiving partner unambiguously gains.

Wonnacott and Lutz (1987), Krugman (1991) and Summers (1991) propose geographic proximity between partner countries as important in ensuring that preferential liberalization improves welfare. Specifically, they suggest that countries entering into preferential arrangements with geographically proximate countries are likely to do better than in agreements with distant countries, because the former are more likely than the latter to be trade creating, leading to a larger improvement in welfare. Proximate countries are thus argued to be 'natural' partners for preferential trade. Bhagwati and Panagariya (1996) and Panagariya (1997b), however, provide a number of examples in which, between two otherwise identical potential partners, a country achieves a superior outcome by granting trade preferences to the distant partner. For instance, it may be that a preference granted to a distant partner leads to a smaller transfer (loss) of tariff revenue with a closer country, since, with an initial non-discriminatory tariff, the liberalizing country imports less from the more distant partner. Thus, the 'natural trading partners' hypothesis is shown to lack general theoretical validity.

Numerous studies have attempted to evaluate quantitatively the trade creation and trade diversion effects of PTAs. Recently, focusing on the effects of PTAs on excluded countries, Chang and

Winters (2002) show how Mercosur was associated with significant declines in the prices of nonmembers' exports to the region. Yeats (1998) shows how under Mercosur the greatest increases of intra-union trade flows were in goods in which the member countries had the least comparative advantage, confirming the trade diversionary effects of preferential liberalization.

Srinivasan et al. (1993) note that the econometric frameworks used in most *ex post* studies of trade flows generally lack microeconomic underpinnings, making an evaluation of the associated welfare consequences difficult. Krishna (2003) develops an econometric framework for the analysis of PTAs with a strong welfare-theoretic foundation, so that the estimated parameters relating to trade creation and trade diversion effects fit directly into theoretically derived welfare expressions. His application of this framework to the evaluation of the natural trading partners hypothesis does not find any support in US data.

Necessarily Welfare-Improving Preferential Trade Areas

The generally ambiguous welfare results provided by the theoretical literature raised an important question relating to the *design* of necessarily welfare-improving PTAs. A classic result stated independently by Kemp (1964) and Vanek (1965) and proved subsequently by Ohyama (1972) and Kemp and Wan (1976) provides a welfare-improving solution for the case of CUs. Starting from a situation with an arbitrary structure of trade barriers, if two or more countries freeze their net external trade vector with the rest of the world through a set of common external tariffs and eliminate the barriers to internal trade (which implies the formation of a CU), the welfare of the union as a whole necessarily improves (weakly) and that of the rest of the world does not fall. A Pareto-improving preferential trade agreement may thus be achieved. The logic behind the Kemp–Wan theorem is as follows. By fixing the combined, net extra-union trade vector of member countries at its pre-union level, non-member countries are guaranteed their original level of welfare. Moreover, if we take the extra-union trade vector as an endowment, the joint welfare of the union is maximized by allowing free trade of goods internally (thus equating the marginal rate of substitution and marginal rate of transformation for each pair of commodities to each other and across all agents in the union). The PTA thus constructed has a common internal price vector, implying further a common external tariff for member countries. This customs union is (weakly) welfare improving; the rest of the world is no worse off and the welfare of member countries is jointly improved (weakly). Welfare improvement is achieved even if additional 'non-economic' objectives (such as maintaining the output of a sector or its employment of a factor) are introduced, as Krishna and Bhagwati (1997) show. The Kemp–Wan–Ohyama design, by freezing the external trade vector and thus eliminating trade diversion, offers a way to sidestep the complexities and ambiguities inherent in the analysis of PTAs. It has played an important role in shaping the way that economists think about issues relating to the design and implementation of PTAs.

The Kemp–Wan–Ohyama analysis of welfare improving CUs does not extend easily to FTAs, however. In the case of an FTA, member-specific tariff vectors imply that the domestic-price vectors differ across member countries and the FTA generally fails to equalize marginal rates of substitution across its members. Without a common internal price vector, however, the Kemp–Wan–Ohyama methodology lacks application. Nevertheless, Panagariya and Krishna (2002) have provided a corresponding construction of necessarily welfare-improving FTAs. The Panagariya–Krishna FTA, in complete analogy with the Kemp–Wan CU, freezes the external trade vector of the area, with the essential difference that the trade vector of each member country with the rest of the world is frozen at the pre-FTA level. Since, in FTAs, different member countries impose different external tariffs, it is necessary to specify a set of rules of origin to prevent a subversion of FTA tariffs by importing through the lower-tariff member country and directly transshipping goods to the higher-tariff country (which, if allowed, would bring the FTA arbitrarily close to a CU). The Panagariya–Krishna solution requires that all goods for which *any* value is added within the FTA are to be traded freely. Importantly, the proportion of domestic value added in final goods does not enter as a criterion in the rules of origin.

Theory thus suggests that ensuring welfare improvement requires that, along with elimination of internal barriers, external tariff vectors should eliminate trade diversion – member countries should continue to import the same amounts from the rest of the world as they did initially. There have, however, been significant departures in practice. While Article XXIV of the GATT stipulates that internal restrictions be eliminated on 'substantially all trade', the qualifier 'substantially' is vulnerable to abuse. Numerous goods are typically exempt from internal liberalization by member countries. Furthermore, restrictive rules of origin also serve to ensure a level of protection from both intra- and extra-union imports, as Krueger (1999) notes. On external tariffs, Article XXIV requires that external barriers not be more restrictive than initially. For FTAs, since countries retain individual tariff vectors, this could be taken to imply that no tariff is to rise. For CUs, since a common external tariff is to be chosen and initial tariffs on the same good likely vary across countries, the tariff vector would necessarily change for each country. The expectation is that that the 'general incidence' of trade barriers should not be higher or more restrictive than before. As Bhagwati (1993) notes, it is clear that Article XXIV's ambiguity in this regard leaves plenty of room for protectionist behaviour by member countries. The 1994 'Understanding on the Interpretation' of Article XXIV issued by the GATT provides greater clarity on the issue of measurement and choice of the common external tariff – indicating that the GATT secretariat would compute weighted average tariff rates and duties collected in accordance with the methodology used in the assessment of tariff offers in the Uruguay Round of trade negotiations and examine trade flow and other data to arrive at suitable measures of non-tariff barriers. Nevertheless, it may be observed that leaving external barriers at their initial level and removing internal barriers do not eliminate trade diversion. Indeed, with this configuration, some trade diversion is practically guaranteed.

Preferential Trade Agreements and Multilateral Free Trade

Recent analysis in the literature has focused on issues concerning the expansion of trade blocs, the endogenous determination of policy (relating to trade preferences within a PTA and extra-union trade), and the effects of preferential agreements on the multilateral trade system (that is, whether trade blocs will serve as 'building blocs' or 'stumbling blocs' in the path to multilateral free trade, in Bhagwati's 1993, phrasing).

Krugman (1993) analyses the welfare consequences of exogenously formed and expanded trade blocs. Considering a fully symmetric structure of countries, each specialized in production in a differentiated product variety, Krugman asks how world welfare is affected by the expansion of trade blocs if member countries liberalize fully their mutual trade but apply optimal tariffs against non-members. As the (symmetric) trade blocs increase in size, their market power increases and so do the (optimal) tariffs they impose on non-members. On the other hand, increasing the number of countries within a bloc increases the volume of goods that is traded freely. Krugman finds that the net effect on world welfare is non-monotonic in bloc size. Specifically, world welfare (which is maximized with global free trade) falls as the world is divided up into trade blocs but rises again as bloc sizes decrease and the trade diversion losses (relative to trade creation gains) fall. Bond and Syropoulos (1996) show how generalizing the assumptions of Krugman's model relating to consumption preferences and the pattern of production and trade may alter the relationship between optimal tariffs and market size so that optimal tariffs fall as bloc size increases. More severely, Deardorff and Stern (1994) and Srinivasan (1993) question the robustness of Krugman's conclusions concerning non-monotonicity of welfare itself, demonstrating a substantial divergence in results when Krugman's assumptions regarding the structure of endowments and comparative advantage are changed.

A different strand of the literature has examined endogenously determined trade blocs and the internal political and economic incentives (if any) for their successive expansion. Taking the 'interest-group' approach to trade policy determination, Grossman and Helpman (1995) and Krishna (1998) both model the influence of powerful producers in considering entry into a PTA. While the models and analytic frameworks differ in detail, they come to a similar and striking conclusion, namely, that PTAs that divert trade are more likely to win internal political support. This is so because governments must respond to conflicting pressures from their exporting sectors, which gain from lower trade barriers in the partner, and from their import- competing sectors, which suffer from lower trade barriers at home, when deciding on whether to form or enter a PTA.

As Krishna (1998) argues, trade diversion effectively shifts the burden of the gain to member-country exporters from member-country import-competing sectors and onto non-member producers, who have little political clout inside the member countries. Krishna (1998) also argues that such PTAs will lower the within-union incentives for any subsequent multilateral liberalization – producers in trade-diverting PTAs may oppose multilateral reform since this would take away the gains from benefits of preferential access that they enjoyed in the PTA that diverted trade to them. Under some circumstances, the within-union incentives for further multilateral liberalization are completely eliminated.

Levy (1997) models trade policy as being determined by majority voting. Countries are assumed to differ in their endowments of factors (labour and capital). Countries are also assumed to produce different varieties of goods – so that trade reform will result in gains to individuals due to the greater number of varieties that are available for consumption. However, it should also be clear that any changes in trade policy result in changes in the distribution of income (by altering the relative rewards to the different factors of production). The arguments that emerge out of this framework are as follows. First, preferential trade integration with partners with similar relative factor endowments (that is, with similar capital–labour ratios) is more likely to receive majority support – since this results in minimal income redistribution and still provides variety gains from trade. Second, bilateral agreements could render infeasible multilateral liberalization (which, even if it brings greater variety gains, would involve trade with countries with quite different relative endowments of capital and labour and could therefore result in much more drastic income redistribution).

McLaren (2002) provides an analysis of the role of sunk costs and trade policy determination. He argues, roughly speaking, that the expectation of a preferential trading agreement could induce agents in the economy to undertake costly and irreversible investments that makes the members within the bloc more specialized towards each other and less so towards the rest of the world. In other words, they increase dependence on each other and lower it towards the rest of the world, and thus reduce their desire to liberalize trade with other countries. Exploring the first-mover advantage that member countries gain having invested in sunk costs, Freund (2000a) finds that with preferential trade member countries gain and that nonmembers lose relative to multilateral outcomes, with the former dominating the latter in magnitude.

A parallel literature has raised the question of what external tariffs will be chosen by member countries, examining, in particular, whether external tariffs can be expected to rise or fall following a PTA. No clear answers to this question emerge. Panagariya and Findlay (1996) finds that external tariffs rise after tariff preferences are granted, as political lobbying for protection is directed away from imports from the partner country to imports from the rest of the world. Emphasizing tariff revenue competition between FTA members, Richardson (1995) finds that external tariff may fall in an FTA as welfare-minded member countries competitively reduce tariffs (so as to retain the source of extra-union imports and earn tariff revenue). In a general equilibrium context, with political lobbying over tariffs, Cadot et al. (1999) reach a similar conclusion for FTAs, while finding that CUs are likely to raise their external tariffs. Cadot et al. (2002) confirm these results for the case of quantity restrictions, where the protective effect that a quantity restriction imposed by a member country has in partner country markets proves central to the analysis. However, when collective action problems over lobbying for external trade policy dominate, FTAs may choose higher external tariffs than CUs, as Richardson (1994) argues. Finally, in a symmetric three-country oligopoly model, Ornelas (2005) finds that a PTA's endogenously determined external tariff may be lower than the pre-union MFN tariff. Cho and Krishna (2006) demonstrate that the opposite may obtain with asymmetric costs across partner countries, the likelihood of the external tariff being higher than the pre-union MFN tariff increasing with the inefficiency of the partner country relative to non-members.

Empirical analysis has offered mixed results as well. Bohara et al. (2004) report evidence of lowered external tariffs following Mercosur, while Cho (2006) finds tariffs in Mexico to be higher on average following NAFTA and systematically higher in goods in which its trading partners were inefficient suppliers (as proxied by pre-FTA export levels relative to the rest of the world). Limão (2006) examines data on US trade barriers and finds that those imported goods on which any partners received preferences were subject to smaller (subsequent) multilateral liberalization than others, suggesting a negative effect of preferences on multilateral reforms.

The economic incentives of non-member countries are considered by Baldwin (1995), who argues that PTA expansion could have 'domino' effects – increasing the size of a bloc increases the incentive for others to join it (as they then gain preferential access to increasingly large markets). On the assumption of open- membership rules (that is, insiders do not oppose the entry of new members who abide by the same rules as the members), the successive expansion of the PTA could then lead to multilateral free trade – a conclusion that is also reached by the work of Yi (1996), which develops a model of endogenous coalition formation to addresses this question.

Aghion et al. (2004) analyse the links between bilateral and multilateral negotiations over trade policy as a dynamic bargaining game in which a leading country endogenously decides whether to sequentially negotiate free trade agreements with subsets of countries or engage in simultaneous multilateral bargaining with all countries at once. They show that, if a coalition formed between the leading country and a follower generates a negative effect on outsiders (that is, there are negative coalitional externalities), the leader prefers sequential bargaining to multilateral bargaining. Conversely, positive coalition externalities imply that multilateral bargaining is preferred. Importantly, while political economy pressures may cause bilateral agreements to impede multilateral agreement, as in Levy (1997) and Krishna (1998), examples where bilateral agreements enable multilateral agreement are also found.

Self-enforcing trade agreements (which work by balancing any benefits that member countries may achieve by deviating from the agreement with the future losses they suffer due to punishments for the deviation) have been variously analysed in the international trade literature. Since bilateral (multilateral) agreements may alter both the benefits of deviating from an existing multilateral (bilateral) agreement and the future punishment costs of this deviation, self-enforcing agreements provide a context in which the links between preferential trade agreements and multilateralism may be studied. Bagwell and Staiger (1997a, b) consider the impact of FTAs and CUs on multilateral tariff cooperation during a transition period when the exogenously agreed-upon lowering of tariffs within FTAs and CUs is implemented. Saggi (2006) shows how exogenously specified FTAs and CUs may undermine self-enforced multilateral tariff cooperation, the former by lowering the cooperation incentives of non-member countries and the latter by lowering the cooperation incentives of members. Freund (2000b) finds that exogenous multilateral liberalization may encourage and help sustain self-enforcing PTAs, thus explaining the recent trend towards bilateralism as a causal response to multilateralism. A similar causal link is explored by Cadot et al. (2001), which views bilateral agreements as an endogenous (protective) response to the competitive pressures that domestic producers face with multilateral liberalization.

Conclusions

A half-century of research has advanced significantly our understanding of the implications of trade discrimination even if the frequently equivocal theoretical and empirical results have established among economists and policymakers an ambivalent attitude towards preferential trade agreements. However, concerns regarding the fragmentation of the world trade system have grown with the rapid proliferation of preferential trade in recent years. Several hundred PTAs are currently in existence. Indeed, many countries belong to multiple PTAs – resulting in a confusing criss-crossing of trade preferences that Bhagwati (1995) has aptly described as 'spaghetti-bowl' regionalism. Several more preferential agreements are in process. With this inexorable erosion of non-discriminatory disciplines within the trade system, research on preferential trade is certain to remain central to the field of international trade policy for many years to come.

See Also

▶ Mercosur
▶ North American Free Trade Agreement (NAFTA)
▶ Trade Policy, Political Economy of

Bibliography

Aghion, P., P. Antras, and E. Helpman. 2004. Negotiating free trade. Working paper no. 10721. Cambridge: NBER.

Bagwell, K., and R. Staiger. 1997a. Multilateral cooperation during the formation of free trade areas. *International Economic Review* 38: 291–319.

Bagwell, K., and R. Staiger. 1997b. Multilateral cooperation during the formation of customs unions. *Journal of International Economics* 42: 91–123.

Baldwin, R. 1995. A domino theory of regionalism. In *Expanding European regionalism: The EU's new members*, ed. R. Baldwin, P. Haaparanta, and J. Kiander. Cambridge: Cambridge University Press.

Berglas, E. 1979. Preferential trading: The n commodity case. *Journal of Political Economy* 87: 315–331.

Bhagwati, J. 1971. Trade-diverting customs unions and welfare improvement: A clarification. *Economic Journal* 81: 580–587.

Bhagwati, J. 1993. Regionalism and multilateralism: An overview. In *New dimensions in regional integration*, ed. J. de Melo and A. Panagariya. Cambridge: Cambridge University Press.

Bhagwati, J. 1995. U.S. trade policy: The infatuation with free trade areas. In *The dangerous drift to preferential trade agreements*, ed. J. Bhagwati and A. Krueger. Washington, DC: AEI Press.

Bhagwati, J., and A. Panagariya. 1996. *Free trade areas or free trade? The economics of preferential trade areas*. Washington, DC: AEI Press.

Bohara, A., K. Gawande, and P. Sanguinetti. 2004. Trade diversion and declining tariffs: Evidence from Mercosur. *Journal of International Economics* 64: 65–88.

Bond, E., and C. Syropoulos. 1996. The size of trading blocs, market power and world welfare effects. *Journal of International Economics* 40: 411–437.

Cadot, O., J. de Melo, and M. Olarreaga. 1999. Regional integration and lobbying for tariffs against non-members. *International Economic Review* 40: 635–658.

Cadot, O., J. de Melo, and M. Olarreaga. 2001. Can bilateralism ease the pains of multilateral trade liberalization? *European Economic Review* 45: 27–44.

Cadot, O., J. de Melo, and M. Olarreaga. 2002. Harmonizing external quotas in an FTA: A step backward? *Economics and Politics* 14: 259–282.

Chang, W., and A. Winters. 2002. How regional blocs affect excluded countries: The price effects of Mercosur. *American Economic Review* 92: 889–904.

Cho, M.-J. 2006. *On the external trade barriers of preferential trade agreements: Empirical evidence from Mexican manufacturing industries*. Mimeo: Brown University.

Cho, M.-J., and P. Krishna. 2006. *On the external trade barriers of preferential trade agreements*. Mimeo: Brown University.

Deardorff, A., and R. Stern. 1994. Multilateral trade negotiations and preferential trading arrangements. In *Analytical and negotiating issues in the global trading system*, ed. A. Deardorff and R. Stern. Ann Arbor: University of Michigan Press.

Freund, C. 2000a. Different paths to free trade: The gains from regionalism. *Quarterly Journal of Economics* 115: 1317–1341.

Freund, C. 2000b. Multilateralism and the endogenous formation of preferential trade agreements. *Journal of International Economics* 52: 359–376.

Gehrels, F. 1956–1957. Customs union from a single–country viewpoint. *Review of Economic Studies* 24: 61–64.

Grossman, G., and E. Helpman. 1995. The politics of free trade arrangements. *American Economic Review* 85: 667–690.

Kemp, M. 1964. *The pure theory of international trade*. Englewood Cliffs: Prentice-Hall.

Kemp, M., and H. Wan. 1976. An elementary proposition concerning the formation of customs unions. *Journal of International Economics* 6: 95–98.

R

Krishna, P. 1998. Regionalism and multilateralism: A political economy approach. *Quarterly Journal of Economics* 113: 227–250.
Krishna, P. 2003. Are regional trading partners 'natural'? *Journal of Political Economy* 111: 202–231.
Krishna, P., and J. Bhagwati. 1997. Necessarily welfare-improving customs unions with industrialization constraints: The Cooper–Massell–Johnson–Bhagwati proposition. *Japan and the World Economy* 154: 169–187.
Krueger, A. 1999. Free trade agreements as protectionist devices: Rules of origin. In *Trade, theory and econometrics: Essays in honor of John S. Chipman*, Studies in the modern world economy, ed. J. Melvin, R. Moore, and R. Riezman, Vol. 15. London: Routledge.
Krugman, P. 1991. The move to free trade zones. In *Policy implications of trade and currency zones*, Federal Reserve Bank of Kansas City. Kansas City: Federal Reserve Bank of Kansas City.
Krugman, P. 1993. Regionalism versus multilateralism: Analytical notes. In *New dimensions in regional integration*, ed. J. de Melo and A. Panagariya. Cambridge: Cambridge University Press.
Levy, P. 1997. A political economic analysis of free trade arrangements. *American Economic Review* 87: 506–519.
Limão, N. 2006. Preferential trade agreements as stumbling blocks for multilateral trade liberalization: Evidence for the United States. *American Economic Review* 96: 896–914.
Lipsey, R. 1957. The theory of customs unions: Trade diversion and welfare. *Economica* 24: 40–46.
McLaren, J. 2002. A theory of insidious regionalism. *Quarterly Journal of Economics* 117: 571–608.
Meade, J. 1955. *The theory of customs unions*. Amsterdam: North-Holland.
Mundell, R. 1964. Tariff preferences and the terms of trade. *Manchester School of Economic Studies* 32: 1–13.
Ohyama, M. 1972. Trade and welfare in general equilibrium. *Keio Economic Studies* 9: 37–73.
Ornelas, E. 2005. Endogenous Free trade agreements and the multilateral trading system. *Journal of International Economics* 67: 471–497.
Panagariya, A. 1997a. The Meade model of preferential trading: History, analytics and policy implications. In *International trade and finance: New frontiers for research: Essays in honor of Peter Kenen*, ed. B. Cohen. Cambridge: Cambridge University Press.
Panagariya, A. 1997b. Preferential trading and the myth of natural trading partners. *Japan and the World Economy* 9: 471–489.
Panagariya, A. 2000. Preferential trade liberalization: The traditional theory and new developments. *Journal of Economic Literature* 161: 316–360.
Panagariya, A., and R. Findlay. 1996. A political economy analysis of free trade areas and customs unions. In *The political economy of trade reform: Essays in honor of Jagdish Bhagwati*, ed. R. Feenstra, D. Irwin, and G. Grossman. Cambridge: MIT Press.
Panagariya, A., and P. Krishna. 2002. On necessarily welfare-enhancing free trade areas. *Journal of International Economics* 57: 353–367.
Richardson, M. 1994. Why a free trade area? The tariff also rises. *Economics and Politics* 6: 79–95.
Richardson, M. 1995. Tariff revenue competition in a free trade area. *European Economic Review* 39: 1429–1437.
Riezman, R. 1979. A 3 × 3 model of customs unions. *Journal of International Economics* 9: 341–354.
Saggi, K. 2006. Preferential trade agreements and multilateral tariff cooperation. *International Economic Review* 47: 29–57.
Srinivasan, T. 1993. Discussion on regionalism vs multilateralism: Analytical notes. In *New dimensions in regional integration*, ed. A. Panagariya and J. de Melo. Washington, DC: World Bank.
Srinivasan, T., J. Whalley, and I. Wooton. 1993. Measuring the effects of regionalism on trade and welfare. In *Regional integration and the global trading system*, ed. K. Anderson and R. Blackhurst. New York: St Martin's Press.
Summers, L. 1991. Regionalism and the world trading system. In *Policy implications of trade and currency zones*, Federal Reserve Bank of Kansas City. Federal Reserve Bank of Kansas City: Kansas City.
Vanek, J. 1965. *General equilibrium of international discrimination*. Cambridge: Harvard University Press.
Viner, J. 1950. *The customs unions issue*. New York: Carnegie Endowment for International Peace.
Wonnacott, P., and M. Lutz. 1987. Is there a case for free trade areas? In *Free trade areas and US trade policy*, ed. J. Schott. Washington, DC: Institute for International Economics.
Yi, S.-S. 1996. Endogenous formation of customs unions under imperfect competition: Open regionalism is good. *Journal of International Economics* 41: 153–177.
Yeats, A. 1998. Does Mercosur's trade performance raise concerns about the effects of regional trade arrangements? *World Bank Economic Review* 12: 1–28.

Regional Development

Richard S. Eckaus

Persisting and even increasing regional differentials in the course of national development are important economic phenomena that have continued to stimulate research. Most features of regional development analysis are related to these patterns of unequal progress, especially the studies and modelling of the regional location of particular types of industry, of urbanization and transportation systems, the interrelations of agricultural and industrial growth and interregional and intraregional migration. The grand and overall issue, however, continues to be that of unequal regional development. It is striking that research on this issue, which was in the mainstream of the early years of development economics, has, with only a few exceptions, retired to the regional science journals.

Early empirical work by Williamson (1965) established the perception that national economic growth has been accompanied by increasing regional income differentials for substantial periods of time, followed by a trend toward regional equality. While the first impressions and systematic research were based on the growth experience of advanced countries, subsequent research has confirmed that developing countries have been repeating the previously observed pattern.

Attempts to explain regional growth patterns analytically have always recognized both the potential contribution of international trade theory and its inadequacies. The implications of conventional trade theories applied to regional development are that there would be continual pressures for equalization of regional product prices and factor incomes. Unique mineral and land resources would receive differential rents but the migration of mobile factors to higher income regions would equalize their returns. Practically speaking, it has been recognized that a substantial amount of time might be required for this theoretical prediction to work out. The long lifetimes and slow depreciation of immobile real capital, particularly buildings and civil works, would delay the movement of other factors which could use the capital.

Since standard trade theory cannot explain increasing regional inequality, alternative hypotheses have been generated outside that theory. One of the first of these argued that the existence of different export industries within regions would explain their differential development. Suppose that a new export industry, based on some regionally unique endowments, or the accident of a local improvement in production technology, expanded rapidly. Incomes in the newly exporting region would also expand faster than those in other regions. But persisting regional differences would require some additional conditions, however, or the differences would be eliminated by goods and factor mobility.

Another early hypothesis was that there are, in fact, regional differences in regional production functions, contrary to the assumptions of Hecksher–Ohlin–Samuelson trade theory, and also in the relative prices of non-traded goods. If one or more technologies used in producing goods and services in one region remain persistently inferior to those in other regions, that region would remain permanently poorer. It makes no difference whether or not the ouput is supplied to an export industry. The same result follows if, for any other reasons, the products produced in different regions make more or less effective use of their comparative advantage. Historically, differences in the levels and applications of technology have had a profound role in the explanation or regional incomes. Allocational inefficiencies arising from particular government policies have been widely recognized as retarding national growth and must have analogous implications for growth in particular regions, as their effects are not uniform (World Bank 1983).

Such explanations can all be placed under the headline of ‘structuralist’ economic hypotheses: those which adduce non-economic influences or economic features outside the conventional characterizations to explain why predictions derived from the assumptions of perfect markets will be in error. Although for some economists a resort to explanations of this type is, virtually, ‘the last refuge of a scoundrel’, some of these explanations have become theoretically respectable when given the titles of ‘adjustment costs’, ‘imperfect information’, ‘risk’ and ‘incomplete markets’.

It has often been alleged that regional biases in central government policy in taxing or supplying public services result in differences in regional economic development. This has, for example, been the most conventional explanation of the persisting economic retardation of the Italian Mezzogiorno after national unification and has been argued with respect to the US South after the Civil War, subjects that have attracted a relatively large amount of attention. Differences in regional savings rates or market structures can also work to the same end.

The idea that regional economic activity tends to concentrate in a '*pole de croissance*' is a relatively old one in the regional literature. It has a forerunner in Losch (1939) but was expressed most suggestively in Perroux (1955). The reasons advanced by Perroux included economies of scale in production and the clustering of technological innovations and investment. The descriptions of the process were suggestive but were not successfully embodied in a rigorous theory. Nonetheless, the idea became widely cited and used to justify methods which never became effective tools of regional economic policy, which is why 'growth pole theory' has been called one of the development literature's fads (Higgins 1983).

The use of regional input–output tables as a means of determining the patterns of regional demands due to the expansion of a particular sector is one of the major features of 'growth pole' analysis. The criticism of the use of this tool, as not being able to capture the full range of nonlinear and dynamic interactions which occur in the process of development (Hansen 1972), is really a criticism of the state of regional development theory. Nonetheless, regional input–output tables have helped in understanding some of the characteristic differences that exist and also in assessing the sources and consequences of industrial location decisions and population movements, that have been major preoccupations of the regional development field.

The hypothesis that increasing returns to scale play a major role in explaining diverging regional development is a 'purely' economic argument of long standing and, perhaps, the most powerful. The concept is intuitively appealing. Suppose there are initial differences in the levels of output in one or more sectors of different regions. If the output is produced with increasing returns to scale, that will mean that the sectors with a 'head start' will have lower costs and can effectively capture the entire market, unless constrained by regulation or private agreement. That would appear to lead to a 'cumulative and circular causation process' which favours the advanced region. Unless these are offsetting differences in savings rates or other structural characteristics, the poorer region could not catch up.

Nonetheless, the argument by itself is ultimately unsatisfactory as an explanation of regional differentials. There is nothing which localizes the benefits of increasing returns as long as all produced goods are traded. But, with the addition of another ingredient, the increasing returns to scale hypothesis does become a complete economic explanation of regional differentials. That ingredient, provided by Faini (1984), is the hypothesis that the increasing returns to scale occur in the production of non-traded intermediate inputs. That would mean that final product industries, using those nontraded inputs at falling prices, would have a continuing cost advantage, once an initial differential were created. The migration of capital toward the region with an initial advantage in producing a nontraded intermediate good would only increase the scale of that production and its cost advantage. As a result, the interregional migration of capital would not equalize its returns. Interregional labour mobility would equalize wages but not permit local labour to share in the rents in the nontraded goods sector created by the increasing returns to scale. That would leave the rents to capital in the nontraded goods sector as a permanent source of regional income differentials.

'Shift-shares' analysis, which is another approach to regional development theory, is simply the use of an identity to divide total growth in employment in a region in separate parts: that due to growth in the aggregate of national employment, that part due to overall growth in the particular sectors in a region and that part due to the special comparative advantage of the region itself. The last source is the one of particular interest but 'shift-share' analysis has had relatively few new insights to offer on the subject (Andrikopoulos 1980).

Grand theorizing about regional development has diminished in favour of studies of particular features of the process. The regional science journals provide much interesting reading which is distinguished by its empirical detail. These offer many insights into the processes of regional development, though they cannot be fitted together in an overall framework.

A continuing preoccupation with regional development theory reflects a worthy desire for an operational general theory. In principle, of course, that 'simply' requires an extension of existing economy-wide general equilibrium models or models of the world economy. The major obstacles are not mainly conceptual: most of the elements of a satisfactory theory are well known. The analytical problems are formidable, however, because of their high dimensionality. The addition of regional details to any model quickly makes it analytically cumbersome. International trade theory, for example, cannot easily manage more than two goods and two factors and two regions.

A computational approach to obtain both analytical and empirical insights is possible. Computable international trade models handle much more detail than the analytical models. Some of the sectorally detailed economy-wide computable models already distinguish rural and urban sectors and are embryonic regional models. Conceptually, it is a small step to full regional disaggregation. However, addition of the essential non-convexities of increasing returns to scale would now be a major barrier to computational solutions.

Nonetheless, the major and continuing obstacle to progress in developing overall regional analyses and policy is lack of detailed regional trade data. If that were available, modellers would be hard at work on their computers.

See Also

▶ Inequality Between Nations
▶ Poles of Development
▶ Uneven Development

Bibliography

Andrikopoulos, A.A. 1980. A synthesis of the production function and the shift-share model. *Regional Science and Urban Economics* 10(4): 539–560.

Borts, G.H. 1960. The equalization of returns and regional economic growth. *Economic Journal* 70: 319–347.

Borts, G.H., and P. Stein. 1964. *Economic growth in a free market*. New York: Columbia University Press.

Faini, R. 1984. Increasing returns, non-traded inputs and regional development. *Economic Journal* 94: 308–323.

Fukuchi, T. Growth and stability of multiregional economy. *International Economic Review* 14(1): 509–520.

Hansen, N. 1972. *Growth centers in regional economic development*. New York/London: Free Press/ Collier–Macmillan.

Higgins, B. 1983. From growth poles to systems of interaction. *Growth and Change* 14(4): 3–13.

Lösch, A. 1939. *Die räumliche Ordnung der Wirtschaft*. Jena: Gustav Fischer Verlag. Translated as *The economics of location*. New Haven: Yale University Press, 1954.

North, D.C. 1955. Location theory and regional economic growth. *Journal of Political Economy* 63: 243–258.

Perroux, F. 1955. Note sur la notion de 'pôle de croissance'. *Cahiers de l'Institut de Science Economique Appliqée*, Series D No. 8.

Williamson, J.G. 1965. Regional inequality and the process of national development: A description of the patterns. *Economic Development and Cultural Change* 13(4), Pt 2: 1–84.

World Bank. 1983. *World development report*. Washington, DC: International Bank for Reconstruction and Development.

Regional Development, Geography of

Jeffrey D. Sachs and Gordon C. McCord

R

Abstract

New theoretical work on spatial concentration of industry – particularly the 'new economic geography' – has significantly helped our understanding why some regions develop more than others, why cities arise and where they are located. However, this work rarely incorporates Adam Smith's observation that spatial differences in economic activity also reflect variations in physical geography, which make some places more productive than others at particular times; nor has it accommodated regional development policy – the use of economic incentives to

attract industry to particular locations. A full theory of regional development would integrate theories of agglomeration economies with physical geography and with public economics.

Keywords

Agglomeration; Agricultural productivity; Backward and forward linkages; Cities; Climate; Clusters; Coastal proximity; Comparative advantage; Competitive advantage; Complementarities; Core–periphery model; Demand smoothing; Foreign direct investment; Geography and economic development; Growth poles; Human capital accumulation; Industrial policy; Knowledge spillovers; Learning region; Linkages; Location choice; Monopolistic competition; New economic geography; Nutrition and development; Preferential attachment; Regional development; Returns to scale; Search costs; Spatial concentration; Special Economic Zones (China); Transport costs

JEL Classification

O16

Differences in economic activity across regions have interested economists since Adam Smith, who argued that high overland transport costs in the interior of Africa and Asia 'seem in all ages' to have had hindered economic development. However, economists' attraction to the study of spatial variations in economic activity has fluctuated over time. Standard trade theory based on comparative advantage helps to explain how the location of economic activity is affected by the spatial distribution of primary resources (such as land, labour and water), but standard trade theory says little about the interdependence of location decisions by economic agents, nor does it consider in any depth the more detailed aspects of physical geography (climate, soils, topography, disease epidemiology).

Neoclassical growth models focus on the accumulation of physical, human and technological capital, which individually or together complement raw labour and land as factors of production, but only recent theory (particularly in the work dubbed the 'new economic geography') has begun to grapple with location choices and the spatial concentration of industry (Henderson 1988; Krugman 1991; Fujita et al. 1999). While these newer theories have contributed importantly to our understanding of why some regions develop more than others, and why cities arise and where they are located, they rarely incorporate Smith's observation that spatial differences in economic activity are also related to variations in physical geography, which intrinsically make some places more productive than others at particular points in time. Nor do they yet go into depth on regional development policy, that is, the use of economic incentives to attract industry to one location or another. A full theory of regional development will integrate theories of agglomeration economies with physical geography and with public economics.

Theories of Agglomeration

Economic activity and population around the globe are concentrated in highly dense metropolitan areas, which suggest that there is an important economic benefit of economic agglomeration (spatial co-location of economic agents). Alfred Marshall (1920) suggested that spatial concentration happens because of knowledge spillovers, larger markets for specialized skills, and backward and forward linkages associated with large local markets.

The initial literature to tackle the intractability of modelling economic geography grew from the von Thünen model (1826), which begins with the existence of a city and derives characteristics about land rents and land use surrounding the city; the resulting unplanned, efficient outcome is a concentric ring pattern of production referred to as 'von Thünen cones'. The model doesn't, however, attempt to explain the *raison d'être* of the city itself.

Later models aimed to explain why population and economic activity tend to agglomerate in the first place. Spatial concentration occurs because production is cheaper due to the large amount of nearby economic activity in agglomeration economies. These increasing returns to scale exist for several reasons: larger markets support more highly specialized products; efficiency increases as a large number of producers and consumers allows for less idle time (a source of increasing returns called demand smoothing); economies of scale of intermediate inputs make production cheaper even for sectors without increasing returns; externalities diffuse learning and expertise, as people can see each others' products and work methods; and search costs are lowered when the search process is spatially concentrated. Florida (1995) pioneered the concept of the 'learning region': to minimize transport costs and maximize learning, firms benefit from spatially concentrating their activities, and thus firms looking to augment their capabilities have strong incentive to locate in these learning regions.

New Economic Geography

The 'new economic geography' of recent decades grew from the Dixit and Stiglitz (1977) model of monopolistic competition under increasing returns to scale. Though admittedly a special case, this model became a workhorse in many fields, and a foundation for the new economic geography. The theoretical backbone of new economic geography is the core–periphery model in Krugman (1991), which looks at three effects: the 'market-access effect' (monopolistic firms locate in big markets and export to small markets), the 'cost-of-living effect' (cost of living is cheaper where there are more firms, due to low transport costs), and the 'market-crowding effect' (imperfectly competitive firms look to locate in regions with few competitors). The model was an important step forward in understanding spatial dynamics, but has the downside of being difficult to manipulate analytically and requires numerical simulations (instead of explicit expressions) to derive results.

Another important concept in the location of economic activity is that of clusters, especially in the work of Porter (1990, 1995, 1998a, b). A cluster is a group of interconnected companies and institutions in a particular location (perhaps a city, or a state, or even a group of neighbouring countries). Companies in a cluster benefit from important complementarities, spillovers and a relationship with public institutions, which improve productivity and productivity growth, and stimulate new business formation. The important contribution of this literature is that a firm's comparative advantage (or 'competitive advantage' in the business phrase) can include characteristics outside the firm itself; often geography and location have important implications on how firms or industries can compete in the market.

One of the striking implications of the new economic geography is that spatial concentration arises in a homogeneous region, where there is no fundamental geographical advantage to locating in one place or another. The precise location of firms is accidental. Early advantages in agglomeration can lead to a snowball effect. First movers in regional development can achieve a lasting competitive advantage by attracting other mobile workers and investors. Growth proceeds with 'preferential attachment' to the places that get an early start.

The Role of Physical Geography

In addition to the new economic geography models of agglomeration, a second basic approach seeking to shed light on growth poles and regional development is based on intrinsic geographical advantages. The assumption of homogeneous space is abandoned, and the role of coasts, hinterlands, rivers, mountains and a vast array of other geographical variables is brought to the fore. Adam Smith himself asserted that the division of labour is limited by the extent of the market, so that coastal regions, by virtue of their ability to engage in sea-based trade, enjoy a wider scope of the market than interior regions. More recently, climatic conditions have been found to have pervasive effects on regional development through disease ecology, agricultural productivity, transport costs, vulnerability to natural hazards, water stress and other factors that may affect economic performance.

Several studies (Gallup et al. 1999; Bloom and Sachs 1998) have noted that tropical areas are consistently poorer than temperate-zone areas, and hypothesize that this may be related to the effects of tropical ecology on human health and agricultural productivity. Tropical infectious diseases, for example, impose very high burdens on human health that in turn may lead to shortfalls in economic performance much larger than their direct short-run effects on health. Another study (Gallup and Sachs 2000) found that, after purchased inputs such as capital, labour and fertilizers are controlled for, the average productivity of tropical food production falls short of the productivity of temperate-zone food production. In the course of economic development, this poor performance in food productivity may have had serious adverse effects on nutrition levels, with adverse consequences for human capital accumulation, labour productivity and susceptibility to infectious disease. These geographical penalties can often be compensated by other kinds of interventions (such as malaria control or improved agronomic practices), but, since those interventions require added resources, affected regions may persistently lag behind more fortuitously located regions.

Geographical advantages can trigger subsequent agglomeration based on increasing returns to scale. The agglomeration is then self-reinforcing, even after the initial spatial advantage loses some of its importance. For example, Chicago's port is not as important as when it was the main driver of the city's growth in the middle of the nineteenth century. Glaeser (2005) illustrates that New York's rise in the nineteenth century was due to a technological change that moved ocean shipping from a point-to-point system to a hub and spoke system, and the city's geography made it the natural hub. Today, however, New York's pre-eminence is based not mainly on the port, but on the legacies of the earlier success: finance, business, remarkable infrastructure and the benefits of agglomeration.

Changing Dimensions of Geography

It is important to stress the changing nature of a region's geographic advantage as technology changes. In early civilizations, when transport and communications were too costly to support much interregional and international trade, regional advantage came from agricultural productivity and local transport rather than from access to oceans. As a result, early civilizations almost invariably emerged in highly fertile river valleys such as those around the Nile, Indus, Tigris, Euphrates, Yellow and Yangtze rivers. These civilizations produced high-density populations that in later eras were often disadvantaged by their remoteness from international trade. As the advantages of overland trade between Europe and Asia gave way to oceanic commerce in the sixteenth century and later, and as the trade routes to the Americas were discovered, economic advantage shifted from the Middle East and eastern Mediterranean to the North Atlantic. In the nineteenth century, the high costs of transporting coal for steam power meant that industrialization almost invariably depended on proximity to coal fields.

In the late twentieth century, air transport and telecommunications have reduced the advantages of coastlines relative to hinterlands. The telecommunications sector, in particular, is deeply affecting the global division of labour and the nature of agglomeration economies. The disadvantages of interior and distant regions may well be eased or eliminated by the advances in telecommunications which allow for more disbursed production and new growth poles far from traditional trade routes. It is notable that Bangalore has become a booming centre of global information technology, despite being an inland city in southern India, and despite the weakness of India's roads and ports at the time of Bangalore's ascendancy. The examples of Bangalore and of course California's Silicon Valley show that today's competitive advantage has to do much more with the location of excellent universities and an attractive living environment for highly skilled and mobile information workers, much like the 'learning regions' described by Florida (1995).

Regional Policy Design

The presence of agglomeration economies, increasing returns and clusters suggests that countries can identify areas of potential growth poles and use policy tools and public investment to trigger these processes. Special policy instruments such as export-processing zones and special tax promotion schemes have helped developing countries to establish clusters in textiles and apparel, electronics, consumer appliances, software and automotive components, to name just a few industries where active industrial policy has played a hand. In the case of growth poles in the knowledge economy (such as Silicon Valley and Bangalore), the importance of government support for higher education and R&D and for the creation of science parks is especially apparent. Spillovers from military technology may play a role as well.

It is clear, however, that the successful development strategies of some countries cannot produce the same salubrious results when implemented in very different settings. When China opened some coastal pockets for foreign direct investment, these Special Economic Zones quickly blossomed into vibrant export platforms and created backward linkages with the immediate hinterland. When landlocked Mongolia turned the entire country into a free trade and investment zone in the late 1990s, however, the inflow of foreign capital was a trickle compared with China's experience, and was based mainly on primary commodities (such as copper). Even within China, the coastal provinces in the east have boomed relative to the interior provinces of western China. Physical geography therefore continues to condition economic development. Geographical determinism should be avoided, however; special geographical hindrances may well call for special compensating investments (in roads, disease control, telecommunications, and so on), or for promotion of a judicious choice of industries (those that can be sustained in the face of high transport costs, for example).

Empirical Studies

Empirical evidence supports the idea that economies of scale, agglomeration forces (Davis and Weinstein 1998, 1999; Midelfart-Knarvik et al. 2000; Overman and Puga 2002; Hanson 2005), and backward and forward linkages (Midelfart-Knarvik and Steen 1999) help explain why economic activity clusters together, and that the von Thünen model helps explain economic dynamics near cities (Fafchamps and Shilpi 2003). The traditional core–periphery model has considerable empirical support, given that the core regions of the global economy (particularly North America, Western Europe, and Japan), enjoy ever-increasing levels in productivity. At a smaller scale, studies of wages in the United States and in developing countries show that *ceteris paribus* workers earn much more in urban areas than rural areas, reflecting their higher productivity (Glaeser and Mare 1994; Bairoch 1988).

While looking for the presence of increasing returns to scale yields insights, it does not address the constraints physical geography may place upon economic growth. For example, Adam Smith's observations on the role of access to navigable water still hold. Cross-country empirical research affirms that the level and growth rate of per capita income continue to be strongly positively correlated with geographic variables such as climate and coastal proximity (Gallup et al. 1999; Mellinger et al. 2000), while within-country differences in growth rates in India and China are clearly related to geography as well (Demurger et al. 2002; Sachs et al. 2002). Smith's observations also implicitly underscore the highly favourable economic geography enjoyed by the nations of western Europe. Extensive ocean shorelines host a succession of natural harbours, and numerous navigable rivers penetrate deep into the interior. In addition, despite the large landmass of the United States, 57% of income was generated in counties within 80 km from the coast, though these counties account for only 13% of land mass (Rappaport and Sachs 2003).

Future theoretical and empirical work in understanding regional development should aim to disentangle the forces of differential geography and self-organizing agglomeration economies. Policy studies should examine in depth how regional development policy has been used in the past, and which instruments are particularly important. Economists and business specialists should aim to provide new tools to help specific regions identify appropriate instruments for regional development, including which kinds of industries are likely to flourish in which kinds of spatial settings.

See Also

- ▶ Location Theory
- ▶ Marshall, Alfred (1842–1924)
- ▶ New Economic Geography
- ▶ Spatial Economics
- ▶ Thünen, Johann Heinrich von (1783–1850)
- ▶ Urban Agglomeration

Bibliography

Bairoch, P. 1988. *Cities and economics development: From the dawn of history to the present*. Chicago: University of Chicago Press.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud. 2003. *Economic geography and public policy*. Princeton: Princeton University Press.

Bloom, D., and J.D. Sachs. 1998. Geography, demography, and economic growth in Africa. *Brookings Papers on Economic Activity* 1998(2): 207–295.

Clark, G., M.P. Feldman, and M. Gertler. 2000. *The Oxford handbook of economic geography*. Oxford: Oxford University Press.

Davis, D.R., and D.E. Weinstein. 1998. *Market access, economic geography, and comparative advantage: An empirical assessment*, Working paper no. 6787. Cambridge, MA: NBER.

Davis, D., and D.E. Weinstein. 1999. Economic geography and regional production structure: An empirical investigation. *European Economic Review* 43: 379–407.

Demurger, S., J.D. Sachs, T.-W. Wing, S. Bao, G. Chang, and A. Mellinger. 2002. Geography, economic policy, and regional development in China. *Asian Economic Papers* 1: 146–197.

Dixit, A.K., and J.E. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67: 297–308.

Dunning, J.H. (ed.). 2002. *Regions, globalization, and the knowledge-based economy*. New York: Oxford University Press.

Fafchamps, M., and F. Shilpi. 2003. The spatial division of labor in Nepal. *Journal of Development Studies* 39(6): 23–66.

Florida, R. 1995. Towards the learning region. *Futures* 27: 527–536.

Fujita, M., P. Krugman, and A.J. Venables. 1999. *The spatial economy: Cities, regions, and international trade*. Cambridge, MA: MIT Press.

Gallup, J.L., and J.D. Sachs. 2000. Agriculture, climate, and technology: Why are the tropics falling behind? *American Journal of Agricultural Economics* 82: 731–777.

Gallup, J.L., J.D. Sachs, and A.D. Mellinger. 1999. Geography and economic development. *International Regional Science Review* 22: 179–232.

Glaeser, E.L. 2005. *Urban colossus: Why is New York America's largest city?* Working paper no. 11398. Washington, DC: NBER.

Glaeser, E., and D. Mare. 1994. *Cities and skills*. Working paper no. E94-11, Hoover Institution.

Hanson, G.H. 2005. Market potential, increasing returns, and geographic concentration. *Journal of International Economics* 67: 1–24.

Henderson, J.V. 1988. *Urban development: Theory, fact, and illusion*. Oxford: Oxford University Press.

Henderson, J.V., and J.-F. Thisse (eds.). 2004. *Handbook of regional and urban economics*, vol. 4. Amsterdam: North-Holland.

Krugman, P. 1991. Increasing returns and economic geography. *Journal of Political Economy* 99: 483–499.

Marshall, A. 1920. *Principles of economics*, 8th ed. London: Macmillan.

Mellinger, A.D., J.D. Sachs, and J.L. Gallup. 2000. Climate, coastal proximity, and development. In *Oxford handbook of economic geography*, ed. G.L. Clark, M.P. Feldman, and M.S. Gertler. Oxford: Oxford University Press.

Midelfart-Knarvik, K., and F. Steen. 1999. Self-reinforcing agglomerations? An empirical industry study. *Scandinavian Journal of Economics* 101: 515–532.

Midelfart-Knarvik, K.H., H.G. Overnman, S.J. Redding, and A.J. Venables. 2000. *The location of European industry*, Economic papers no. 142. Brussels: European Commission Directorate-General for Economic and Financial Affairs.

O'Flaherty, B. 2005. *City economics*. Cambridge, MA: Harvard University Press.

Overman, H.G., and D. Puga. 2002. Unemployment clusters across European regions and countries. *Economic Policy* 34: 115–147.

Porter, M. 1990. *The competitive advantage of nations*. New York: Free Press.

Porter, M. 1995. The competitive advantage of the inner city. *Harvard Business Review* 73: 55–71.

Porter, M. 1998a. Clusters and competition: New agendas for companies, governments and institutions. In *On competition*. Boston: Harvard Business School Press.

Porter, M. 1998b. The microeconomic foundations of economic development. In *Global competitiveness report 1998*, ed. World Economic Forum. Geneva: WEF.
Rappaport, J., and J.D. Sachs. 2003. The United States as a coastal nation. *Journal of Economic Growth* 8: 5–46.
Sachs, J.D., N. Bajpai, and A. Ramiah. 2002. Understanding regional economic growth in India. *Asian Economic Papers* 1(3): 32–62.
Saxenian, A.L. 2006. *The new argonauts: Regional advantages in a global economy*. Cambridge, MA: Harvard University Press.
von Thünen, J.H. 1826. *Der Isolierte Staat in Beziehung auf Landtschaft und Nationalökonomie*. Hamburg. Trans V.M. Wartenberg as *Von Thünen's isolated state*. Oxford: Pergamon Press, 1966.

Regional Distribution of Economic Activity

Barry Moore and John Rhodes

Introduction: Regions and Nations

Our starting point for discussing the regional distribution of economic activity within nations is the recognition that economists interested in the emergence and persistence of economic disparities between regions have drawn heavily on theories relevant for understanding international economic relations between countries. There are, of course, important differences between regions and sovereign nation states, but they are often differences of degree rather than kind. Critically important for a region is its openness and interdependence with other regions. This arises because of the generally greater importance of trade to the region compared with the nation but also because of its greater dependence on investment brought in from other regions. Regions within a nation also share a common currency and this removes the possibility of currency adjustments. Inevitably this puts more weight on internal price flexibility and factor mobility between regions as means by which regions adjust to economic circumstances than is the case with nations. The mobility of labour and capital is greater between regions than countries and as we shall see this has been a major route by which regions adjust to change.

Limitations of the Neoclassical Approach

Perhaps the most influential of the theories of international trade which regional economists have used to explain the regional distribution of economic activity is that grounded in the neoclassical approach. A central conclusion of this theory is that each region will tend to specialize in producing what it is comparatively good at producing, even though it may be inferior (or superior) in producing *all* goods and services. This theory is based on a number of highly unrealistic assumptions, most notably the absence of factor mobility between regions and perfect internal mobility of resources; perfect markets – so that prices are perfectly flexible and markets clear; the absence of increasing returns to scale; perfect foresight on the part of economic agents. Not surprisingly, neoclassical theory has been severely criticized as a theoretical framework for understanding how regional economies function.

When applied to regions, the neoclassical approach is flawed for several reasons. Firstly, inter-*regional* labour and capital mobility is a key feature of regional adjustment processes. This observation not only contradicts one of the underlying assumptions of the neoclassical approach, it implies that a region's growth is, to a significant extent, demand determined where factor supplies adjust to demand via inter-regional mobility.

Secondly, factor price flexibility within each region is a necessary condition to ensure specialization of production and full employment in each region. The reason a region can specialize in the production of goods in which it has a comparative advantage, but not an absolute advantage, is that its real wages are much lower than in other regions. If real wages are not flexible and are at the same level in all regions then production will tend to concentrate in the high productivity regions. Workers in turn will either migrate from the low to high productivity regions in search of jobs or, in the absence of migration, unemployment will result in low productivity regions. In other words, real factor price flexibility is critical for the theory of comparative advantage as an explanation of the regional distribution of economic activity. Real factor returns must match the level of factor productivity region by region.

Perhaps a more fundamental criticism of the neoclassical approach to explaining regional economic activity is that it assumes instantaneous clearing of the labour market, or that wage adjustment and labour mobility are a feature of both the short and long run. Thus in the short run, regional wage flexibility clears the labour market which then induces inter-regional movements of labour and capital, until eventually (in the long run) a new equilibrium is reached. Unfortunately for the neoclassical approach, there is no guarantee that such a new equilibrium will be reached (Canning 1985). The argument is quite simple. If real wages are flexible then they will fall in regions experiencing unemployment until full employment is reached. Low relative wages will attract firms from other regions and workers will seek jobs in other regions. The result is an excess demand for labour which eventually equalizes real wages. In other words the economy must be at over-full employment. However, over-full employment induces real wages to rise above their equilibrium level, relative to other regions, and as a result unemployment starts to rise again, eventually driving real wages down again. Thus when the dynamics of adjustment are examined no simple neoclassical answer to the question of regional growth emerges.

Real Wage Rigidities and Absolute Advantages

In most countries little evidence exists that real wages are sufficiently flexible to secure full employment or meet the conditions required for the theory of comparative advantage to provide an explanation of the regional distribution of economic activity. Quite apart from the role that labour unions play in setting national wage rates across all regions, the potential mobility of skilled labour will tend to equalize wages across regions (a point that Hicks made over half a century ago). Thus when internal labour markets are considered, firms are forced to pay the national (or even international) going wage to avoid losing their internal labour force. In those circumstances where real wages flexibility is small, the regional location of economic activity will be determined by the principle of absolute advantage rather than comparative advantage. Regions with an absolute advantage will grow relative to regions with an absolute disadvantage. Firms will move to areas with an absolute advantage and existing indigenous firms will expand as they exploit their absolute advantages and out-complete firms in other regions. If interregional migration is relatively easy, the workforce will also concentrate in these regions to satisfy their growing demand for labour. If inter-regional migration is difficult, perhaps because of housing constraints or union activity, the growth of the favoured region will be supply constrained and under-utilization of labour and other resources will persist in the disadvantaged region.

Absolute advantage may be enhanced if we admit the possibility of increasing returns to scale. In these circumstances, regions with small producers will face an increasingly uphill task in competing because their market penetration does not enable them to secure sufficient internal economies of scale.

The Competitiveness of a Region's Export Base

This analysis strongly suggests that the regional distribution of economic activity depends to a significant extent on the growth of demand that regions face rather than on the principle of comparative advantage implied by neoclassical theory. Neither is the growth of a region limited by labour or capital shortages since the availability of these two factors is constrained only insofar as there are barriers to inter-regional factor mobility. Of critical importance is the growth of demand that a region enjoys, particularly for the goods and services which it exports to other regions and to other nations. Regional economists have long recognized the importance of the traded goods sector in explaining regional growth. Manufacturing activity is traditionally regarded as the most important of the basic sectors but agriculture, mining, tourism and certain producer and financial services are also significant in the export sector of some regions. Other sectors such as construction, population-related services, transport and communications are largely dependent on income generated by the export sector.

The critical question therefore is 'what determines the growth of a region's export sector?'. Regions favourably endowed with climate and geological conditions will tend to specialize in the production of raw materials and commodities in which they have an absolute advantage. Once a region's specialization is established, the immobility of key factor inputs implies that changes in exports become important in determining the growth of demand for these outputs. The growth of output will however be constrained by the region's opportunities for increasing crop yields or mining productivity.

The role of manufacturing industry in the process of regional growth is more complex than the role of raw materials and agricultural output. In the first place, it is important to recognize that the export base depends on both the location decisions of manufacturing firms and the competitive performance of the indigenous sector (existing firms and newly spawned companies). With respect to existing firms, regional economists have pointed to the inherited structure of the region's industry in explaining regional differences in economic growth. Thus changes in the pattern of demand nationally and internationally, changes in production technology, and changes in the organizational structure of firms will influence different industries in quite diverse ways. How individual regions fare in this process of change will depend critically on which industries they possess and their importance to the region. In this sense it is difficult to deny that 'structural' factors must play some role in determining a region's economic performance. However, it is also clear that a region's growth might influence its structure. Structure and growth are interdependent.

Even when allowance is made for structural differences across regions at a relatively high level of disaggregation, significant differences in the regional growth performance persist. Differences in unit input costs between regions, combined with the locational flexibility permitted by improvements in transport and communications are one possible explanation. There are in many countries significant regional differences in operating costs, notably wages and transport costs. Such cost differences might be expected to lead to changes in a region's net export growth as well as to changes in location (investment) decisions in favour of relatively low cost regions.

There has also been a long term declining trend in the employment/floorspace ratio. This means that manufacturing firms located in urban centres which were established in earlier periods of industrialization, no longer find such locations appropriate or necessary. Expansion is often difficult in highly urbanized areas and non-manufacturing firms are willing to pay more for an urban location than manufacturing firms. As a result of this, regions dominated by urban concentrations of manufacturing are experiencing a relative decline as firms decentralize their activities to other more rural/small town regions where costs are lower and suitable industrial floorspace more easily and cheaply acquired.

This process of decentralization by manufacturing firms (observed in many advanced industrialized countries) is paralleled by population dispersal. It is difficult to judge the extent of cause and effect in this case, however, since there is evidence of a change in residential preferences in favour of small cities and towns, which in turn could influence the location decisions of firms. This is particularly so in the residential choices being made by professional and scientific workers. It is also clear that relocation decisions by firms provoke changes in the residential locations of households. Whatever the precise balance of cause and effect between population and employment dispersal from large cities, it is clear that regions are growing differentially, at least in part, as a result of their urban/rural structure.

Changes in the regional distribution of certain manufacturing industries have been influenced by the growth of large multi-regional, multinational corporations. A region's fortunes no longer depend merely on trends in its dominant sectors at the national and international level. Rather, companies are selecting locations according to the functional divisions of the company. Mass production and assembly activities will be drawn to regions with a surplus of unskilled or semi-skilled labour, where wages are relatively low and where female workers are more readily available. Control functions, product research and development are more typically located in or close to larger cities in the more favoured regions. Apart from the obvious implications for the spatial division of labour and regional growth it is important to recognize that large corporations are now able to organize their regional allocation of resources in a way that maximizes their advantages but at the cost of increasing uncertainty in the regions. In other words the process of industrial restructuring is now proving of critical importance to the emerging pattern of regional economic activity.

Closely related to the above notion of a regional functional division of labour is the concept of the product life-cycle whereby different regions are characterized by different points in the life-cycle. Initially, during the research and development phase, uncertainty requires good communications and flexibility in decision making afforded by proximity to key management personnel. Once the innovative monopoly advantage is exhausted and product development has largely ceased, cost advantages associated with internal economies of scale and the use of less skilled labour become more important. These advantages are to be found in regions with adequate space for high capital intensity or abundant supplies of low skilled labour. In this respect the product life-cycle process maps in closely with earlier observations on the significance of regionally differentiated input costs, the urban/rural shift of manufacturing activity and the functional division of labour across regions. Insofar as regions cannot maintain their competitive advantage at different stages in the product life-cycle they will suffer a loss of export markets, increased import penetration and relative economic decline. This pattern of events seems characteristic of many old industrialized regions in countries that industrialized early. Such regions often relied heavily on products such as steel, coal and textiles for their early economic development. These regions now face a major competitive threat from overseas, where low labour costs, higher productivity and greater economies of scale are securing significant price competitive advantages. Other industries such as engineering and motor vehicles are also increasingly vulnerable to such competitiveness as are the regions which depend on these industries to any significant extent.

Cumulative Regional Growth and Decline

The discussion above points to a number of factors important in the changing net export performance of a region. Recent theories of cumulative causation, espoused initially by Myrdal (1959) and developed later by Kaldor (1971), attempt to provide a view of regional growth which emphasizes both the role of net export growth and the dynamic and interdependent processes of regional development. Of central importance to this view is the denial of comparative static models with their equilibriating tendencies resulting from the free play of market forces. Rather, cumulative causation models raise the possibility of increasing concentrations of economic intensity in favoured regions.

In cumulative causation models once growth becomes firmly established in a region, a virtuous circle is established whereby the participants enjoy the advantages of external economies associated with innovation, skill development, and an entrepreneurial culture which feeds on itself, generating yet greater externalities and dynamic economies of scale. At the same time the need for public infrastructure, public services and housing gives an added stimulus to the virtuous circle of growth. By contrast in declining areas the economic and social structure stagnates as population leaves and industries decline and wither away. Within this model export growth can provide the stimulus whereby faster output and productivity growth, greater product and process innovation and greater investment are secured. In particular Kaldor's (1966) incorporation of the Verdoorn relationship linking output growth and productivity growth is an important feature of the cumulative process whereby faster output growth gives rise to faster productivity growth, which in turn increases export competitiveness which feeds back into faster output growth and so on. Finally, although balance of payments disequilibrium is not manifest in terms of (for example) a growing deficit in a region's current account, it is manifest through a slower growth of real income and net emigration. The opposite occurs in regions enjoying cumulative growth.

Although much of the theoretical literature has emphasized the importance of net exports and private sector investment as sources of exogenous demand to a region, it is important to note that public expenditure has been of growing importance in the post-war period. There are several ways in which public expenditure supports economic activity differentially across a nation's regions. Firstly, there are net fiscal transfers to regions experiencing above average unemployment through unemployment benefits and lower tax revenues from regional expenditure and income. It is in this way that the implicit current account deficit of regions is often largely financed by surplus regions. Secondly, the provision of public sector infrastructure (including housing) may be important particularly in regions which are in the early stages of economic development. Thirdly, public sector procurement policies, related for example to a nation's defence expenditure, very often have pronounced regionally differentiated effects. The latter are also significant for regional development in that they will tend to encourage technology transfer between the public sector and private firms, thereby reinforcing the region's growth potential.

Regional Differences in Structural Adaptation

One important question which remains partly unanswered in our discussion so far is why structural adaptation of regions varies. Some regions are apparently able to adjust relatively easily to changing competitive conditions but others experience severe difficulties. Models which emphasize absolute advantage rely heavily on relative cost disadvantages as the major difficulty facing companies in satisfying export demand whilst non-cost factors operating on the supply side are largely ignored. The cumulative causation models include an endogenous supply response to demand but little is said about factors constraining the supply of inputs when the pattern of demand changes. In this respect there is evidence for regionally differentiated process and product innovation, with the latter of particular importance as a factor underpinning the success of the more rapidly growing regions. The rate of new firm formation is also important in securing a flexible supply response. The size structure of firms has been shown to be related to the rate of new firm formation, with regions dominated by small firms spawning new firms faster than regions dominated by larger firms.

The Contribution of Regional Policy

Regional policy is designed to influence the geographical distribution of economic activity either by restraining economic development in fully employed regions and/or by encouraging expansion in problem regions through financial inducements to the private sector or a differentially high rate of public expenditure to improve the infrastructure. Such policies have been followed in the majority of OECD countries from the 1960s onwards.

The operation of such policies presented new challenges on how to disentangle their impact in promoting economic development in backward regions from all the other factors causing changes in the regional distribution of economic activity. Until this was possible the cost effectiveness of regional policies could not be properly assessed. Methods of evaluation based upon a variety of quantitative and qualitative techniques are gradually being developed.

The results of this work suggest that regional policy had a substantial influence on the regional distribution of economic activity, particularly in the 1960s and early 1970s, when the more advanced industrial economies were growing relatively quickly. The impact of policy declined, however, with the onset of slower growth and recession from the mid-1970s onwards. In the case of Britain, the most recent evaluation of policy suggested that it had contributed about half a million new jobs to the assisted areas during 1960–1981 – a very significant contribution.

In recent years, however, there has been growing dissatisfaction with regional policy and increasing pressure for reforms which would render policy more appropriate to the needs of backward regions and the economic conditions prevailing in the 1980s. Dissatisfaction arose in part from the seeming inability of policy to fully solve the economic problems of backward regions in which the policy-induced development fell far short of the needs of regions with large declining sectors. As the 1980s recession emerged, this gap between the needs of the problem regions and what policy could achieve became progressively wider.

The criticism was also expressed that regional policy had diverted too many branch plants into the problem regions, thus making those areas more vulnerable to subsequent closure in times of recession. In addition, such policies did not encourage decision-making in the region or enhance indigenous development which would lead to self-sustained growth from within the area. Branch plant economies, in addition to discouraging the development of entrepreneurship and the start-up of new small firms, also distorted the occupational structure towards lower skilled manual workers, leaving those with technical, professional and managerial skills to seek jobs in more prosperous regions.

A further point of contention raised by the critics of regional policy as pursued in the last two decades is that too much emphasis has been placed on attempting to influence the geographical distribution of manufacturing industries, which have been in decline or growing relatively slowly, and too little emphasis has been placed on attempting to encourage service industries to move to the depressed regions. This apparent imbalance in policy stance, which has only partially been rectified in recent years, was justified on two grounds. Firstly, most service industries depend on local income and population and therefore do not constitute part of the export base of a region. Secondly, those service industries which serve a wider regional, national or international market, such as financial and business services, are limited in their choice of location by the need for face to face contact with each other and with other national institutions and the need for instant information which can only be found in the larger capital cities.

Another issue concerning the efficacy of policy is the choice of regional policy inducements to private firms, between those which subsidize the use of labour and those which subsidize capital. Neoclassical theory suggests that labour subsidies should be favoured on the grounds that it is the use of the factor of production in excess supply which should be subsidized (i.e. labour) and not the factor of production which is in scarce supply (i.e. capital). Empirical work however suggests that capital subsidies are the more cost effective. Labour subsidies have to be applied continuously, year after year, to maintain jobs which have been created and they rely for their impact on marginal changes in costs and prices and the subsequent responses of those who purchase the region's output. Unless the labour subsidy is large and is passed on to the buyer in the form of lower prices, its impact on competitiveness and the volume of orders is limited. Capital subsidies, on the other hand, can influence company decisions at crucial times when large capital investment projects are being planned, and can influence their location in favour of high unemployment areas, thus capturing with one grant a relatively large number of new jobs for the entire length of the project's life which may be ten or twenty years.

Partly in response to such criticisms and partly in the search for improvements in the cost effectiveness of policy and the need to cut public expenditure, central governments have tended to reduce the priority afforded to regional policies in the 1980s, just at a time when the economic problems of backward regions have worsened. As a consequence, local authorities have become more active in the promotion of economic development in their areas, which encourages the 'leap-frogging' of financial inducements offered, leading to wasteful competition between authorities. The authorities with most natural advantages of location and environment tend to have most success.

Meanwhile the search continues for a regional policy which could be expected to meet the needs of problem regions more effectively. Regional Development Agencies form the basis of one set of proposals. Following the precedents of New Town Corporations and the Scottish and Welsh Development Agencies, such bodies can coordinate the activities of public and private sectors, raise funds from both public and private sources, assess local and regional needs more carefully and with more commitment, and generally act as a catalyst for indigenous growth and eventually self-sustaining development. The emphasis is thus shifting more towards increasing the amount of supply-side flexibility to improve a region's export competitiveness.

There have also been calls for improved monitoring of the role of the public sector in causing or reducing regional disparities through its own procurement and expenditure policies. But no type of regional policy will solve the problems of depressed areas effectively and efficiently unless the major industrialized economies return to an era of rapid growth and full employment.

See Also

- ► Location of Economic Activity
- ► Poles of Development
- ► Spatial Economics

Bibliography

Armstrong, H., and J. Taylor. 1978. *Regional economic policy*. Oxford: Philip Allan.

Borts, G. 1960. The equalization of returns and regional economic growth. *Economic Journal* 50: 319–347.

Canning, D. 1985. *The dynamics of regional wage adjustment*. Cambridge, UK: Department of Applied Economics.

Dixon, R.J., and A.P. Thirlwall. 1975. A model of regional growth rate differentials along Kaldorian lines. *Oxford Economic Papers* 27(2): 201–214.

Isard, W. 1960. *Methods of regional analysis: An introduction to regional science*. Cambridge, MA: MIT Press.

Kaldor, N. 1966. *The causes of the slow rate of economic growth of the United Kingdom*. Cambridge: Cambridge University Press.

Kaldor, N. 1970. The case for regional policies. *Scottish Journal of Political Economy* 17(3): 337–348.

Moore, B., J. Rhodes, and P. Tyler. 1986. *The effects of government regional economic policy*. London: HMSO.

Myrdal, G. 1957. *Economic theory and underdeveloped regions*. London: Duckworth.

Regional Economics

Ronald J. Krumm and George S. Tolley

The major problem of regional economics, which is to explain the location of production and population within a national economy, has often been approached as if it were not amenable to the usual tools of economic analysis. That regional economics is a distinct field can be seen by comparing the similarities and differences between a region and a nation. A region is like a nation in that goods are traded between it and the rest of the world. Locational differences in factor supplies lead to differences in goods produced among regions and hence to trade among regions, just as with nations. The similarities extend partly – but not wholly – to behaviour of the factors of production. Capital is a factor of production displaying mobility among nations and regions.

One of the most important differences between a nation and a region is with regard to another factor of production, labour. Labour is restricted from moving freely among nations, whereas the *sine qua non* of regions within a nation is that labour can move without restriction among regions. Beyond this crucial difference are differences in policy instruments. Unlike a nation, a region cannot undertake independent monetary and trade policies.

The differences make international trade and regional economics distinct from one another. In contrast to the nation, the demand for goods in a region depends on the endogenous number of people in the region. A determinant of demand as well as what is produced in the region is labour supply which is affected by degree of labour mobility, local goods demand and supply determined cost of living differences among regions, and amenities in the region. Regional economics is very much concerned with the resulting effects on the location of industry. The major policy concerns in regional economics are with effects of local taxation and expenditure policies and with efforts to influence the distribution of activities among regions. All this is in contrast to the concerns in international trade with tariffs, exchange rates and monetary policies not found in regional economics.

While regional economics has been relatively neglected, it has led other fields of economics in distinguishing between traded and nontraded goods, a distinction only recently of importance in international trade literature but long made in local area multiplier analysis. Furthermore, regional economics has recognized more fully than the rest of economics that the use of goods as intermediate inputs is fundamental to production, helping determine the location of economic activity. These and other strands of regional analysis are discussed below, followed by consideration of a comprehensive framework for analysing regions.

Received Regional Analysis

Early Location Theory

An early lasting contribution is Von Thünen's (1826) theory of agricultural land use around a city. Transport cost savings near a city lead to higher bids for land and thus more intensive land use. Formally this model is the same as used in the 20th century to explain land use within a city as a function of distance from the central business distinct. Von Thünen did not provide a theory of regions because the demand for agricultural products emanating from the city was not explained. However, he stressed competition for land in an area, an ingredient missing in most attempts to understand regions.

Weber (1929) attempted to explain the spatial location of industry as the choice of production location which minimizes input hauling costs. Weber did not explain how much production will occur at each location and ignored output price as an influence on location choice. Hoover (1937, 1948) modified the Weberian framework by allowing for a more complicated transportation cost structure, substitution of inputs and economies of scale in production. Isard (1956, 1975) added substitution between transportation and other inputs in production. Moses (1958) considered input substitutions more fully and showed that the optimum location is sensitive to the homogeneity characteristics of the production function. An unexplored implication of these analyses is that the value of certain locations is greater than others and that firms minimizing costs will compete with one another for a location leading to land rents as in the Von Thünen model which would be a further influence on location choice.

Losch (1940) extended industry location analysis to more than one industry, analysing sizes and types of cities with no resource differentials among regions, with population distributed uniformly, and with input goods available at all places in perfectly elastic supply. Spatial concentrations of production result from differences in scale economies and costs of transportation of finished goods. Mills and Lav (1964) showed that Losch's hexagon-shaped market areas are inappropriate and that free entry of firms need not result in market areas that completely exhaust all space.

Losch, like Von Thünen, Weber and Hoover did not recognize the effect of production on local labour demand and hence on the geographic distribution of population and the demand for output. Beckman's (1958) extension of Losch's work allows demands for local goods to influence city sizes in a hierarchy but is driven by uniformly distributed agricultural production and does not allow for serious analysis of production location.

Economic Activity in a Single Region

North (1955) exemplifies the export base approach to regional analysis which is concerned with induced effects on a region deriving from the existence of so-called basic industries which export their output out of the region. The induced effects depend on additional employment attracted to the region due to the fact that demands of workers and firms in the basic industries can be partly satisfied by local production. The effects lead to a local employment multiplier. The reasons for the location, size and type of export industries to begin with are not explained in the export base approach, nor is the rule of prices considered. On the other hand, the approach provides a beginning for considering regional demand endogeneity not found in the industry location work discussed above.

The export base approach provides a rationale for a larger body of practically oriented work quantifying the regional effects of (unexplained) increases in exports out of a region, culminating in input–output models which adapt Leontief inter-industry analysis to regions as exemplified by Isard (1975).

More Recent Advances in Regional Analysis

Borts and Stein (1964) took a long step forward in considering flows of factors of production with analysis of labour, capital and labour–capital ratios applying tools from the mainstream of economics. Labour supply shifts are the major exogenous movers in their model stemming from an emphasis on high birth-rates and technological change in agriculture acting to eject labour from agriculture in different regions, particularly the South, during the decades just before and after World War II. The analysis does not bring in regional comparative advantage in production or amenities affecting where labour wishes to locate.

More recently, some urban models have recognized the fact that migration occurs between cities and have posited that what is truly exogenous to an area is the utility level necessary to induce people to locate in the region. Tolley (1974) developed a model of an area in which there is consumption and production of goods that are traded between cities, as well as non-traded goods for which area production equals area consumption. Differences in costs of production and demand for non-traded goods were shown to result in unequal money wages between areas even though utility is equalized. Henderson (1974) addressed the existence of cities and hierarchies of cities in view of economies of scale and population mobility, and Upton (1981) further analysed equilibrium city sizes. Tolley et al. (1979) examined the effects of externalities on city size, and Henderson and Ioannides (1981) applied similar principles to growth and change in systems of cities.

The strands in the contributions that have been cited provide the beginnings of a conceptual framework for investigating regional growth and decline phenomena, including the empirical issues brought up in Muth (1971), Engle and Hutchins (1978), Howland (1979), and Schmenner (1982), many of which are as of yet unresolved. In the next section we expand on these contributions in order to present a more comprehensive framework

Ingredients of a Comprehensive Theory of Regions

Production

Let the production of good i in a region be given by $x_i = F_i(\mathrm{X}_i, N_i, \mathrm{L}_i, \mathrm{A})$, where $\mathbf{X}_i$ is a vector of intermediate goods used in production of x_i, N_i is labour, $\mathbf{L}_i$ is location-specific capital including land, mineral deposits and port sites, and $\mathbf{A}$ is a vector of location-specific amenities. Amenities include climate characteristics and differ from region-specific capital in that they are public goods for all firms and house-holds in the region. Non-region-specific capital, also an argument of the production function, is suppressed here with the idea that it exhibits little price variation among regions.

Given the vector of goods prices, $\mathbf{p}$, and the labour wage, W, the region supply curve of any x_i is given by

$$x_i^s = f_i^s(\mathbf{p}, W, \mathbf{A}) \tag{1}$$

R

where the entire vector of prices rather than simply own price, p_i, enters because of the use of goods as inputs whose prices affect the profitability of producing x_i. Accompanying the supply curve are demands for each good used in production of x_i

$$x_i^{\mathrm{dd}} = f_{i,j}^{\mathrm{dd}}(\mathbf{p}, W, \mathbf{A}) \tag{2}$$

and the demand for labour to produce x_i

$$N_i^{\mathrm{d}} = N_i^{\mathrm{d}}(\mathbf{p}, W, \mathbf{A}). \tag{3}$$

Equations (1), (2) and (3) represent the producer part of regional theory giving producer responses to prices, wage and regional amenities. Amenities will be taken here to be exogenous, though in extensions effects of output on pollution and congestion and effects of government on local services treated as public goods could be considered as making amenities endogenous. Given the amenities, prices $\mathbf{p}$ and the wage W are determined in an equilibrium depending not only on the producer behaviour conditions but also on product demand behaviour and labour supply behaviour, which are the other parts of the regional theory to be considered in later subsections.

The producer behaviour conditions have in common with regional input–output models that goods may be used as intermediate inputs. However, in input–output models the goods are used in fixed proportions, and the amount exported out of the region or some other concept of final demand is taken as exogenous. In contrast, a major purpose of the producer behaviour conditions here is to offer a price-theoretic explanation of how inputs and outputs are determined.

To derive Eqs. (1), (2) and (3), multiply the marginal product of each input, obtained by partial differentiation of the production function, by output price p_i and set equal to input price which is an intermediate good price p_j, wage W, or price of location specific capital R depending on the input being considered. Together with the production function the result is $Q + 1$ equations determining output x_i and the Q inputs used in its production. Total demand for each intermediate input is the sum of the demands in production of each x_i. Finally, a supply function may be posited for location-specific capital, $L_i = L(R_i)$, showing how its price is endogenous raised as more of it is demanded, adding one more equation to the system.

While the solutions for L_i and R_i are not shown, they influence Eqs. (1), (2) and (3) since they are part of the same equation system. Consider, for example, the response of regional output of x_i in Eq. (1) to an increase in output price p_i. An increase in p_i would raise the marginal revenue product of all inputs leading to indefinite increase of output if all input prices remained the same. But, as region-specific capital L_i is expanded its price R_i rises leading to a rise in the marginal cost of producing x_i, stopping the increase in output when the rise in marginal cost has come to equal the increase in output price. A reason why supply curves of regional outputs slope upward is seen to be that there are upward sloping supply curves of location-specific capital, even though price and quantity of the region-specific capital have been substituted out in the solutions and thus do not explicitly appear in the regional producer behaviour conditions.

Some of the region-specific capital consists of land at various distances from production and consumption sites within the region. As demand for access to these sites grows people and firms will either have to locate more densely or locate further away from the sites. Rising access costs within the region are one of the regions for upward sloping supply curves for regional outputs. Access within an urban economy has been a subject of much urban economics analysis and is a feature linking urban and regional analysis.

Labour Supply

If people can choose without restriction where within a nation to live, a tendency can be expected for labour to respond to any wage differentials among regions that would permit a bettering of well-being from moving. In equilibrium such differentials will be eliminated so that the wage in each region will equate utility of a marginal worker in the region to that attainable elsewhere. The nominal wage will be such that

$$\ln W = \ln W_0 + \mathbf{k}_{\mathbf{x}}''[\ln(p/p_0)] - \mathbf{k}_{\mathbf{A}}'(\mathbf{A}/\mathbf{A}_0) \quad (4)$$

where W_0 is the nominal wage available in the rest of the economy, $\mathbf{k}_{\mathbf{x}}$ is a vector of expenditure shares on elements of $\mathbf{x}$, $\mathbf{p}_0$ is the vector of market good prices in the rest of the economy, $\mathbf{k}_{\mathbf{A}}$ is a vector of the marginal evaluation of each amenity, and $\mathbf{A}_0$ is the set of amenity levels available in the rest of the economy. This specification indicates that the nominal wage differential between one region and any other depends on differences in the comprehensively defined cost of living consisting of the usual weighted sum of market or private goods prices plus the amounts of amenities or public goods weighted by marginal valuations attached to them.

Given W_0, p_0, and A_0, the wage W at which labour is supplied to a region is increased by higher market good prices contained in $\mathbf{p}$ and lower amounts of amenities contained in $\mathbf{A}$. This condition still holds if some degree of labour immobility is allowed for, introducing N on the right-hand side of the labour supply equation and not affecting the formal solution for region equilibrium.

Demand within the Region

Turning to product demand in the region, a part of the demand for regional output already considered is the demand for goods as intermediate inputs in production as given by Eq. (2), to which must be added demands by households in the region.

The amount of each good demanded per household in the region, based on received consumer theory, is a function of the goods price vector $\mathbf{p}$, the wage W as a determinant of household income, and amenities $\mathbf{A}$ in view of their complementarity or substitutability with market goods. Multiplying per household demand by the number of households, demand for the ith good by households in the region is

$$x_i^{\mathrm{d}} = d_i(\mathbf{p}, W, \mathbf{A})^* N \quad (5)$$

which presumes strict proportionality between employment and population and between the nominal wage and nominal earnings with no non-wage income influences on demand. Extensions could introduce behaviour determining fraction of the population in market employment and family size, as well as property income which is particularly important to where retirees and others not receiving wage income live.

Supply and Demand from Outside the Region

The price of any good supplied from outside the region will be the price elsewhere in the economy plus the cost of transporting it into the region. Similarly the price received for a good exported to satisfy demands outside the region will equal the price elsewhere less the cost of transporting it to the destination outside the region. Possible values of each p_i are bounded below by the export price and above by the import price, since producers will send supplies of the good outside the region rather than sell below the export price and demanders will import rather than pay more than the import price.

Goods may be divided into two categories: (a) traded goods, whose prices are at one of the two bounds and hence are either exported or imported and (b) nontraded goods, whose prices are between the bounds.

Region Equilibrium

For traded goods any difference between the amount demanded and supplied of a good is taken up by exports or imports. Equality between demand and supply within the region is not a condition of equilibrium. Moreover, the prices of traded goods may be treated as a first approximation as being exogenous. For nontraded goods, on the other hand, quantities demanded and supplied within the region must be equal. The prices of nontraded goods are thereby endogenously determined, occurring at the intersection of demand and supply within the region.

Let the price vector $\mathbf{p}$ be separated into two parts, one containing prices of traded goods and the other containing prices of nontraded goods. One subset of the system of equations determining regional equilibrium consists of the supply Eq. (1) pertaining to nontraded goods. Another subset consists of the demands for each nontraded good. The demand for a nontraded good is obtained by summing the demand for it as an intermediate input, which is to say summing Eq. (2) over all j, and adding to this sum the household demand for the good given by Eq. (5).

To the foregoing equations for supply and demand of nontraded goods are added two additional equations, one for the demand for labour consisting of the sum of the labour demands given by Eq. (3) over all traded and nontraded goods produced in the region and the other for the supply of labour which is given by Eq. (4). The result is a system of $2R + 2$ equations which is the basic regional system determining the prices and quantities of the R nontraded goods, the amount of labour employed in the region, N, and the region wage rate, W. Given the solution of these equations the resulting local goods prices may be used in the supply curves for traded goods to calculate the production of the traded goods.

Illustration: Export Trade-Induced Growth

As an illustration giving giving the flavour of adjustments implied by the framework, consider the response in a region to a rise in price of an export good. In addition to the initial expansion moving along the supply curve for the good as determined by Eq. (1) leading to induced demand increases because of the appearance of N in the household demand Eq. (5), costs and hence prices of other goods in the region using the exported good as an input would rise, leading to a decline in production of these other goods and an associated decline in employment partly offsetting the initial rise in employment, along with a rise in the wage because of the higher cost of living due to the higher prices. The rise in the wage would in turn affect the costs and hence prices of all goods produced in the region, leading to still further feedbacks. In the process, some goods could change as between being traded or nontraded. The example shows how the framework calls attention to market responses and leads to a great number of issues for empirical investigation.

Conclusion

Technical change broadly defined is at the heart of much regional change showing up as production function shifts within a region or changes in prices received or paid as a result of events in other regions. Historically the shift from water power to fossil and other fuel sources amounted to a pervasive lowering of costs of power sources in regions with little water power. As might be expected, transportation costs play a major role in regional change by altering prices paid and received. The coming of the railroad and then the automobile, followed by superhighways, have changed the regional distribution of activity directly and indirectly. Lower transportation costs have made it less costly to respond to climate and other regional amenities, helping explain differences between 19th- and 20th-century regional growth including shifts to the South and West in the United States.

Extension of the framework to a long-term dynamic version of regional theory is required to understand lagging regions and the failure to adjust instantaneously. These topics involve capital investment and intergenerational considerations in labour mobility as younger people move out of disadvantaged regions more readily then older people, with income disparities possibly exacerbated by induced effects on local government functions, cultural milieu and leadership including entrepreneurship.

Taxes, transfers between different levels of government and transportation measures are examples of policies that can be examined applying the framework. A stated purpose of some policies is to expand regional output, though even here the interest is in the more ultimate beneficiaries such as labour or capital in the region. Policies lead to regional expansion by acting on the profitability of firms or amenities to which labour supplies respond. To the extent policies are financed from within the region there are further taxation and price effects reducing profitability of firms already in the region and raising wages that must be paid. Inter-regional effects determine national consequences. A purpose of a comprehensive regional framework is to call attention to the parameters in the producer, household demand and labour supply behaviour relations determining policy effects.

See Also

- ▶ Location of Economic Activity
- ▶ Spatial Economics

Bibliography

Beckmann, M.J. 1958. City hierarchies and the distribution of city size. *Economic Development and Cultural Change* 6: 243–248.

Borts, G.H., and J.L. Stein. 1964. *Economic growth in a free market*. New York: Columbia University Press.

Engle, R., and A. Hutchins. 1978. *Some evidence on the sources of Metropolitan growth*. San Deigo: University of California. Discussion Paper 78–16.

Henderson, J.V. 1974. The sizes and types of cites. *American Economic Review* 64(4): 640–656.

Henderson, J.V., and Y.M. Ioannides. 1981. Aspects of growth in a system of cities. *Journal of Urban Economics* 10(1): 117–139.

Hoover, E. 1937. *Location theory and the shoe and leather industry*. Cambridge, MA: Harvard University Press.

Hoover, E. 1948. *The location of economic activity*. New York: McGraw-Hill.

Howland, M. 1979. The business cycle and long run regional growth. In *Interregional movements and regional growth*, ed. W. Wheaton. Washington, DC: Urban Institute.

Isard, W. 1956. *Location and space-economy*. Cambridge, MA: MIT Press.

Isard, W. 1957. The value of the regional approach in economic analysis. In *Regional income: Studies in wealth and income*, ed. W. Isard. Princeton: Princeton University Press.

Isard, W. 1975. *Introduction to regional science*. Englewood Cliffs: Prentice-Hall.

Krumm, R.J., and G. Tolley. 1983. On the regional labor supply relation. In *The Urban economy and housing*, ed. R. Grieson. Lexington: Lexington Books.

Lösch, A. 1940. Die räumliche Ordnung der Wirtschaft. Jena: Gustav Fischer. Trans. W.H. Woglom as The Economics of Location. New Haven: Yale University Press, 1954.

Mills, E., and M. Lav. 1964. A model of market areas with free entry. *Journal of Political Economy* 72: 278–288.

Moses, L. 1958. Location and the theory of production. *Quarterly Journal of Economics* 72: 259–272.

Muth, R.F. 1971. Migration: Chicken or egg? *Southern Economic Journal* 37(3): 295–306.

North, D.C. 1955. Location theory and regional economic growth. *Journal of Political Economy* 63: 243–258.

Schmenner, R. 1982. *Making business location decisions*. Englewood Cliffs: Prentice-Hall.

Tolley, G. 1974. The welfare economics of city bigness. *Journal of Urban Economics* 1(3): 324–345.

Tolley, G., P. Graves, and J. Gardner. 1979. *Urban growth policy in a market economy*. New York: Academic Press.

Upton, C. 1981. An equilibrium model of city size. *Journal of Urban Economics* 10(1): 15–36.

von Thünen, J.H. 1826. *The Isolated State*. Trans. C.M. Wartenburg. London: Pergamon Press, 1966.

Weber, A. 1929. *Theory of location of industries*. Trans. C. Friedrich. Chicago: University of Chicago Press.

Regression and Correlation Analysis

D. V. Lindley

Correlation is a tool for understanding the relationship between two quantities. Regression considers how one quantity is influenced by another. In correlation analysis the two quantities are considered symmetrically: in regression analysis one is supposed dependent on the other, in an unsymmetric way. Extensions to sets of quantities are important.

Suppose that for each value of a quantity x, another quantity y has a probability distribution $p(y \mid x)$, the probability of y, given x. The mean value of this distribution, alternatively called the expectation of y, given x, and written $E(y \mid x)$, is a function of x and is called the regression of y on x. The quantity x is often called the independent variable, though a better term is regressor variable: y is the dependent variable. The regression tells us something about how y depends on x. The simplest case is linear regression, where $E(y \mid x) = \alpha + \beta x$ parameters α and β: the latter is called the regression coefficient (of y on x). Other features of the conditional distribution $p(y \mid x)$ are usually considered in addition to the mean. The variance (or standard deviation) measures the spread of the y – values, for fixed x. A common case is where this is constant over x: the regression is then said to be homoskedastic. A further common assumption is that $p(y \mid x)$ is normal, or Gaussian. Then y is normally distributed about $\alpha + \beta x$ with constant variance σ^2.

The regression concept of y on x does not involve a probability distribution for the regressor x. If it does have one, $p(x)$, then x and y have a joint distribution given by $p(x, y) = p(y \mid x)p(x)$. This joint distribution yields variances, σ_{xx} and σ_{yy}, for x and y, and a covariance σ_{xy}. The correlation between x and y is then defined as $\rho_{xy} = \sigma_{xy}/(\sigma_{xx}\sigma_{yy})^{1/2}$. It is the ratio of the covariance to the product of the standard deviations and is clearly unaffected by a change of scale in either x or y (and since the variances and covariance are unaffected, by a change in origin). It is easy to show that $-1 \leq \rho_{xy} \leq 1$, and if x and y are independent, ρ_{xy} is zero. When $\rho_{xy} = 0$, x and y are said to be uncorrelated. The correlation measures the association between x and y. If x and y have a joint distribution, then not only is there a regression of y on x, considered above, but also of x on y.

The linear, homoskedastic case is easily the most common one used in practice and has several important properties. We may write $y = \alpha + \beta_x + \epsilon$, where ε has zero mean and variance σ^2. If x has a distribution, then the factorization $p(x, y) = p(y \mid x)\, p(x)$ shows ε is independent of x and therefore ϵ and x are uncorrelated. Averaging we have $\mu_y = \alpha + \beta\mu_x$, relating the means, μ_x and μ_y, of x and y. A change of origin enables both of these to be put equal to zero, when $\alpha = 0$ and $E(y \mid x) = \beta x$, or $y = \beta_x + \epsilon$. Multiplying this last result by x and taking expectations, $\sigma_{xy} = \beta\sigma_{xx}$ as ε and x are uncorrelated. Consequently the regression coefficient of y on x equals σ_{xy}/σ_{xx}. Similarly the regression coefficient of x on y (if that regression is also linear homoskedastic) is σ_{xy}/σ_{yy} and the square of the correlation coefficient equals the product of the regression coefficients.

Returning to the relation $y = \beta x + \varepsilon$ and considering the variances of both sides, we obtain $\sigma_{yy} = \beta^2\sigma_{xx} + \sigma^2$ (again using the lack of correlation between x and ϵ). Hence $\sigma^2 = \sigma_{yy} - \sigma_{xy}^2/\sigma_{xx}$, on using $\beta = \sigma_{xy}/\sigma_{xx}$, and we have the important relationship that $\sigma^2 = \sigma_{yy}\left(1 - \rho_{xy}^2\right)$, showing that the variance σ^2, of y about the regression, is a proportion $\left(1 - \rho_{xy}^2\right)$ of the total variance of y, σ_{yy}. In the form $\sigma_{yy} = \beta^2\sigma_{xx} + \sigma^2$, we have the result that the total variance of y is made up of two additive components, that due to x, $\beta^2\sigma_{xx}$, and that about the regression line. The former is called the component of variance ascribable to x: the latter is the residual variance and, as we have just seen, is a proportion $\left(1 - \rho_{xy}^2\right)$ of the total. That ascribable to x is a proportion ρ_{xy}^2. This decomposition of variance is at the heart of analysis of variance techniques.

The ideas of regression and correlation are due to Galton and Pearson. The classic example has x the height of a father and y that of his son. Both regressions are linear, homoskedastic and normal, having positive regression coefficients which are less than one. Galton noticed that tall (short) fathers have sons who are, on average, shorter (taller) than themselves. This follows since, centering the values at the mean, or average height, $E(y \mid x) = \beta x < x$ if $x > 0$ corresponding to tall fathers, $\beta x > x$ if $x < 0$ for short ones. This is the phenomenon of regression (of heights) towards the mean and is necessary if the variability in heights is not to increase from one generation to the next. An illustration from economics might have x as the price of an item and y the number sold. There β will be negative reflecting the average decrease in numbers sold as the price increases. Here x might not have a probability distribution but be at the control of the seller.

The modern tendency is to make increasing use of regression and less of correlation. Part of the explanation for this is the importance of dependency relations, instead of associations, between quantities. Another reason is that in so many examples (as item price) the regressor variable is not random, so that σ_{xx} and σ_{xy} are meaningless and correlation ideas are unavailable. A third consideration is that correlation can be misleading. As an illustration of this let x be a quantity, symmetrically and randomly distributed about zero. Let $y = x^2$. Then $\sigma_{xy} = E(xy) = E(x^3) = 0$ by the symmetry about zero. Hence the correlation is zero whilst y and x are highly associated, one being the square of the other. Correlation ideas work well when all variables are normally distributed but less well otherwise. (If $y = x^2$, y cannot be normal.)

The ideas and definitions extend to the case where there are several regressor variables $x_1, x_2, \ldots, x_m$. Write $\mathbf{x} = (x_1, x_2, \ldots, x_m)$. Then $E(y \mid \mathbf{x})$ is the (multiple) regression of y on $\mathbf{x}$. In the linear case with means at zero, $E(y \mid x) = \Sigma\beta_i x_i$ and β_i is the partial regression coefficient of y on x_i. The notation and nomenclature here are too brief and can be misleading, for β_i only measures the dependence of y on x_i in the presence of the other quantities in $\mathbf{x}$. Were, say x_m, to be omitted β_i, $i < m$, would typically change: indeed, the regression might not be linear. The cumbersome notation exemplified by $\beta_{2.134}$ ($i = 2$, $m = 4$) is sometimes used. In words, the coefficient of y on x_2, allowing for x_1, x_3 and x_4. The variance about the regression remains and the homoskedastic case, where this is constant, is the one usually considered.

In the linear case $E(y \mid \mathbf{x}) = \boldsymbol{\Sigma}\beta_i P_i(x)$. the x's can be functionally related. A common case is where $x_i = x^i$, the powers of a single quantity x. This is referred to as polynomial regression. It is usually more convenient to work with polynomials $P_i(x)$ of degree i in x which are orthogonal with respect to some measure. Then $E(y \mid \mathbf{x}) = \boldsymbol{\Sigma}\beta_i P_i(x)$. Another possibility is where the x_i are periodic, say cos it. Notice that the linearity is in the terms $P_i(x)$ – or the coefficients β_i – not in x.

If the regressor variables have a joint distribution then the covariances σ_{yi}, between y and x_i, and σ_{ij} between x_i and x_j are available. With more than one regressor variable additional concepts can be introduced. For example, if all the x's are held fixed except for x_i there is a conditional joint distribution of y and x_i given all the x's except x_i. This has a correlation, defined as above as the ratio of the conditional covariance to the product of the conditional standard deviations, and is called the partial correlation between y and x_i. The notation is exemplified by $\rho_{y2.134}$. This will, in general, depend on the fixed values of the regressor variables but is normally only used when it is constant. This happens if the joint distribution of y and $\mathbf{x}$ is multivariate normal.

In the case of a single regressor variable we saw that $1 - \rho_{xy}^2 = \sigma^2/\sigma_{yy}$, where σ^2 is the residual variance of y, conditional on x. In the multiple case, continue to define σ^2 in this way conditional on all the quantities in $\mathbf{x}$. Then define R^2 by $(1 - R^2) = \sigma^2/\sigma_{yy}$, in analogy with the single variable case. The positive square root R is called the multiple correlation coefficient (of y on $\mathbf{x}$). As before, we may write $\sigma_{yy} = \sigma^2 + R^2\sigma_{yy}$ expressing the total variance of y additively in terms of the residual variance σ^2 and that due to the regression on $\mathbf{x}$. It is more common nowadays to work in terms of the variance components than R^2.

The mathematical theory of regression and correlation is now well understood. Centering at the means, all the concepts depend on the matrix of variances and covariances of y, the dependent variable, and $\mathbf{x}$, the set of regressor variables: σ_{yi} and σ_{ij}. The calculations are merely ways of rearranging these elements in convenient forms: correlations and components of variance in regression are just two possibilities. The real difficulty, and the real interest in regression lies in the interpretation of the results.

As an illustration consider the simple case of linear, homoskedastic regression of y on a single regressor variable x, written $y = \beta x + \varepsilon$, with β as the regression coefficient and ε as the residual variation, with zero mean and variance σ^2. All this says is that for any fixed x, y has mean βx and variance σ^2: and it is only this aspect of the dependence of y on x that is described. Suppose a large amount of data consisting of pairs (x_i, y_i) is collected and the fit $y = 2x + \varepsilon$ with $\sigma^2 = 2$ is established. (We discuss how this might be done below.) This shows a fairly close association between y and x. In order therefore to increase y it might be thought reasonable to set x to a high value. Suppose this is done, will this cause y necessarily to increase? Surprisingly, not so. Suppose there is another quantity z and the real relationships are that $y = -x + z + \varepsilon_1$, and $x = \frac{1}{3}z + \varepsilon_2$ so that z is the basic quantity determining the situation. This clearly yields $y = 2x + \varepsilon$, with $\varepsilon = \varepsilon_1 - 3\varepsilon_2$, the observed relation. If now x is controlled at a large value without affecting z which is, under natural conditions, the main determinant of x, the effect will be to decrease y through $y = -x + z + \varepsilon_1$. Consequently a strong positive relationship between y and x need not imply an increase in y when x is increased. There can be an enormous difference between the association of y with x, when x is uncontrolled and allowed to vary freely, and the association when x is controlled. And the reason is the presence of another quantity z whose influence on x in the free system is disturbed by the control.

Whenever the regression of y on a set of quantities $\mathbf{x}$ is discussed, one has to beware of the possible presence of other, unobserved quantities $\mathbf{z}$ that could affect the relationship. A laboratory scientist, or even a social scientist doing a planned survey, can often guard against such hidden quantities by careful design or by appropriate randomization; but an economist, or anyone who has to rely on data from unplanned studies, has always to be on his guard against their effects. Another way of describing the difficulty is to distinguish carefully between association and causation. All regression and correlation analyses can do is study association: the underlying causal mechanism is not necessarily revealed. It is remarkable how little attention has been paid by statisticians to the meaning of causation, and to how it can be revealed by statistical analysis. Economists have had to rely on statistical analyses of randomly obtained data and some of the causal inferences they have drawn are totally unjustified by that data and the analyses.

We now consider the nature of these statistical analyses, confining ourselves predominantly to the case of homoskedastic, linear regression $y = \Sigma\beta_i x_i + \varepsilon$, ε having mean zero and variance σ^2. There the means have been supposed zero. There is usually no difficulty over this as the mean of each variable can ordinarily be estimated by the sample means, $y.$ and x_i. The quantities being discussed here are, in terms of the original data, the deviations, $y - y$ and $x_i - x._i$, from the sample means. The standard method of estimating the β's and σ^2 is least squares. This has been in use for two centuries and is still adopted by almost all data analysts. If that data is (y_j, x_{ji}: $i = 1, 2, \ldots m$; $j = 1, 2, \ldots n$) consisting of n independent observations of y and the m regressor variables, then the least-squares estimates of β_i are provided by minimizing the sum of squares of residuals $y - \Sigma\beta_i x_i$ for each of the n observations: that is $\Sigma_j(yj - \Sigma_i\beta_i x_{ji})^2$. Matrix notation is most convenient. Write $\boldsymbol{y} = (y_1, y_2, \ldots y_\mathrm{n})^\mathrm{T}$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_\mathrm{m})^\mathrm{T}$ and X as the matrix with elements x_{ji}, observation j on variable x_i. Then $\mathbf{y} = \boldsymbol{X\beta}$ + residual and the sum of squares to be minimized over β is $(\boldsymbol{y} - \boldsymbol{X}\beta)^\mathrm{T}$ $(\boldsymbol{y} - \boldsymbol{X}\beta)$ with minimum given by $\widehat{\beta} = \left(X^T X\right)^{-1} X^T y$. The variance σ^2 is estimated by the sum of squares at $\widehat{\beta}$ divided by $(n - m)$. The $\widehat{\beta}_i$ are called the least-squares estimates of β_i.

The method is deservedly popular because it is relatively easy to use and interpret, and many convenient computer programs are available. Its long and successful history testifies to its merits. Unfortunately it has been discovered that there can be very real difficulties when m, the number of variables is large. With the availability of fast computers capable of handling a lot of data, it is not uncommon to have 40 or more variables. The difficulties then become noticeable. Before the arrival of such computing power, least squares was only used with few variables and the difficulties are scarcely noticeable. It is easy to appreciate what could go wrong: it is not so easy to correct it. Consider the case where the sum of squares is $\Sigma_j(y_j - \beta_j)^2$. This apparently special and degenerate case is, in fact, a canonical form for least squares and any multiple regression situation can be transformed to it by linear transformations. (In so doing, the meanings of the y's and β's will change.) The minimization is trivial with estimate $\widehat{\beta}_j = y_j$, and the minimum value is zero. But we know that y_j differs from its expectation, here β_j but in general $\Sigma_i\beta_i x_{ji}$, by an amount which has variance σ^2, so the average of $\Sigma_j\left(y_j - \widehat{\beta}_j\right)^2$ ought to be about σ^2, and indeed this is the usual estimate of σ^2 as mentioned above. But here this estimate is zero, which is absurd. This first, rigorous demonstration that least squares is unsatisfactory was given by Charles Stein. He showed that whenever the number of variables exceeds two, there is an estimate which is, for every value of the regression coefficients, better than least squares. Better here means having smaller mean-square error, though the statement remains true under many other meanings. The efficiency varies with the true values of the β's. The result just quoted says that it is always less than one. It can be as low as $2/m$: with $m = 40$ this gives only 5% efficiency, a rather serious loss.

It is surprising how little attention Stein's result has received outside of a small group, largely of theoreticians, yet its practical value could be enormous. Stein, and others, have produced estimates which improve on least squares but none has had much acceptance. Fairly early in the use of computers for regression analysis, it was appreciated that difficulties could arise when the matrix X^TX, which has to be inverted to obtain the least-squares estimates, is illconditioned, with determinant near zero. This is the matrix of sample variances and covariances of the regressor variables, a typical element being $\boldsymbol{\Sigma}_r x_{ri} x_{rj}$ where the x's are deviations from their means, x_i. It will be ill-conditioned if, in the data, there is a near linear relationship between the regressor variables. One suggestion was to put the matrix into correlation form, dividing each row and each column by the sample standard deviation of the variable corresponding to that row or column, so making all diagonal elements one and each off-diagonal element equal to a sample correlation coefficient between x_i and x_j, and then subtracting a constant λ from each unit diagonal element. This leads to ridge regression estimates and ways of choosing λ have been proposed. It often works well but can fail.

These ideas all lie within a frequentist school of inference. In principle, a solution is available with the Bayesian paradigm for inference. Here, in addition to the distribution of y, given $\mathbf{x}$, is included a probability distribution for the regression parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)$. Inference is then made by calculating the revised probability distribution of $\boldsymbol{\beta}$ given the data. This procedure always avoids Stein's criticism provided the original distribution of $\boldsymbol{\beta}$ has total integral unity. (Least squares results from this procedure only if all the values of $\boldsymbol{\beta}$ are equally probable, a form which is not finitely integrable.) The practical difficulty is the choice of a distribution for $\boldsymbol{\beta}$. The ridge method can be produced for certain types of exchangeable distributions for $\boldsymbol{\beta}$. In the case of polynomial regression, a reasonable possibility is to suppose that the coefficients of the higher degree polynomials are likely to be smaller than those of lower degree. When the regressor variables refer to different quantities, a possibility is to suppose that few of them have an appreciable coefficient, and therefore influence y, but it is not known which are the determining ones.

This idea that only a few regressors matter has led to a lot of work on the choice of which to include in the regression. There are two broad ways to proceed. One can fit all the quantities available and then discard them one by one as long as the discarding has little effect. Or one can proceed in the reverse direction, introducing them one at a time only if they have an appreciable effect. In both of these methods it has to be decided how the effect is to be measured. The usual criterion is the change in the variance of y ascribable to $\mathbf{x}$; the quantity denoted above by $\boldsymbol{R}^2\boldsymbol{\sigma}_{yy}$. Alternatively expressed, this is the change in the multiple correlation coefficient. For example, in the method where the variables are discarded, R^2 will decrease when a variable is omitted from the regression. Only if this decrease is small will the omision be granted. There are two difficulties here. First, it is possible for two quantities, separately to have little effect, but jointly to be of considerable importance, so that tests of them one at a time may be misleading. (The possibility of computing all 2^m regressions is too extravagant.) Second, it is not clear what is meant by saying the change in R^2 is "small": how small? One possibility is to use an ordinary significance test, here a t-test. If significant the regressor causing the change can be included: if not, it is omitted. This is for some suitably chosen significance level. This has been thought to be unsatisfactory by some and other criteria have been proposed. It is here that the Bayesian and frequentist views part company. The usual Bayesian criterion for 'small' depends on the assumed distribution for the regression coefficients, but, in general, it seems to need more evidence to introduce a regressor when using the Bayesian approach than when employing a significance test. The former has been accused of favouring the hypothesis that the variable is not worth including. The Bayesian reply is that some 'significant' effects are spurious. Multiple regression techniques are so widely used today that one wonders how many effects of x_i on y reported in the literature are meaningful.

Regression concerns a relation, to take the linear, one variable form, $y = \beta x + \varepsilon$ between y and x. This treats y and x asymmetrically and does not lead to $y = \beta^{-1}y + \varepsilon'$ with ε' unrelated to y. There is, however, a symmetric form that is sometimes useful. Suppose two quantities, ξ and η, are exactly linearly related, $\eta = \beta\xi$, or equally $\xi = \beta^{-1}\eta$. Suppose that each is measured with error giving $y = \eta + \varepsilon$, $x = \xi + \varepsilon'$. Then the pair (x, y) may have linear regressions but the real interest lies in β, the coefficient of the exact relationship. This is often referred to as the case where both variables, independent and regressor, are subject to error. Ordinary least-squares techniques, even with a single regressor variable, require modification.

Linear multiple regression is part of the general theory of linear models in which, to use the notation above, $\boldsymbol{E}(y \mid \boldsymbol{X}) = \boldsymbol{X\beta}$, the linearity being in the parameter $\boldsymbol{\beta}$. Least squares and its Steintype modifications are the standard techniques for analysis, together with the analysis of variance.

See Also

- Bayesian Inference
- Estimation
- Heteroskedasticity
- Least Squares
- Maximum Likelihood
- Non-linear Methods in Econometrics
- Outliers
- Residuals
- Statistical Inference

Bibliography

Efron, B., and C. Morris. 1975. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* 70: 311–319.
Hoerl, A.E., and R.W. Kennard. 1970. Ridge regression: Biased estimation of non-orthogonal problems. *Technometrics* 12: 55–67.
Seber, G.A.F. 1977. *Linear regression analysis*. New York: Wiley.
Vinod, H.D., and A. Ullah. 1981. *Recent advances in regression methods*. New York: Dekker.
Zellner, A. 1971. *An Introduction to Bayesian inference in econometrics*. New York: Wiley.

Regression-Discontinuity Analysis

Wilbert van der Klaauw

Abstract
In recent years regression discontinuity analysis has grown into a popular approach for evaluating causal relationships in empirical economics. The method takes advantage of a discontinuity in the probability of treatment as a function of a continuous variable to identify a meaningful average treatment effect. This article summarizes the regression discontinuity approach to identifying and estimating causal effects and describes several validity tests.

Keywords
Assignment mechanisms; Control function approach; Identification; Instrumental variables; Kernel estimators; Propensity score; Regression discontinuity analysis; Selection bias; Semiparametric estimation; Treatment effect

JEL Classification
C14

The regression discontinuity (RD) data design is a quasi-experimental evaluation design first introduced by Thistlethwaite and Campbell (1960) as an alternative approach to evaluating social programmes. The design is characterized by a treatment assignment or selection rule which involves the use of a known *cut-off* point with respect to a continuous variable, generating a discontinuity in the probability of treatment receipt at that point. Under certain comparability conditions, a comparison of average outcomes for observations just left and right of the cut-off can be used to estimate a meaningful causal impact. While interest in the design had previously been mainly limited to evaluation research methodologists (Cook and Campbell 1979; Trochim 1984), the design is currently experiencing a renaissance among econometricians and empirical economists (Hahn et al. 1999, 2001; Angrist and Krueger 1999; Porter 2003). Among the main econometric contributions have been the formal derivation of identification conditions for causal inference and the introduction of semiparametric estimation procedures for the design. At the same time, a large and rapidly growing number of empirical applications are providing new insights into the applicability of the design, which have led to the development of several sensitivity and validity tests.

The popularity of the RD design in applied economic research can be linked to several of its features. First, the assignment rules in many existing programmes and procedures for allocating social resources, frequently lend themselves to RD evaluations. In many cases, programme resources are allocated based on some type of formula that has a cut-off structure. One area of economic research where the design has proven especially fruitful in recent years has been the evaluation of educational interventions. Education programmes are frequently assigned to schools or students who score below a cut-off on some scale (student performance, poverty), and school and programme funding decisions are often based on allocation formulas containing discontinuities. Similarly, the design has proven useful in evaluating the socio-economic impacts of a diverse set of government programmes and laws, many of which use eligibility cutoffs or funding formulas involving thresholds in allocating scarce resources to those potential recipients who need or deserve them most (see van der Klaauw 2007a). A second attractive feature of the design is that it is intuitive and its results can be easily communicated, often with a visual portrayal of sharp changes in both treatment assignment and average outcomes around the cut-off value of the assignment variable (Bloom et al. 2005). Third, a researcher can choose from among several different methods to estimate effects that have credible causal interpretations (Hahn et al. 2001).

Consider the general problem of evaluating the impact of a binary treatment on an outcome variable, using a random sample of individuals where for each individual i we observe an outcome measure y_i and a binary treatment indicator t_i, equal to one if treatment was received and zero otherwise. The evaluation problem that arises in determining the effect of t on y, is due to the fact that each individual either receives or does not receive treatment and is never observed in both states. Let $y_i(1)$ be the outcome given treatment, and $y_i(0)$ the outcome in absence of treatment. Then the actual outcome we observe equals $y_i = t_i y_i(1) + (1-t_i)y_i(0)$. A common regression model representation for the observed outcome can then be written as

$$y_i = P + \alpha_i t_i + u_i \qquad (1)$$

where $\alpha_i = y_i(1) - y_i(0)$ and $y_i(0) = E[y_i(0)] + u_i = \beta + u_i$. Non-random assignment or selection into treatment implies that a comparison of average outcomes of treatment recipients and non-recipients ($E[y_i(1)|t_i = 1]$ and $E[y_i(0)|t_i = 0]$) would generally not provide us with a valid treatment effect estimate.

Hahn et al. (2001) analysed the conditions under which a discontinuity in the treatment assignment or selection rule can be exploited to solve the selection bias problem and to identify a meaningful causal effect. Following Trochim (1984) they distinguish between two different forms of the design, depending on whether the treatment assignment is related to the assignment variable by a deterministic function (*sharp design*) or a stochastic one (*fuzzy design*). In the case of a sharp RD design, individuals are assigned to or selected for treatment solely on the basis of a cut-off score on an observed continuous variable x. This variable, alternatively called the assignment, selection, running or ratings variable, could represent a single characteristic or a composite variable constructed using several characteristics. Those who fall below some distinct cutoff point $\bar{x}$ are placed in the control group ($t_i = 0$), while those on or above that point are placed in the treatment group ($t_i = 1$) (or vice versa). Thus, assignment occurs through a known and measured deterministic decision rule: $t_i = t(x_i) = 1\{x_i \geq \bar{x}\}$ where $1\{.\}$ is the indicator function. As the assignment variable itself may be correlated with the outcome variable, the assignment mechanism is clearly not random.

R

However, if we have reason to believe that persons close to the threshold with very similar x values are comparable, then we may view the design as almost experimental near $\bar{x}$, suggesting that we could evaluate the causal impact of treatment by comparing the average outcome for those with ratings just above to those with ratings just below the cutoff. More formally, consider the following *local continuity* (*LC*) *assumption*:

$E[u_i|x]$ and $E[\alpha_i|x]$ are continuous in x at $\bar{x}$, or equivalently, $E[y(1)|x]$ and $E[y(0)|x]$ are continuous at $\bar{x}$,

then on the assumption that the density of x is positive in a neighbourhood containing $\bar{x}$,

$$\begin{aligned}\lim_{x\downarrow\bar{x}} E[y_i|x] - \lim_{x\uparrow\bar{x}} E[y_i|x] &= \lim_{x\downarrow\bar{x}} E[\alpha_i t_i|x] \\ &+ \lim_{x\uparrow\bar{x}} E[\alpha_i t_i|x] + \lim_{x\downarrow\bar{x}} E[u_i|x] \\ &- \lim_{x\uparrow\bar{x}} E[u_i|x] = E[\alpha_i|x=\bar{x}].\end{aligned} \tag{2}$$

The RD approach therefore identifies the average treatment effect for individuals close to the discontinuity point. Note that the continuity assumption formalizes the idea that individuals just above and below the cut-off need to be 'comparable', requiring them to have similar average potential outcomes when receiving and when not receiving treatment. While in the absence of additional assumptions (such as a common effect assumption) one could learn about treatment effects only for a sub-population of persons near the discontinuity point, as pointed out by HTV this local effect is highly relevant to policymakers who are contemplating less restrictive eligibility rules and marginal expansions of programmes via a change in the cut-off.

The continuity assumption required for identification is not innocuous. Even if treatment receipt is determined solely on the basis of a cut-off score on the assignment variable, this is not a sufficient condition for the identification of a meaningful causal effect. The continuity assumption rules out coincidental functional discontinuities in the x–y relationship such as those caused by other programmes employing assignment mechanisms based on the exact same assignment variable and cut-off. In addition, the continuity restriction generally rules out certain types of behaviour both on the part of potential treatment recipients who exercise control over their value of x and programme administrators in choosing the assignment variable and cut-off point. Lee (2007) analyses the conditions under which an ability to manipulate the assignment variable may invalidate the RD identification assumptions. He shows in the context of a sharp RD design that as long as individuals do not have *perfect* control over the position of the assignment variable relative to the cut-off score, the continuity assumption will be satisfied.

While in the sharp RD design treatment assignment is known to depend on the selection variable x in a deterministic way, in the case of a fuzzy design (Campbell 1969), treatment assignment depends on x in a stochastic manner but in such a way that the propensity score function $\Pr(t=1|x)$ is again known to have a discontinuity at $\bar{x}$. Instead of a 0–1 step function, the selection probability as a function of x would now contain a jump smaller than 1 at $\bar{x}$. The fuzzy design can occur in case of misassignment relative to the cut-off value in a sharp design, with values of x near the cut-off appearing in both treatment and control groups. This situation is analogous to having no-shows (treatment group members who do not receive treatment) and/or crossovers (control group member who do receive the treatment) in a randomized experiment. This could occur if in addition to the position of the individual's score relative to the cut-off value, assignment is based on additional variables observed by the administrator, but unobserved by the evaluator.

A comparison of average outcomes of recipients and non-recipients, even if near the cut-off, would not generally lead to correct inferences regarding an average treatment effect. However, as shown by HTV, one can again exploit the discontinuity in the selection rule to identify a causal impact of interest by noting that under the LC assumption and with a locally constant treatment effect ($\alpha_i = \alpha$ in a neighbourhood around $\bar{x}$), the treatment effect α is identified by

$$\lim_{x\downarrow\bar{x}}E[y_i|x] - \frac{\lim_{x\uparrow\bar{x}}E[y_i|x]}{\lim_{x\downarrow\bar{x}}E[t_i|x] - \lim_{x\uparrow\bar{x}}E[t_i|x],} \quad (3)$$

where the denominator is always non-zero because of the known discontinuity of $E[t\mid x]$ at $\bar{x}$.

In the case of varying treatment effects, HTV show that under the local continuity assumption, and a local conditional independence assumption requiring to be independent of α_i conditional on x near $\bar{x}$, the ratio above identifies $E[\alpha_i|x=\bar{x}]$, the average treatment effect for cases with values of x close to $\bar{x}$. The conditional independence assumption is a strong assumption which may be violated if individuals self-select into or are selected for treatment on the basis of expected gains from treatment. HTV show that, under a weaker local monotonicity assumption similar to that assumed by Imbens and Angrist (1994), the ratio (3) will instead identify a local average treatment effect (LATE) at the cut-off point, which represents the average treatment effect of the 'compliers', that is, the subgroup of individuals whose treatment status would switch from non-recipient to recipient if their score x crossed the cut-off. More recently Battistin and Rettore (2003) considered the special case where an eligibility rule divides the population into eligibles and non-eligibles according to a sharp RD design, and with eligible individuals self-selecting into treatment. In this case the LC assumption alone is sufficient for the ratio to identify $E[\alpha_i|t_i=1,x=\bar{x}]$, the average treatment effect on the treated, for those near the cutoff.

As indicated by these identification results, estimation of treatment effects in an RD design involves estimating boundary points of conditional expectation functions. The most common empirical strategy in the literature has been to adopt parametric specifications for the conditional expectations functions. Consider the following alternative representation of outcome Eq. (1) in case of a sharp RD design:

$$y_i = m(x_i) + \delta t_i + e_i, \quad (4)$$

where
$e, -= y_i - E[y_i|t_i,\ x_i]$, $t_i = 1\{x_i,\ \geq \bar{x}\}$,
$m(x) = E[u_i|x] + (E[\alpha_i|x] - E[\alpha_i|\bar{x}])1\{x\geq\bar{x}\}$

Then under the local continuity assumption $m(x)$ will be a continuous function of x at $\bar{x}$, and $\delta = E[\alpha_i|\bar{x}]$ (the average treatment effect at $\bar{x}$) will measure the discontinuity in the average outcome at the cut-off. This suggests that if the correct specification of $m(x)$ were known, and was included in the regression, we could consistently estimate the treatment effect for the sharp RD design. This idea of including a specification of $m(x)$ in the regression of y on t in order to correct for selection bias caused by selection on observables, is in the econometrics literature known as the control function approach (Heckman and Robb 1985). A popular choice among empirical researchers has been to use global polynomials or to use splines (piecewise polynomials) which, even though globally continuous, have a knot at the cut-off (Trochim 1984; van der Klaauw 2002; McCrary 2007).

In the case of a fuzzy RD design, when assuming local independence of t_t and α_i conditional on x, then in a neighbourhood of $\bar{x}$,

$$y_i = m(x_i) + \delta E[t_i|x_i] + w_i, \quad (5)$$

where $w_i = y_i - E[y_i|x_i]$ and $m(x) = E[u_i|x] + (E[\alpha_i|x] - E[\alpha_i|\bar{x}])E[t|x]$. With the local continuity assumption again implying that $m(x)$ will be continuous at the cutoff, and with $E[t_i|x_i]$ being discontinuous at $\bar{x}$, δ in this regression will measure the ratio in (3), which in this case equals the average local treatment effect $E[\alpha_i|x]$. Similarly, δ can be interpreted as a local average treatment effect if we replaced the local independence assumption with the local monotonicity condition of Imbens and Angrist (1994).

This naturally leads to the two-stage procedure adopted by van der Klaauw (2002), where in the first stage we estimate the propensity score function specified as

$$t_i = E[t_i|x_i] + v_i = f(x_i) + \gamma 1\{x_i \geq \bar{x}\} + v_i$$

where $f(\cdot)$ is continuous in x at $\bar{x}$ and γ measures the discontinuity in the propensity score function at $\bar{x}$. In the second stage the control function-augmented outcome equation is then estimated

R

with t_i replaced by the first-stage estimate of $E[t_i|x_i] = \Pr[t_i = 1|x_i]$ as in Maddala and Lee (1976). With correctly specified $f(x)$ and $m(x)$ functions, this two-stage procedure yields a consistent estimate of the treatment effect. The approach is similar in spirit to those proposed earlier in the RD evaluation literature by Spiegelman (1979) and Trochim and Spiegelman (1980). Note that in case of a parametric approach, if we assume the same functional form for $m(x)$ and $f(x)$, then the two-stage estimation procedure described here will be equivalent to two-stage least squares (in case of linear-in-parameter specifications) with $1\{x_i \geq \bar{x}\}$ and the terms in $m(x)$ serving as instruments. Because of the popularity of this particular parametrization, the RD approach is often interpreted as being equivalent to an instrumental variable approach, as it implicitly imposes an exclusion restriction by excluding $1\{x_i \geq \bar{x}\}$ as a variable in the outcome equation.

Valid parametric inference for the estimation of the treatment effect requires a correct specification of the control function $m(x)$ and of $f(x)$ in the treatment equation. To mitigate the potential for misspecification bias, several semiparametric estimation procedures have been proposed for estimating $m(x)$ and $f(x)$, or equivalently for estimating the limits $\lim_{x\downarrow\bar{x}}E[z|x]$ and $\lim_{x\uparrow\bar{x}}E[z|x]$ in (3) semiparametrically. These methods rely on less-restrictive smoothness conditions away from the discontinuity, with estimates based mainly on data in a neighbourhood on either side of the cut-off point. Asymptotically this neighbourhood needs to shrink, as with usual nonparametric estimation, implying that we should expect a slower than parametric rate of convergence in estimating treatment impacts. HTV considered the use of kernel and local linear regression estimators, while Porter (2003) proposed estimating the limits using local polynomial regression and partially linear model estimation. RD estimators based on local polynomial regression and partially linear model estimation have better boundary behaviour than the kernel-based estimator and as shown by Porter, achieve the optimal rate of convergence. This result is based on a known degree of smoothness of the conditional expectation functions. Sun (2005) proposed an adaptive estimator to first estimate the degree of smoothness in the data prior to implementing either estimator.

The internal validity of the RD approach relies on the local continuity of conditional expectations of potential outcomes around the discontinuity point. While this assumption is fundamentally untestable, a number of validity tests have been developed to bolster the credibility of the RD design. First, economic behaviour may lead to sorting of individuals around the cut-off point, where those below the cut-off may differ on average from those just above the cut-off. Such precise sorting around the cut-off would generally be accompanied by a discontinuous jump in the density of the assignment variable at the cutoff. Several approaches have been used for assessing this possibility (McCrary 2007; Lee 2007; Chen and van der Klaauw 2007; Lemieux and Milligan 2004). Second, one can test for evidence that individuals on either side of the cut-off are observationally similar by directly comparing average characteristics (McEwan and Urquiola 2005) or by repeating the RD analysis treating the characteristics as outcome variables (van der Klaauw 2007b). Alternatively, one can test for an imbalance of relevant characteristics by assessing the sensitivity of RD estimates to the inclusion of observed characteristics as controls (van der Klaauw 2002; Lee 2007). Third, in some applications data are available from a baseline period in which the programme did not yet exist, or for a group of individuals that was not eligible for treatment. In such a case the credibility of the design can be significantly enhanced by repeating the RD analysis with such data. Finding a zero treatment effect in such a falsification test would suggest that a non-zero post-programme effect was not an artifact of the specific RD model specification, estimation approach chosen or caused by another programme using the same cut-off and assignment variable.

Finally, while this exposition has focused on the binary treatment case with a selection rule containing a single discontinuity at a known cut-off, the approach can be readily extended to one where there are multiple treatment dose levels and multiple cut-offs or 'cut-off ranges' within which the treatment dose varies continuously (van der Klaauw 2007a). Similarly, the approach can be modified to cover cases where the assignment or selection variable is discrete instead of continuous (Lee and Card 2006).

See Also

- Causality in Economics and Econometrics
- Natural Experiments and Quasi-Natural Experiments
- Propensity Score
- Selection Bias and Self-Selection
- Semiparametric Estimation
- Treatment Effect

Bibliography

Angrist, J.D., and A.B.. Krueger. 1999. Empirical strategies in labor economics. In *Handbook of labor economics*, vol. 3, ed. O. Ashenfelter and D. Card. Amsterdam: North-Holland.

Battistin, E., and E. Rettore. 2003. *Another look at the regression discontinuity design*. Working Paper No. CWP01/03, CeMMAP, Institute for Fiscal Studies.

Bloom, H.S., J. Kemple, B. Gamse and R. Jacob. 2005. *Using regression discontinuity analysis to measure the impacts of reading first*. Paper presented at the annual conference of the American Educational Research Association, Montreal, Canada, April.

Campbell, D.T. 1969. Reforms as experiments. *American Psychologist* 24: 409–429.

Chen, S., and W. van der Klaauw. 2007. The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*.

Cook, T.D., and D.T. Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton-Mifflin.

Hahn, J., P. Todd, and W. van der Klaauw. 1999. *Evaluating the effect of an antidiscrimination law using a regression-discontinuity design*, Working paper no. 7131. Cambridge, MA: NBER.

Hahn, J., P. Todd, and W. van der Klaauw. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69: 201–209.

Heckman, J.J., and R. Robb. 1985. Alternative methods for evaluating the impact of interventions. In *Longitudinal analysis of labor market data*, ed. J. Heckman and B. Singer. New York: Cambridge University Press.

Imbens, G.W., and J. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–476.

Lee, D.S. 2007. Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics.* (forthcoming).

Lee, D.S., and D. Card. 2006. *Regression discontinuity inference with specification error*, Technical working paper no. 322. Cambridge, MA: NBER.

Lemieux, T., and K. Milligan. 2004. *Incentive effects of social assistance: A regression discontinuity approach*, Working paper no. 10541. Cambridge, MA: NBER.

Maddala, G.S., and L. Lee. 1976. Recursive models with qualitative endogenous variables. *Annals of Economic and Social Measurement* 5: 525–545.

McCrary, J. 2007. Testing for manipulation of the running variable in the regression discontinuity design. *Journal of Econometrics.*

McEwan, P.J., and M. Urquiola. 2005. *Economic behavior and the regression-discontinuity design: Evidence from class size reduction*. Working paper, Columbia University.

Porter, J. 2003. *Estimation in the regression discontinuity model*. Unpublished manuscript, Harvard University.

Spiegelman, C.H. 1979. *Estimating the effect of a large scale pretest posttest social program*. Proceedings of the Social Statistics Section, American Statistical Association, pp. 370–373.

Sun, Y. 2005. *Adaptive estimation of the regression discontinuity model*. Working paper, University of California, San Diego.

Thistlethwaite, D., and D. Campbell. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51: 309–317.

Trochim, W.K. 1984. *Research design for program evaluation: The regression- discontinuity approach*. Beverly Hills: Sage.

Trochim, W., and C.H. Spiegelman. 1980. *The relative assignment variable approach to selection bias in pretest-posttest group designs*. Proceedings of the Survey Research Section, American Statistical Association, pp. 376–380.

van der Klaauw, W. 2002. Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* 43: 1249–1287.

van der Klaauw, W. 2007a. *Regression-discontinuity analysis: A survey of recent developments in economics*. Unpublished manuscript, Federal Reserve Bank of New York.

van der Klaauw, W. 2007b. Breaking the link between poverty and low student achievement: An evaluation of title I. *Journal of Econometrics*.

Regular Economies

Egbert Dierker

JEL Classifications
D5

General equilibrium theory describes those states of an economy in which the individual plans of many agents with partially conflicting interests are compatible with each other. Such a state is called an equilibrium. The concept of an equilibrium simply being based on a consistency requirement lends itself to the study of specific questions of quite different character. Indeed, equilibrium theory provides a unifying framework for the analysis of questions arising in various branches of economic theory. In our opinion it is fruitful to view equilibrium theory as a method of thinking applicable to a variety of problems of different origin.

Ideally one would like to have general principles which ensure that equilibria exist, that they are unique, and that, therefore, the equilibria resulting from different policy measures can unequivocally be compared. Moreover, one would like to know whether equilibria have some desirable properties when no single agent can exert an essential influence on the global outcome to his personal advantage. These welfare questions are particularly interesting because the concept of an equilibrium itself is not based on the well-being of the economic agents. Finally, although the concept of an equilibrium as described above is static in nature, one would like to have a dynamic theory according to which some equilibrium is approached in the course of time.

These and related questions such as the computability of equilibria have been studied in the past with different degrees of success. There are general principles which yield the existence of an equilibrium in an astonishingly large variety of cases. Furthermore, the welfare properties of equilibria are well understood. However, it is easy to construct examples of economies with an infinite number of equilibria and it appears to be very difficult to provide conditions which lead, without being artificial or and hoc, to the uniqueness of an equilibrium. As a consequence, comparative statics does not have a basis which makes it generally a well-defined problem. Also, the difficulties encountered when studying the uniqueness issue present severe obstacles for the development of a dynamic theory.

The theory of regular economies may be viewed as an effort to advance general equilibrium theory in the absence of a satisfactory uniqueness result. The seminal paper is Debreu (1970). Debreu explicitly allows for the multiplicity of equilibria. However, he requires each equilibrium to be *locally* unique. Each equilibrium is well determined and robust in the sense that it is not destroyed by a small change in the parameters.

A regular economy is an economy with a certain, finite number of equilibria, all of which respond continuously to small parameter changes. Hence each of these equilibria can be traced for some while during a parameter change. Thus there is a basis for doing comparative statics locally, that is to say as long as the equilibrium under consideration stays robust. If, at a certain point, it ceases to be robust, a drastic change is to be expected, the size and direction of which are probably hardly predictable. The focus of the theory of regular equilibria is more on the continuous behaviour of robust equilibria than on drastic changes.

It is most remarkable that Debreu (1970), by using concepts and techniques developed in the mathematical field of differential topology, has introduced a new kind of thought into economic analysis. In the meantime this way of thinking has penetrated many areas of economic theory at different levels. One of the first applications has occurred in the technically advanced area of core theory, where the continuous dependence of the set of price equilibria on the characteristics of the agents, which is guaranteed in a regular economy, plays an important role. An application on a purely conceptual level in oligopoly theory is incorporated in the notion of a demand function which an oligopolist faces in the Cournot–Nash context. The graph of this function is considered as given by the equilibria of an exchange economy with initial endowments as varying parameters.

The dependence of the equilibria on initial endowments will be discussed in detail in the next section because this case is particularly suited to illustrate basic ideas of the theory of regular economies.

Debreu's Theorem on Regular Equilibria

The purpose of this section is to describe the kind of reasoning typical for the theory of regular economies in a prototypical situation. It is desirable to deal with parameter variations taking place in some Euclidean space because the mathematical structures to be used are most familiar in this case. We shall study exchange economies which differ by the allocation of initial endowments.

There are l commodities and m consumers. Individual initial endowments are supposed to be positive in each component. If we denote the strictly positive orthant in $\mathbb{R}^l$ by P, then an initial allocation is a vector $(e_1,\ldots,e_m) \in P^m$. Since the demand function f_i of each consumer i is considered as fixed, an economy E is fully specified by $(e_1,\ldots,e_m)$. The space of all economies under consideration can thus be identified with P^m, an extremely simple subset of a Euclidean space. We want to examine how the exchange equilibria of an economy-there may be several such equilibria–depend on the particular economy $E \in P^m$.

We assume that all goods are desired so that attention may be restricted to strictly positive relative prices. Price systems are normalized; to be specific we consider price systems in

$$S = \left\{ p = (p_1 \ldots p_2) \gg 0 || |p|| = \left(\sum_{k-1}^{l} p_k^2 \right)^{1/2} = 1 \right\}.$$

If consumer i initially possesses the commodity bundle e_i, his demand at the price system p is $f_i(p, p \cdot e_i) \in \mathbb{R}_+^l$, where $p \cdot e_i = w_i > 0$ is i's wealth. Hence the aggregate excess demand of the economy E, given by the initial allocation $(e_1,\ldots,e_m) \in P^m$, at p is

$$Z_E(p) = \sum_{i=1}^{m} [f_i(p, p \cdot e_i) - e_i].$$

We assume Walras's Law which states that the value $p \cdot Z_E(p)$ of the excess demand is identically equal to zero. Furthermore, every f_i is supposed to be continuous.

The desirability of all commodities will be captured in the following condition, which is always satisfied when consumers have strictly monotone preferences.

(D) If the price of at least one good approaches zero and the wealth $w_i > 0$ of every agent stays away from zero, then

$$\sum_{i=1}^{m} \|f_i(p, w_i)\|$$

tends to infinity.

An *equilibrium price system* of E is a price system $p \in s$ at which the consumption plans $f_i\,(p, p \cdot e_i)$ of all agents are consistent, i.e. a zero of the excess demand function Z_E. It is not difficult to show the following consequence of the desirability assumption (D) by a fixed point argument:

Every economy $E \in P^m$ has at least one equilibrium if (D) *holds.*

We would like to know how the equilibrium prices vary when the initial allocation is modified. Therefore we look at the graph Γ of the correspondence ('multi-valued function') $\prod$ which assigns to every economy $E \in P^m$ its set $\{p \in S + Z_E(p) = 0\}$ of equilibrium price systems. Defining $Z : P^m \times S \to \mathbb{R}^l$ by $Z(E, p) = Z_E\,(p)$ we get

$$\text{graph}\left(\prod\right) = \Gamma = Z^{-1}(0).$$

Since Z is a continuous function, Γ is a closed set. It is well known that, in the case of a (single-valued) function, the closedness of the graph is intimately related to the continuity of the function. Here, where $\prod$ is a correspondence rather than a function, we obtain the following continuity result: *the graph* Γ *of the equilibrium price correspondence* $\prod$ *is upper hemi-continuous and compact-valued, if* (D) *holds.*

This is tantamount to the following explicit statement. If (E_n) is a sequence of economies in P^m converging to $E \in P^m$ and if $p_n \in \prod(E_v)$ is an equilibrium price system to E_n for all n, then the sequence (p_n) has a subsequence which converges to an equilibrium price system of the limit economy E, provided (D) holds.

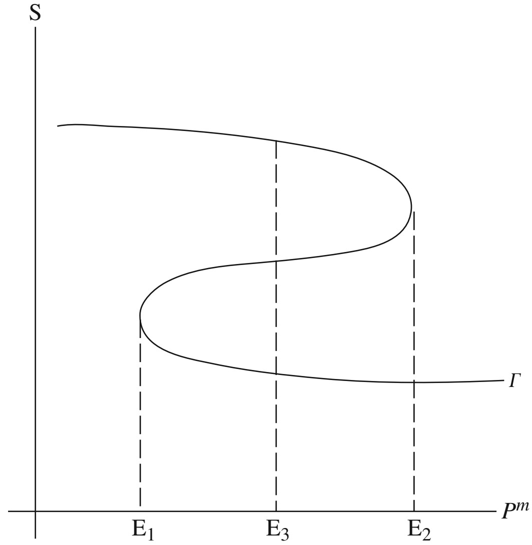
To improve our understanding of Γ, we assume that the demand functions f_i are continuously differentiable (C^1 for short) and we invoke the implicit function theorem in the following manner. Walras's Law allows us to disregard one market, say the lth, and to concentrate on

$$\widehat{Z} : P^m \times S \to \mathbb{R}^{l-1}$$

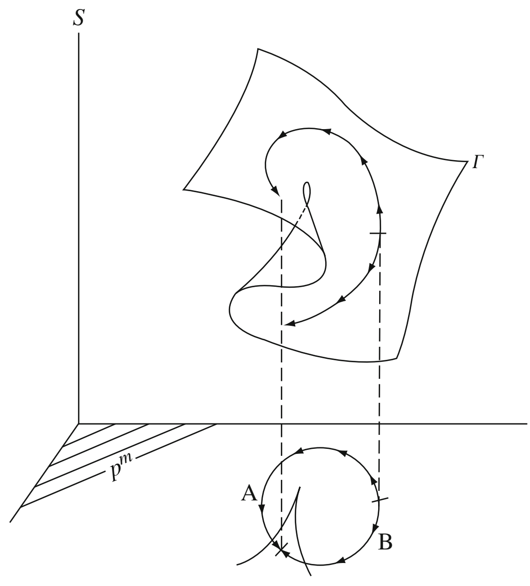
which is obtained from Z by deleting the last component. Let p be an equilibrium price system of E, i.e. $\widehat{Z}(E, p) = 0$. A simple calculation yields that the derivative $d\hat{Z}(E, p)$ has maximal rank at (E, p). Therefore, the graph Γ is given by a smooth surface of dimension lm. That is to say each point in Γ has a neighbourhood in Γ which can be mapped onto an open subset of $\mathbb{R}^{lm}$ by a C^1 diffeomorphism, i.e. a C^1 map with a C^1 inverse. Such a locally Euclidean space is called a C^1 manifold: see Fig. 1.

We have seen that *the graph* Γ *of the equilibrium price correspondence* $\prod$ *is a* C^1 *manifold,* but Fig. 1 suggests more. In Fig. 1, Γ is not only locally Euclidean, there is even a global diffeomorphism between Γ and $\mathbb{R}^{lm}$. Indeed, one can show that this global equivalence holds (see Balasko 1975).

The equilibrium price correspondence is continuous except at two points, E_1 and E_2. If a parameter variation leads through E_1 (or E_2) the equilibrium may be forced to jump. The equilibrium reached after the jump, however, is robust in the sense that no sudden change must occur when one passes through E_1 (or E_2) again. One can imagine a situation such as in E_1 takes place when a slight reduction in the supply of an important raw material leads to a drastic increase in its price. If later on the supply begins to increase again prices perhaps vary but stay at their high level. A reversion of this phenomenon may occur when the supply has reached the much higher level corresponding to E_2.



Regular Economies, Fig. 1



Regular Economies, Fig. 2

In Fig. 2 we have drawn a two-dimensional parameter space. The following remarkable phenomenon may happen here.

There are two paths, A and B, in the parameter space which have their starting point and endpoint in common. Following either path there is no need for the equilibrium to jump. However, the two equilibria reached at the end are quite different equilibria of the same economy. In other words, if two or more policy variables are at one's disposal one must be aware of the possibility that the final outcome depends very well on the order in which the variables are utilized.

The economies E_1 and E_2 in Fig. 1 are characterized by the fact that the graph Γ has a vertical tangent above E_1 and above E_2. Similarly, in Fig. 2, Γ has vertical tangents above all points on the cusp drawn in the bottom plane, which represents P^m. Apparently qualitative changes of the equilibrium price set at an economy E are associated with vertical tangents of Γ above E. This motivates the following definitions. A *critical point* of the projection pr: $\Gamma \rightarrow P^m$ is a point in Γ at which the derivative of pr has rank less than dim $P^m = lm$. A *critical value* of pr: $\Gamma \rightarrow P^m$ is the image of a critical point. A *regular value* of pr: $\Gamma \rightarrow P^m$ is a point in P^m which is not a critical value. Figures 1 and 2 suggest that almost all points in P^m are regular values. Indeed, the concepts introduced above are defined in differential topology in a quite general context and Sard's theorem, an analytical tool of great importance, asserts that the critical values of a sufficiently differentiable mapping are rare. More precisely, *Sard's theorem* applied to our particular problem yields that the set of critical values of pr: $\Gamma \rightarrow P^m$ is a (Lebesgue) null set.

Null sets are small in a probabilistic sense. At this point we make essential use of the space of economies P^m being part of a Euclidean space. If, for instance, consumers' demand functions or preferences are allowed to vary instead of consumers' endowments, it is not clear how null sets are to be defined. However, one can express quite easily when two demand functions or preference orderings are close to each other. That is to say metric structures are very often naturally given when there is no obvious way to define null sets. A set can then be defined to be small in a topological sense if its closure is nowhere dense.

Furthermore, if the concepts of smallness in the probabilistic and in the topological sense are both well-defined, as they are in the case of variable initial endowments, one has to be aware of the fact that the two variants of the intuitive notion of smallness apply to quite different sets. Defining a *critical economy* $E \in P^m$ as a critical value of pr: $\Gamma \rightarrow P^m$ and *regular economy* as a regular value of pr we ask ourselves whether the null set of critical economies has a null closure. We know already that the desirability assumption (D) implies that the equilibrium price correspondence is upper hemicontinuous and compact-valued or, in more intuitive terms, that Γ has only finitely many layers above some compact ball B of economies in P^m. Hence the points in Γ which lie above B and have a vertical tangent form a compact set. Projecting this set down to B yields a compact set, the set of critical economies in B. Since this set is also null by Sard's theorem, it is nowhere dense. We obtain:

The set of critical economies in P^m is a closed null set if (D) holds.

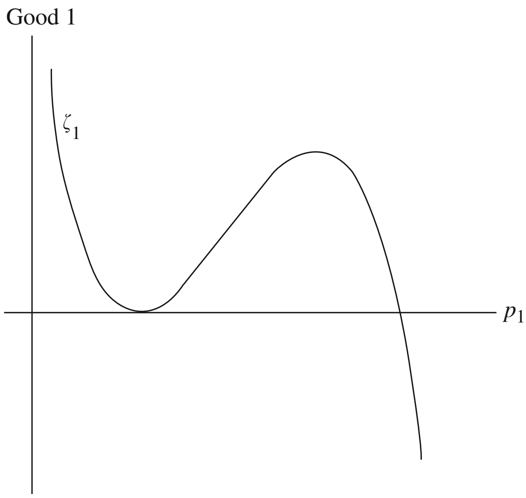
Let $E \in P^m$ be a regular economy. Then E has a finite number of equilibria and this number is locally constant. If E has r equilibria, then there is a neighbourhood U of E and there are rC^1 functions $g_1, \ldots, g_r$ such that the set $\prod(E')$ of equilibrium price systems of any economy $E' \in U$ is given by $\{g_1(E'), \ldots, g_r(E')\}$. In particular, *the equilibrium price correspondence* $\prod$ *is continuous in a neighbourhood of a regular economy.*

These results, with minor differences, have been obtained by G. Debreu (1970), whose proof, however, differs from the exposition given here.

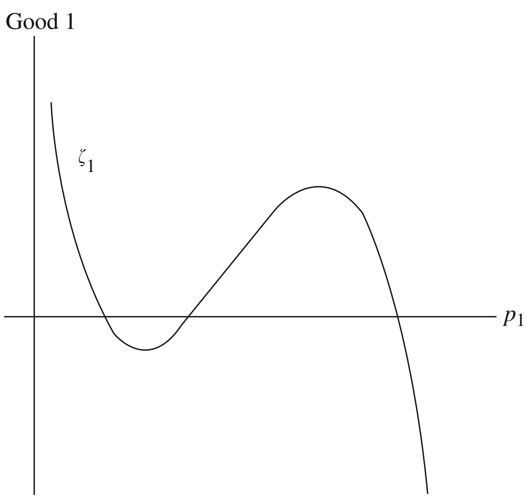
R

Extensions

When one wants to extend the theory of regular equilibria to more general spaces of economies, it is often useful to employ a definition of a regular economy which focuses on the given economy and does not refer to the graph Γ or to the parameter space. To motivate the following definition we contrast Fig. 3, in which the excess demand of a critical economy such as E_1 or E_2 in Fig. 1 is drawn, with Fig. 4, which shows the graph of a regular economy such as E_3. It is assumed that



Regular Economies, Fig. 3



Regular Economies, Fig. 4

there are two goods so that it suffices, according to Walras's Law, to look at the excess demand ζ_1 for good 1.

In Fig. 3 there is one equilibrium at which $d\zeta_1/dp_1$ vanishes. Shifting the graph of ζ_1 a little upwards destroys this equilibrium. In Fig. 4, however, $d\zeta_1/dp_1$ does not vanish at any equilibrium and all equilibria are robust.

Let the excess demand function $\zeta : S \to \mathbb{R}^l$ of an economy E be C^1. A price system $p \in s$ is called a *regular equilibrium price system* if $\zeta(p) = 0$ and the matrix

$$\left[\frac{\partial \zeta_h}{\partial p_k}(p)\right]_{h,k=1,\ldots,l-1}$$

is regular. A *regular economy* is an economy all equilibrium price systems of which are regular. One can show that this definition, introduced by E. and H. Dierker (1972), is independent of the way in which goods are indexed and that it is consistent with the definition given above.

The results on regular economies obtained in various frameworks are quite similar to those established in the previous section. It is shown that almost all economies, in an appropriate sense, are regular. Every regular equilibrium is locally unique and can be traced along its path when the economy varies gradually, as long as it stays regular. Economic models in which results of this kind have been precisely formulated and verified deal with variations in consumption and production (see, in particular, Smale 1974). Also the case of many consumers, that is to say of consumption sectors described by the distribution of consumers' characteristics, has been treated. The basic mathematical tool is always some variant of Sard's Theorem. References can be found in my survey article (Dierker 1982).

The study of regular equilibria has led to a revival of the differentiable viewpoint in general equilibrium theory and related areas. Readers interested in this modern development are referred to the excellent book by A. Mas-Colell (1985), which also contains an extensive, systematic presentation of the theory of regular equilibria.

See Also

► Existence of General Equilibrium
► Global Analysis in Economic Theory
► Mathematics and Economics

Bibliography

Balasko, Y. 1975. Some results on uniqueness and on stability of equilibrium in general equilibrium theory. *Journal of Mathematical Economics* 2: 95–118.

Debreu, G. 1970. Economies with a finite set of equilibria. *Econometrica* 38: 387–392.

Dierker, E. 1982. Regular economies, ch. 17. In *Handbook of mathematical economics*, ed. K. Arrow and M. Intriligator, 795–830. Amsterdam: North-Holland.
Dierker, E., and H. Dierker. 1972. The local uniqueness of equilibria. *Econometrica* 40: 867–881.
Mas-Colell, A. 1985. *The theory of general economic equilibrium: A differentiable approach*, Econometric Society monographs. Cambridge: Cambridge University Press.
Smale, S. 1974. Global analysis and economics IV: Finiteness and stability of equilibria with general consumption sets and production. *Journal of Mathematical Economics* 1: 119–127.

Regulation

Robert Boyer

Abstract

'Régulation' theory analyses the long-term transformation in capitalist economies and their consequences for growth patterns and cyclical adjustments. The degree of coherence of a specific configuration of the major institutional forms – monetary regime, wage-labour relation, form of competition, state–citizen institutionalized compromise and mode of support of the international regime – defines various accumulation regimes and 'régulation' modes. Over one century, several regimes have been observed along with a succession of changing patterns for the related structural crises. The demise of the post-Second World War Fordist regime has been associated with an uncertain process of institutional restructuring and the coexistence of various brands of capitalism.

Keywords

Adjustment; Accumulation regimes; Capitalism; Coevolution; Complementarity; Credit; Diffusion of technology; Fine tuning; Fordism; Hierarchy; Historicism; Information technology; Institutionalism; Isomorphism; Kalecki, M.; Keynesianism; Marxism; Mass production; *régulation*; Stabilization policies; Stagflation; Welfare state

JEL Classifications

K2

Since the 1980s, the term *'régulation'* has suggested state intervention in the name of economic management though its opposite, 'dérégulation', has been more widely used. In the area of economic policy and in accordance with Keynesian precepts, regulation indicates the adjustment of macroeconomic activity by means of budgetary or monetary contra-cyclical interventions. In the area of public management, a complete body of literature, under the name of regulation theory, has investigated the methods for organizing the decentralization of the supply of various public utilities.

This term is also used in physics and biology. In mechanics, a regulator is a means to stabilize the rotary speed of a machine. In biology, regulation corresponds to the reproduction of substances such as DNA. In general terms, the theory of systems involves the study of the role of a set of negative and positive feedback loops in relation to the stability of a complex network of interactions.

Here, a third and different, but not totally unrelated, meaning of the term will be developed. Theories of régulation constitute an area of research which has focused on analysing long-term transformations in capitalist economies. Initially, it focused on American and French capitalisms (Aglietta 1982; Benassy et al. 1979) but it was progressively extended first to major OECD economies (Mazier et al. 1999) then to Latin American countries (Hausmann 1981; Ominami 1985) and ultimately Asian countries (Bertoldi 1989; Boyer 1994). A general presentation of the present state of the theory is to be found in Boyer and Saillard (2002) and a large sample of national case studies in Jessop (2001). Basically, the theory of régulation combines Marxian intuitions and Kaleckian macroeconomics with institutionalist and historicist studies, mobilizing most of the tools of modern economic analysis.

At a primary level, a form of régulation denotes any dynamic process of adaptation of production and social demand resulting from a conjunction of economic adjustments linked to a given

R

configuration of social relations, forms of organization and productive structures (Boyer 1990).

Most economic theories emphasize the general invariables of eminently abstract systems, in which history serves merely as a confirmation or, failing that, as a perturbation. Neoclassical theory studies the shift of a stable equilibrium after an external shock, Keynesian economists stress the role of effective demand and fine tuning whatever the context and the period. Even Marxists tend to extrapolate, as general laws, the quite specific evolutions observed in the early phases of capitalism. In contrast, the régulation approach seeks a broader interaction between history and theory, social structures, institutions and economic regularities (de Vroey 1984).

The starting point is the hypothesis that accumulation has a central role and is the driving force of capitalist societies. This necessitates a clarification of factors that reduce or delay the conflicts and disequilibria inherent in the formation of capital, and which allow for an understanding of the possibility of periods of sustained growth (Boyer and Mistral 1978). These factors are associated with particular regimes of accumulation, namely, the form of articulation between the dynamics of the productive system and social demand, the distribution of income between wages and profits on the one hand, and on the other hand the division between consumption and investment. It is then useful to explain the organizational principles which allow for mediation between such contradictions as the extension of productive capacity under the stimulus of competition on product markets, for labour and finance. The notion of institutional form – defined as a set of fundamental social relations (Aglietta 1982) – enables the transition between constraints associated with an accumulation regime and collective strategies; between economic dynamics and individual behaviour. A small number of key institutional forms, which are the result of past social struggles and the imperatives of the material reproduction of society, frame and channel a multitude of partial strategies which are decentralized and limited in terms of their temporal horizon. Five main institutional forms do shape accumulation regimes.

The forms of competition describe by what mechanisms the compatibility of a set of decentralized decisions by firms and individuals is ensured. They are competitive while the *ex post* adjustment of prices and quantities ensures a balance; they are monopolist if the *ex ante* socialization of revenue is such that production and social demand evolve together (Lipietz 1979). The type of monetary constraint explains the interrelations between credit and money creation: credit is narrowly limited in terms of movement of reserves when money is predominantly metallic; the causality is reversed when on the contrary the dynamics of credit conditions the money supply in systems where the external parity represents the only constraint weighing upon the national monetary system (Benassy et al. 1979). The nature of institutionalized compromises defines different configurations of relations between the state and the economy (André and Delorme 1983; Jessop 1990): the state-as-referee when only general conditions of commercial exchange are guaranteed; as the interfering state when a network of régulation and budgetary interventions codifies the rights of different social groups. Modes of support for the international regime are also derived from a set of rules which organize relations between the nation state and the rest of the world in terms of commodity exchange, migration, capital movements and monetary settlements. History goes beyond the traditional contrast between an open and a closed economy, free trade and protectionism; it makes apparent a variety of configurations (Mistral 1986; Lipietz 1986a). Finally, forms of wage relations indicate different historical configurations of the relationship between capital and labour, that is, the organization of the means of production, the nature of the social division of labour and work techniques, type of employment and the system of determination of wages, and finally, workers' way of life including the welfare state. If, in the first stages of industrialization, wage-earners are defined first of all as producers, during the second stage they are simultaneously producers and consumers.

At this point appears the notion of régulation, as a conjunction of mechanisms and principles of adjustment associated with a configuration of wage relations, competition, state interventions and hierarchization of the international economy. Finally, a distinction between 'small' and 'big' crises is called for (Boyer 1990). The former, which are of a rather cyclical nature, are the very expression of régulation in reaction to the recurrent imbalances of accumulation. The latter are of a structural nature: the very process of accumulation throws into doubt the stability of institutional forms and the régulation which sustains it because the profit does not recover by contrast with conventional business cycles.

Thus, in long-term dynamics as well as in short-term development, institutions are important. Historical research confirms that sometimes institutional forms make an impression on the system in operation; at other times they register major changes in direction. At the end of a period which can be counted in decades, the very mode of development – that is, the conjunction of the mode of régulation and the accumulation regime – is affected: there will be changes in the tendencies of longterm growth and eventually in inflation, specificities of cyclical processes (Mazier et al. 1999).

So a periodization of advanced capitalist economies emerges which is not part of the traditional Marxist theory. Despite the rise in monopoly, the interwar period is still marked by competitive regulation. After the Second World War an accumulation regime without precedent is instituted – that of intensive accumulation centered on mass consumption (Bertrand 1983) – known as Fordist and channelled through monopolist-type regulation.

In fact, the alteration in wage relations – in particular the transition to Fordism, that is, the synchronization of mass production and wage-earners' access to the 'American way of life' – and in monetary management, that is, transition to internally accepted credit money – seems to have played a greater role than the change in modes of competition or conjunctural fine tuning à la Keynes (Aglietta 1982; Aglietta and Orlean 1982; Boyer 1988).

Since the 1960s, many economies have been experiencing a big crisis without historical precedent: stagflation, absence of cumulative depression, breaking-down of most previous economic regularities, length of the period of technological and institutional restructuring (Boyer and Mistral 1978; Lipietz 1985). In consequence, it is logical that former economic policies lose their efficacy (Boyer 1990). First, because the crisis is not cyclical but structural; this invalidates the policy of finetuning; second, because the structural changes which permitted the 1929 crisis to be overcome have become blocked and cannot be repeated (Lipietz 1986b).

Since the formative years, the research programme has been developing both extensively and intensively. The collapse of the Soviet bloc economies has pointed to the need to investigate the necessary and sufficient institutions required for a viable capitalist economy (Emergo 1995; Hollingsworth and Boyer 1997): economic viability depends on the compatibility of a complete set of institutional forms. In the epoch of financialization (Aglietta 1998), information and communication technologies diffusion (Boyer 2004), rise of services (Petit, 1986) and strengthening of foreign competition (Lipietz 1986a), no clear follower to Fordism has yet emerged and diffused. Nevertheless, since the mid-1970s a series of trials and errors concerning the reform of the monetary regime, the tax and welfare system, competition and wage relations has finally delineated a new institutional architecture, quite complex to analyse. Conversion, layering and recomposition of existing institutional forms have replaced the strong synchronization associated with major crises and world wars (Boyer 2005b).

The large number of international comparisons has systematically exhibited the persisting diversity of various brands of capitalism. Within industrialized countries: market dominated, corporate-led, state governed and social democratic versions, with some possible sub-variants, coexist (Amable 2004). An equivalent but different variety is observed for Latin American countries (Quémia 2001). Consequently, the financial crises experienced by Mexico, Brazil and Argentina are quite different, even if they all point out the destabilizing role of global finance upon contrasted domestic accumulation regimes (Boyer and Neffa 2004).

These numerous structural changes call for new directions for the research agenda of régulation theory. Can the concepts of complementarity, hierarchy, isomorphism and coevolution explain how various mixes of institutions can cohere and define a coherent accumulation regime (Boyer 2005a; Socio-Economic Review 2005)? What kind of political economy analysis can explain the emergence and restructuring of institutional forms, especially the choice of monetary regime, the configuration of the welfare state or the nature of insertion into the world economy (Palombarini 1999)? How to analyse multilevel régulation modes, especially in order to understand the complex process of European integration (Boyer and Dehove 2001)? Finally, is not the anthropogenetic model, based on the production of humankind by education, health care and culture, a possible follower of the Fordist regime (Boyer 2004)?

See Also

- ▶ Beveridge, William Henry (1879–1963)
- ▶ Braudel, Fernand (1902–1985)
- ▶ Capitalism
- ▶ Collective Bargaining
- ▶ Competition
- ▶ Contemporary Capitalism
- ▶ Development Economics
- ▶ Growth and Institutions
- ▶ Institutional Economics
- ▶ Kalecki, Michal (1899–1970)
- ▶ Marx, Karl Heinrich (1818–1883)
- ▶ Profit and Profit Theory

Bibliography

Aglietta, M. 1982. *Regulation and crisis of capitalism*. New York: Monthly Review Press.

Aglietta, M. 1998. *Le capitalisme de demain*. Note de la fondation Saint-Simon (101), November.

Aglietta, M., and A. Orlean. 1982. *La violence de la monnaie*. Paris: PUF.

Amable, B. 2004. *The diversity of modern capitalisms*. Oxford: Oxford University Press.

André, Ch., and R. Delorme. 1983. *L'état et l'économie*. Paris: Seuil.

Benassy, J.P., R. Boyer, and R.M. Gelpi. 1979. Régulation des économies capitalistes et inflation. *Revue économique* 30 (3): 397–441.

Bertoldi, M. 1989. The growth of Taiwanese economy 1949–1989: Success and open problems of a model of growth. *Review of Currency Law and International Economics* 39: 245–288.

Bertrand, H. 1983. Accumulation, régulation, crise: un modèle sectionnel théorique et appliqué. *Revue économique* 34: 305–343.

Boyer, R. 1979. Wage formation in historical perspective: The French experience. *Cambridge Journal of Economics* 3: 99–118.

Boyer, R. 1988. Technical change and the theory of 'regulation'. In *Technical change and economic theory: The global process of development*, ed. G. Dosi et al. London: Pinter Publishers.

Boyer, R. 1990. *The regulation school*. New York: Columbia University Press.

Boyer, R. 1994. Do labour institutions matter for economic development? In *Workers, institutions and economic growth in Asia*, ed. G. Rodgers. Geneva: ILO/ ILLS.

Boyer, R. 2004. *The future of economic growth*. Cheltenham: Edward Elgar.

Boyer, R. 2005a. Coherence, diversity and the evolution of capitalisms. *Evolutionary and Institutional Economic Review* 2 (1): 43–80.

Boyer, R. 2005b. How and why capitalisms differ. *Economy and Society* 34: 509–557.

Boyer, R., and M. Dehove. 2001. Théories de l'intégration européenne. *La Lettre de la Régulation* 38: 1–3.

Boyer, R., and J. Mistral. 1978. *Accumulation, inflation, crises*. Paris: PUF.

Boyer, R., and J. Neffa, eds. 2004. *La crisis Argentina (1976–2001)*. Madrid/Buenos Aires: Editorial Mino y Davila.

Boyer, R., and Y. Saillard, eds. 2002. *Regulation theory: The state of the art*. London: Routledge.

de Vroey, M. 1984. A regulation approach interpretation of the contemporary crisis. *Capital and Class* 23 (Summer): 45–66.

Emergo: Journal of Transforming Economies and Societies. 1995. Special issue, 2(4).

Hausmann, R. 1981. *State landed property, oil rent and accumulation in Venezuela*, Ph.D. thesis, Cornell University.

Hollingsworth, R.J., and R. Boyer, eds. 1997. *Contemporary capitalism*. Cambridge: Cambridge University Press.

Jessop, B. 1990. *State theory. Putting capitalist states in their places*. Oxford: Polity Press.

Jessop, B., ed. 2001. *Regulation theory and the crisis of capitalism,* 5 vols. Cheltenham: Edward Elgar.

Lipietz, A. 1979. *Crise et inflation, pourquoi?* Paris: Maspéro.

Lipietz, A. 1985. *The magic world: From value to inflation*. London: Verso.

Lipietz, A. 1986a. New tendencies in the international division of labor. In *Production, work, territory*, ed. A.J. Scott and M. Storper. London: Allen & Unwin.

Lipietz, A. 1986b. Behind the crisis: The exhaustion of a regime of accumulation. *Review of Radical Political Economics* 18 (1–2): 13–32.

Mazier, J., M. Basle, and J.F. Vidal. 1999. *When economic crises endure*. London: M.E. Sharpe.
Mistral, J. 1986. Régime international et trajectoires nationales. In *Capitalisme fin de siècle*, ed. R. Boyer. Paris: PUF.
More information on régulation theory is available from the Association Recherche et Régulation. Online. Available at http://web.upmf-grenoble.fr/regulation. Accessed 19 Oct 2006.
Ominami, C. 1985. *Les transformations dans la crise des rapports nord–sud*. Paris: La Découverte.
Palombarini, S. 1999. *Vers une théorie régulationniste de la politique économique. L'Année de la Régulation 1999*. Vol. 3. Paris: La Découverte.
Petit, P. 1986. *Slow growth and the service economy.* London: Frances Pinter.
Quémia, M. 2001. *Théorie de la régulation et développement: trajectoires latino-américaines. L'Année de la régulation*. Vol. 5. Paris: Presses de Sciences-Po.
Socio-Economic Review. 2005. A dialogue on institutional complementarity. 2 (1): 43–80.

Regulation and Deregulation

Stephen Breyer and Paul W. MacAvoy

Regulation, particularly in the United States, consists of governmental actions to control price, sale and production decisions of firms in an avowed effort to prevent private decision-making that would take inadequate account of the 'public interest'. The Federal Congress established the first national regulatory agency (the Interstate Commerce Commission) in 1887 to control railroad freight rates and passenger fare, and by 1910 had further directed that agency to set limits on charges for long-distance use of the telephone. Since then 14 Federal agencies and upwards of 100 state agencies have been given authority to regulate corporate activity (Domestic Council Review Group on Regulatory Reform 1977).

The legal foundation for regulation consists of statutes allowing the government to grant or condition the right of a company to provide service. Certification or licensing of a common carrier or utility company by a regulatory agency was commonplace by the 1960s in the transportation, communication and energy distribution industries. By approving the tariffs of the licensed company, the regulatory bodies exercised control over almost all railroad, trucking, airline and telephone rates, they also set prices and conditions of sale for electricity and natural gas. Authority to offer service, further, was controlled in respect of commercial or savings banks and liquor distribution (in some states) and even dry cleaners (in Oklahoma; see Plott 1965). But the regulatory mode for achieving public policy objectives in pricing has not been widely used in Europe and Asia; instead, most of these industries have been owned and operated by governmental authorities. Nonetheless, the rationale and the results of regulation in the United States can be used to a degree to gauge the performance of nationalized industry in Europe and Asia.

Federal certification has been extended over other, 'non-utility' industries. Companies that are issuers of securities and commodity contracts have been licensed ostensibly in order to protect depositors and investors. More recently, particularly since the 1970s, the federal government expanded the scope of regulation to encompass health, safety and the environment. It promulgated rules aimed at making safer automobiles, safer consumer products and a safer workplace. These rules often required firms to design their products or production processes to meet specified engineering standards. By 1975 federal regulation significantly controlled important product, price or process decisions by firms in industries accounting for about 24 per cent of gross national product (MacAvoy 1979, p. 25).

These regulatory activities have involved an agency, board, or bureau that (i) contained a governmental bureaucracy and (ii) operated in an adversarial relation with private industry. The agency typically has exercised control by responding to a company request for a licence or tariff change with a decision based on an individual case adjudicatory proceeding or by comparing the request to its own standing rules.

This 'public/private' control relationship has advocates who view it as the method of economic control to be preferred over either the limited forces of competition found currently in private markets or the power in government ownership of enterprises. However, the results from analysis of firm behaviour under regulation do not establish such a preference. Both theoretical and empirical research question the extent to which regulation can achieve the goals for which it has been promulgated.

Reasons for Regulation

Advocates of regulatory programmes assert many different reasons in support of them. Yet a small number of economic arguments are made as 'justification' for kinds of regulation. With market behaviour judged on the efficiency norm that prices equal full social costs at the margin, regulation is required to overcome one or more 'defects' that prevent corporations from operating according to the norm (see Lerner 1964). The 'defects' that have most often led to policy proposals for regulation can be classified as follows.

The Presence of Monopoly Power

When economies of firm scale in a particular market are so extensive as to create a significant cost advantage for a single enterprise, then this 'natural' monopolist can be expected to restrict output or directly set higher than marginal cost prices without concern for entry of a competitor (see Baumol and Bradford 1970). Regulation of such a firm aims to provide cost-driven prices, but it cannot achieve 'efficiency', since to do so would require ending up with per-unit revenues below the level necessary for form survival (with marginal less than average costs, the natural monopolist cannot be required to set prices equal to marginal costs; see Lerner 1964). And where the natural monopolist can discriminate among customers and on sales to a single customer, regulation is unnecessary because profit-maximizing prices on marginal sales will equal marginal costs. Thus the rationale for regulation of the natural monopolist has settled for less than the norm, such as constraining the monopolist to set average prices at average costs (see Kahn 1970, vol. 1).

A variant of the argument has been used to regulate both monopolies and competitive industries. When and where 'rents' paid to scarce resources are unusually large in amount, and are the consequence of sharp, unexpected price increases of a widely used product, regulation has been sought to keep prices down. For example, the regulation of natural gas at the wellhead in the 1960s and of petroleum products at retail in the 1970s were for this purpose, without regard for the competitiveness of the market. But where there is competition, supply will fall short of demand; then regulation has to be justified on grounds that the shortage for a few does less harm than the good to come to the rest of the consumers from keeping prices down (see Kahn 1960).

To Account for Spillover Costs

Regulation has been proposed where a product's price does not reflect important costs inherent in the production process – costs that are imposed on neighbours or others in the economy. The price of electricity may not reflect the full cost of air pollution that results from using coal in power generation. If not, demand for electricity is greater than the norm of economic efficiency would dictate (since buyers do not pay the full marginal costs).

Of course, the harmful results of pollution derive both from the electricity generation process and the fact that people have moved into the area bordering the plant. In theory, electricity consumers and pollution sufferers would agree to share optimally the cost of antipollution devices if they could bargain efficiently among themselves (see Coase 1980). Such bargaining typically has been found by legislatures to be impractical, however, as compared with direct regulation of emissions or discharges. Environmental regulation, as a way to constrain important spillovers, has attacked the pollution-generating process by setting engineering standards for equipment used in the production process – standards of which the aim is to reduce discharge.

The effectiveness of such regulatory controls has been much debated. To what extent they have been successful in taking account of 'spillover' costs in setting standards and in actually reducing pollution or increasing safety have been major issues in applied research. Local controls and penalties in the form of higher insurance charges, court damage awards, etc. existed before Federal environmental and auto safety regulation had been established. Some argue that strict engineering standards imposed through federal regulation have not done much better (see Lave 1981; Crandall 1983; MacAvoy 1986; but see Freeman 1982). Others agree, but state that taxes on pollution or unsafe cars or other incentive-based systems could be effective even if current methods were not (see Breyer 1982; Stewart 1985).

To Compensate for Inadequate Information

Regulation sometimes aims at lowering the costs of obtaining information. In particular, government action has been called for when (i) suppliers profit from misleading consumers whose available legal remedies, such as private court actions, are more costly than regulation; (ii) consumers cannot readily evaluate the information available, and the costs of mistakes are high, such as on potential drug effectiveness or safety of a particular airline; and (iii) the market on the supply side fails for some reason to furnish the information as is demanded (at cost-based prices). Given the first and second reasons, the government has created special commissions to license 'safe' goods and services. For the last reason, the government may seek to provide more, if not better, information or to require producers to supply the information, as in the case of financial or securities disclosures. Most such approaches have been made without specific knowledge of markets for information or empirical studies on the 'failure' of current sources.

Other Justifications

Of course, there are special-interest arguments for regulation. Price and entry regulation of airlines, trucking, and ocean shipping, for example, have been justified to control 'excessive' competition that allegedly would destroy all except one or two firms that would then set noncompetitive prices. This argument is advanced by incumbent firms. The cost functions implicit in these depictions of behaviour are not widely found in these industries, however. Regulated firms have at times advanced a similar argument in an effort to extend regulation of entry to encompass the prices set by competitive, unregulated rivals. In the telephone industry, regulation of entry has been sought to prevent 'predatory' pricing on long-distance services in which the local service Bell Operating Company would allegedly set unregulated long-distance rates below variable costs to drive rivals out of that business, covering its own losses with high regulated returns on local service (see MacAvoy and Robinson 1985; Brock 1981). But unless regulation of local service itself is so defective as to allow 'padding' of regulated costs with losses on other services, firms cannot readily set such prices and profit from continually keeping others out of cross-subsidized markets. The argument is one in which ineffective regulation mandates regulation of another market.

A further justification in support of regulatory policy is that a source other than the consumer makes the purchasing decision, thus leading to inefficient increases in purchases. Medical care is often cited as an example, for it is the patient's insurance company or the government, not the patient, who pays the bill. And government regulation of medical service charges is sometimes advocated for this reason. This rationale may be viewed as a statement of belief that demand is more inelastic with respect to price than if there were only first-party purchases. For such price non-sensitivity to result in adverse market performance, suppliers would have to possess monopoly power, so that this is a version of the first (monopoly) rationale.

Unequal bargaining power sometimes is used as a rationale for regulation that would protect small firms, suppliers or customers from the power of the large firms or buyers with whom they must deal. State regulators, for example, prescribe standard forms for insurance contracts. This again is a monopoly rationale, stated from the viewpoint of the corporate purchaser of goods and services. The operationality of these arguments determines the worth of regulation. There has to be a direct relationship between rationale and results so as to establish a foundation for regulatory policy. Peltzman, whose work is based on earlier developments of G.J. Stigler, has established hypotheses which are based on the 'public interest' rationale, which can be tested against alternative hypotheses based on the claim that the regulatory agencies restrain market operations for those private interests willing to compensate such agencies for those services (Peltzman 1976; Stigler 1971). Thus, which is the 'true' rationale for a particular programme can be determined by examining regulatory methods and results.

Methods of Regulation

As important as rationale are the means used for obtaining the desired results. Although all regulation involves a governmental bureaucracy that 'commands and controls' the individual actions of private firms, one can distinguish between different regulatory programmes according to the specific methods used to effectuate that 'control'. More to the point, certain regulatory methods have been used to attain specific regulatory ends. They include the following.

Cost of Service Ratemaking

This system has been commonly used to set prices in the electricity industry as well as for local and long-distance telephone service industries and for airlines, road transport companies and railways. In principle, the regulator determines a revenue requirement based upon the firm's accounting costs during a 'test year'. Income accounts at issue include operating costs, taxes, an allowance for depreciation and 'allowed returns' defined as a 'reasonable' rate of return times the 'rate base' including the undepreciated portion of investments relevant to regulated operations valued on a historical expenditure basis. Once the revenue requirement is determined, the regulator approves rates in a tariff designed to recover, during the foreseeable future until another proceeding, the amount of the revenue requirement.

A host of economic problems arise in transforming this practice into a set of bureaucratic procedures administerable through adversary hearings (see Kahn 1970, vols 1 and 2). In principle, assuming market and accounting values of assets are the same, the process would set the 'allowed returns' rate equal to the firm's costs of capital, so that the level of prices would equal the firm's long-run average total costs of operation. The anti-monopoly rationale would be achieved, if only in the general sense that the monopoly could not set prices resulting in excessive returns. But this general rule cannot bring about prices equal to marginal costs, nor can it determine optimal departures of individual prices from the marginal costs of specific services (see Averch and Johnson 1962; Baumol and Klevorick 1970). And when applied with accounting data under economy-wide conditions if inflation and low growth of demands for regulated services, it can produce results that deviate widely from even the suboptimal goal inherent in average cost pricing (see Joskow 1974; Joskow and Noll 1981; MacAvoy 1982).

Setting Historically Based Price Ceilings

This system, used during wartime rationing, has been adapted to controlling petroleum product prices, and it has been proposed for controlling hospital care fees (see Kalt 1981; Joskow 1981). It consists of holding prices at their level of a certain past date (e.g. 'last August 1') but then allowing increases above that level justified by changes in operating costs. Ceilings require continuous adjustment as administrators, for example, cope with new products, changing demand and the necessarily resulting shortages and deterioration of service quality.

Issuing Permits

Agencies charged with allocating a commodity in scarce supply, such as television bandwidth or airline landing rights, have developed systems that allow them to choose among applicants after public hearings at which each sets forth qualifications. The system requires the agency first to define the precise commodity awarded, and then to apply standards that weed out those not qualified. Finally, the agency selects among competing applicants, subjectively deciding which will best serve the 'public interest' on the basis of their presentations.

This system could embody a process by which to achieve any of the goals in regulation. Indeed, by refusing to issue a permit, the agency can prevent an enterprise from operating as a monopoly or from causing spillover costs. The creative use of such operating authority has been advocated as the method for achieving the efficiency norms, without rate-base regulation (Demsetz 1968; Bailey and Baumol 1984). But in practice, licensing has not been an important substitute for price or spillover regulations. When applied to allocating such 'public goods' as bandwidth, it has often proved to be difficult or impossible to find any meaningful set of coherent criteria that will allow choice among qualified applicants. Instead, the system tends to spawn complex and extended hearings (all characteristics of applicants are treated as relevant), and the results have been open to charges of inconsistency or special-interest distortion (see Wilson 1979).

Standard Setting

Regulators set increasingly large numbers of standards forbidding, for example, methods or products that are unsafe or spread pollution. Standards typically require the regulator to obtain information from a wide range of interested parties: the industry, consumer groups, suppliers, customers, employees, other government bodies, and so forth. The regulator decides such questions as: (i) Should the standard directly control the relevant evil (such as pollution content) or control a surrogate discharge (such as smokestack or the sulphur content of coal?) (ii) How stringent should the standard be? (iii) To what extent should the standard embody a requirement for equipment performance rather than specification for a particular equipment design?

In practice, the agency considers standards proposed by firms before modifying and finally adopting them. It will receive comments from different parties and then 'negotiate' a final compromise among conflicting proposals, so as to reduce opposition by those regulated that could take the form of court suits or resistance to compliance which generates large enforcement costs.

Each of these four regulatory systems has proved to be controversial – with lengthy arguments about whether their application has costs that make the regulatory 'cure' worse than the 'disease' to be prevented. Operations under rate-base regulation are confounded by the fact that current costs seldom bear any relation to unregulated prices set on the basis of allowed returns on previous investments. The use of price ceilings can be worse, as last-period prices become more anachronistic. With permits and standards, the regulator must deal with the fact that tests for various potential results required for certification are often inconclusive and thus regulatory choices of operating conditions are random. As a result, the case for regulation has been embraced less enthusiastically in recent experience. Also, the results of regulation have cast doubt on how well this system has lined up to its 'public interest' rationale.

The Effects of Regulation

Enacting legislation to establish regulatory commissions is only the beginning. The impact of regulation is largely determined by the way an agency translates statutory goals into operating rules, so as to bring about changes in prices, sales and service quality from company adjustments to these rules. Most of the agencies regulating price and entry had their programmes in place by the 1960s, so that results can be documented.

The public utility and transportation industries showed little price-level effect from regulation before 1970 (see Moore 1970; Jackson 1969). During the 1970s, however, regulation had excessive price-reducing effects. The commission process became more constraining as companies with annual increases in 'justified' costs requested increases in rates that were denied by the regulatory agencies.

Telephone regulation, split between state and Federal commissions, experienced both a rate-level freeze and, as well, the tendency of commissions to shift any increases away from home consumers. Since significant portions of total costs were derived from the common operations of both local and long-distance systems, the agencies were able to assign more 'costs' to interstate long-distance services, thereby allowing the state regulators to hold down local residential charges (see the Federal Communications Commission Report, FCC Docket 20,003). But at the same time, long-distance rates were held constant while costs specific to this service fell, opening up a profit margin to cover the shift in joint costs (MacAvoy and Robinson 1983).

Natural gas, not a monopoly industry, was regulated in the 1960s on the variant rationale of controlling 'rents' with the result that supply shortages had developed of as much as one quarter of total demands by the middle 1970s. The agency maintained wellhead prices at approximately the level that was realized in fixed markets before regulation began through ceiling which 'area rates' kept new contract prices at the level of average historical costs of production from inground reserves in any region. The system could not conceivably have worked to achieve price stability and sufficiency of supplies at the same time: gas demand increases, as a result of lower prices for gas relative to other fuels, exceeded the GNP and total energy consumption growth rates each year; commensurate supply increases were forthcoming only at marginal costs higher than average historical costs, which because controlled price ceilings were based on average costs, would not be undertaken (MacAvoy and Breyer 1974; MacAvoy 1983).

The airline industry showed the same results over the 1970s. In reaction to regulation and inflation, they reduced service quality significantly in the early part of the decade. Throughout this period the number of flights declined, and airline passengers were offered less convenient scheduling or more crowded flights. But the Civil Aeronautics Board also imposed fare structure rules that prevented selective cuts where demands were elastic and lower-than-average cost conditions would lead to expansion of service (Breyer 1982). Airlines at the same time were offering service at rates which in respect of capacity were too low, and rates specifically which were too high (Breyer 1982; Miller 1977). The railroads realized somewhat the same results, but for slightly different reasons. In the presence of rising fuel, labour and capital costs, the Interstate Commerce Commission granted revenue increases that kept the rail rate index in line with average costs. But the commission did not allow reductions of service on lines experiencing greater-than-average cost increases. The railroads thus had to continue to provide for small shippers, those on short-distance lines and those seeking small-volume but frequent service, even though unit costs for these services increased more rapidly than revenues. To meet service requirements and still earn profits, rate-cost margins were increased in high-volume and longdistance markets were subject to incursions from competitive trucking suppliers. Thus railroads were faced with increasing the rate of market attrition to stay at an average price level covering costs (Coleman 1977).

These results are the product of two distinct regulatory processes: (i) in the rate of return case loads created a regulatory lag that began to work against rather than for the unregulated companies; and (ii) the commissions, for political not efficiency reasons, kept down the size of current-dollar rate increases (in keeping with Peltzman hypotheses; see Peltzman 1976). When costs began to rise in the late 1960s, regulatory lag penalized the firm – and the difference between historical and current costs widened as the increase in the number of case applications extended the amount of time required for case decisions. Thus the greater the inflation, and the longer the lag in deciding on increases in regulated prices, the greater the profit-reduction effect of controls on these industries. Further, beyond the clerical problems of regulatory lag, rate increases when granted were not enough to compensate for cost increases. Simply to avoid adverse public reaction, rate-setting agencies would not grant price increases that were very large in billions of dollars. The dollar sizes of proposed additional revenues so concerned regulators that they became reluctant to grant even those increases that were fully justified on the efficiency norm.

As the impact of the over-regulation was beginning to be felt by the public utilities and the transportation and communications industries in the late 1960s, the manufacturing industries were just beginning the new experiment with health, safety and environmental regulation. New agencies had been established to regulate company operations using the permit powers to force detailed equipment specifications. Because such standard setting has been litigious and prolonged, these regulations have proven to be extremely detailed and inflexible. They have forced substantial investments made to meet equipment specifications, which in turn have caused increases in production costs in the industries most subject to the new controls. Subsequent price increases in those industries were greater during the early years of health and safety regulation than in those industries not as subject to such regulation (the exceptions were the electricity utilities and the petroleum-refining industry, which were price-regulated through all or part of the period) (MacAvoy 1982). Consumers in effect paid for the equipment outlays required by health and safety regulations.

These cost increases might have been compensated for by less pollution and fewer industrial and highway accidents. Analysts, however, have been unable to find significant reductions in the unhealthful conditions which were to be dealt with by the new regulatory activities. Research on worker safety regulation has indicated that there have not been widespread reductions in worker accident rates from agency activities (Smith 1976; de Pietro 1976; Viscusi 1983; Nichols and Zeckhauser 1977). Regulation of automobiles produced somewhat the same results. In its early years of operation, the Federal Highway Safety Agency directed its regulatory activities towards improving crash survivability, with a goal being to decrease the fatality rate per 100 million vehicle-miles by one-third (Second Annual Report on the Administration of the National Traffic and Motor Vehicle Safety Act of 1969). But recent highway accident rates have been better explained by (i) the cost of accident insurance, (ii) personal income levels, (iii) driving speed, (iv) driver age, (v) alcoholic intoxication, and (vi) a secular trend than by regulatory actions (Peltzman 1975; Manne and Miller (eds) 1976). When the statistical equation for this relation in the pre-regulation period was used to generate predictions of accidents for the period after mandatory safety devices were introduced, it was found that projected highway fatality rates without regulation differed very little from actual rates under regulation.

The benefits of environmental regulation have not been substantially more important than those from the other major ‘social control’ initiatives. To be sure, environmental controls have realized positive results at certain locations – where the environmental agencies have invoked rules against pollution by a particular company. Certain rivers and air corridors have been made cleaner than they were five years ago because of such actions (Conservation Foundation 1982). How widespread the improvement is, however, is debatable. Since standards for each industry and state have been different and have been enforced to widely differing degrees, it cannot be said that regulation has improved air and water quality nationwide. At the same time, industryspecific standards were being put in place after product and process improvements already underway were beginning to reduce pollution. Pollutants were being reduced because it was profitable, given new technology, to conserve inputs that had been previously discharged as waste (Crandall 1983; Mills and White 1978; see Mills and White 1978, on automative emissions regulation). Thus, recent air- and water-quality changes may have been affected by changes in industrial activity, not simply by regulation.

On the whole, the health and safety regulatory systems have most likely increased prices and reduced GNP in the most regulated industries. Also, there are indications that because of drawing attention towards equipment and away from behaviour, the control system had not brought about improvements in health, safety and environmental quality that could have followed from use of other regulatory methods. In other words, these regulatory agencies have been generating substantial cost effects, but one cannot be certain they have brought about the benefits intended in the enabling legislation.

Deregulation

Beginning in the mid-1970s, public dissatisfaction with the burdens that regulation imposed, combined with economists' criticism of specific regulatory programmes, created a strong political movement bent upon ending many particular regulatory programmes. The first major programme to be dismantled was airline regulation. Economists had long criticized the regulatory regime imposed on airlines, claiming that the 'excessive competition' rationale was a chimera, without application to the structurally competitive airline industry. Since the elimination of the Civil Aeronautics Board in 1978, most of the facts have borne out the evaluation of regulation as operating a decade earlier. Regular fares have risen on regularly scheduled services but travellers have tended to fly on discount fares, and in such cases, where demands are elastic and costs lower than average, fares have fallen (Morrison and Winston 1986; Meyer and Oster 1984; Bailey et al. 1985). By most economic measures, the industry after deregulation has been operating more efficiently (Morrison and Winston 1986). At the same time, even though some firms have faced the threat of bankruptcy, the average firm has increased profitability. The gains from service flexibility, in cost reduction, have allowed the more efficient airlines to offer more service between smaller and medium-sized cities resulting in consumer gains valued by Morrison and Winston at $6 billion.

On the heels of airline deregulation, other deregulatory legislation has followed. To a substantial degree, rate-of-return regulation has been eliminated or partially replaced by increased reliance upon competition as the determinant of prices and sales in trucking, railroads, stockbroking, long-distance telephone service and, to a somewhat lesser degree, banking. Each of these industries, however, satisfied (to a considerable degree) the structural economic preconditions for maintaining competition since their markets were usually large enough to support several competing firms of efficient size. The deregulatory movement has not advanced to the point of reducing the coverage of health, safety and the environment regulation given that the rationale for intervention is stronger and the market alternatives to classical regulation less obviously superior. In these latter areas, 'regulatory reform pressure' has taken the form of advocating, not total deregulation, but rather less restrictive or less burdensome methods of governmental intervention aimed at achieving the relevant regulatory end.

Most prominently, economists have proposed that the government create saleable rights to engage in limited undesirable conduct, such as polluting. With such rights, and a market for their sale, firms would find it costly to pollute. A system of marketable rights allows the firms and the regulatory agency to know in advance the amount of pollution (equal to the sum of unit rights issued), but not to know in advance the price of continuing to use a process that emits pollutants.

At the present time, significant progress has been made in 'deregulating' those industries where the economic case for controls has never been strong, and recent results from regulatory operations have caused significant declines in the quality of service. Whether 'reform' will go further to encompass 'less restrictive alternatives' to regulation is the theoretical and empirical research issue of the 1980s and 1990s. To date, however, the agenda in welfare economics for reform differs significantly from the political agenda to the point where regulation may be present for more than these decades.

See Also

- ▶ Communications
- ▶ Energy Economics
- ▶ Industrial Organization
- ▶ Market Structure

Bibliography

Ackerman, B., and R. Stewart. 1985. Reforming environmental law. *Stanford Law Review* 37: 301.

Averch, H., and L.L. Johnson. 1962. Behavior of the firm under regulatory constraint. *American Economic Review* 52: 1052–1069.

Bailey, E.E., and W.J. Baumol. 1984. Deregulation and the theory of contestable markets. *Yale Journal on Regulation* 1(2): 111–138.

Bailey, E.E., D.R. Graham, and D.P. Kaplan. 1985. *Deregulating the airlines*. Cambridge, MA: MIT Press.

Baumol, W.J., and D.F. Bradford. 1970. Optimal departures from marginal cost pricing. *American Economic Review* 60(3): 265–283.

Baumol, W.J., and A.K. Klevorick. 1970. Input choices in rate-of-return regulation: An overview of discussion. *Bell Journal of Economics* 1(2): 162–190.

Breyer, S. 1982. *Regulation and its reform*. Cambridge, MA: Harvard University Press.

Brock, G.W. 1981. *The telecommunications industry: The dynamics of market structure*. Cambridge, MA: Harvard University Press.

Coase, R.W. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1–44.

Coleman, W.T., Jr. 1977. Time for corrective action. In *Railroad revitalization and regulatory reform*, ed. P.-W. MacAvoy and J.W, 63–70. Snow, Ford Administration Papers on Regulatory Reform, American Enterprise Institute, Studies in Government Regulation.

Conservation Foundation. 1982. *State of our environment*. Washington, D.C.: The Foundation.

Crandall, R. 1983. *Controlling industrial pollution: The economics and politics of clean air*. Washington, DC: Brookings Institution.

Demsetz, H. 1968. Why regulate utilities? *Journal of Law and Economics* 11: 55–65.

de Pietro, A. 1976. An analysis of the OSHA inspection program in manufacturing industries, 1972–1973. *Draft Technical Analysis Paper*, US Department of Labor.

Domestic Council Review Group on Regulatory Reform. 1977. *The challenge of regulatory reform: A report to the president*. Washington, DC: Government Printing Office.

Freeman, A.M. 1982. *Air and water pollution control: A benefit cost assessment*. New York: Wiley.

Harter, P. 1982. Negotiating regulations: A case for malaise. *Georgetown Law Journal* 71(1).

Jackson, R. 1969. Regulation on electric utility rate levels. *Financial Economics* 45: 370–376.

Joskow, P.L. 1974. Inflation and environmental concern: Structural change in the process of public utility price regulation. *Journal of Law and Economics* 17(2): 291–311.

Joskow, P.L. 1981. *Controlling hospital costs: The role of government regulation*. Cambridge, MA: MIT Press.

Joskow, P.L., and R. Noll. 1981. Regulation in theory and practice: An overview. In *Studies in public regulation*, ed. G. Fromm, 1–65. Cambridge, MA: MIT Press.

Kahn, A.E. 1960. Economic issues in regulating the field price of natural gas. *American Economic Review* 50: 507.

Kahn, A.E. 1970–71. *The economics of regulation,* 2 vols. New York: Wiley.

Kalt, J.E. 1981. *The economics of politics of oil price regulation*. Cambridge, MA: MIT Press.

Lave, L. 1981. *The strategy of social regulation*. Washington, DC: Brookings Institution.

Lerner, A.P. 1964. Conflicting principles of public utility rates. *Journal of Law and Economics* 7: 61–70.

MacAvoy, P.W. 1979. *The regulated industries and the economy*. New York/London: W.W. Norton.

MacAvoy, P.W. 1982. *Energy policy: An economic analysis*. New York: W.W. Norton.

MacAvoy, P.W. 1986. The record of the environmental protection agency in controlling industrial air pollution. In *Energy, markets and regulation: What have we learned?* Cambridge, MA: MIT Press.

MacAvoy, P.W., and S. Breyer. 1974. *Energy regulation by the federal power commission*. Washington, DC: Brookings Institution.

MacAvoy, P.W., and K. Robinson. 1983. Winning by losing: The AT&T settlement and its impact on telecommunications. *Yale Journal on Regulation* 1(1): 1–42.

MacAvoy, P.W., and K. Robinson. 1985. Losing by judicial policymaking: The first year of the AT&T divestiture. *Yale Journal on Regulation* 2(2): 225–262.

Manne, H.G., and R.L. Miller (eds.). 1976. *Auto safety regulation: The cure or the problem?* Glenridge: Thomas Horton.

Meyer, J.R., and C.V. Oster Jr. (eds.). 1984. *Deregulation and the new airline entrepreneurs*. Cambridge, MA: MIT Press.

Miller, J.C., III. 1977. The effects of the administration's draft bill on air carrier finances. In *Regulation of passenger fares and competition among the airlines*, ed. P.-W. MacAvoy and J.W. Snow, 181–200. Ford Administration Papers on Regulatory Reform, American Enterprise Institute for Public Policy Research: Studies in Government Regulation.

Mills, E.S., and L. White. 1978. Government policies toward automotive emissions control. In *Approaches to controlling air pollution*, ed. A.F. Friedlaender, 348–409. Cambridge, MA: MIT Press.

Moore, T.G. 1970. The effectiveness of regulation of electric utility prices. *Southern Economic Journal* 36(4): 365–375.

Morrison, S., and C. Winston. 1986. *The economic effects of airline deregulation*. Washington, DC: Brookings Institution.

National Highway Traffic Safety Administration. 1969. *Second annual report on the administration of the National Traffic and Motor Vehicle Safety Act*. Washington, DC: Government Printing Office.

Nichols, A.L., and R. Zeckhauser. 1977. Government comes to the workplace: An assessment of OSHA. *The Public Interest* 49: 39–69.

Peltzman, S. 1975a. *Regulation of automobile safety*. Washington, DC: American Enterprise Institute for Public Policy Research.

Peltzman, S. 1975b. The effect of safety regulation. *Journal of Political Economy* 83(4): 667–725.

Peltzman, S. 1976. Toward a general theory of regulation. *Journal of Law and Economics* 19(2): 211–248.

Plott, C.R. 1965. Occupational self-regulation: A case study of the Oklahoma dry cleaners. *Journal of Law and Economics* 8: 195–222.
Smith, R.S. 1976. *The Occupational Safety and Health Act*. Washington, DC: American Enterprise Institute for Public Policy Research.
Stewart, R. 1985. Economics, environment and the limits of legal control. *Harvard Environmental Law Review* 9: 1.
Stigler, G.J. 1971. The theory of economic regulation. *Bell Journal of Economics* 2(1): 3–21.
Suskind, L., and G.G. McMahon. 1985. The theory and practice of negotiated rulemaking. *Yale Journal on Regulation* 3: 133.
US Senate Commission on the Juduciary, Subcommittee on Administrative Procedure. 1975. *Civil Aeronautics Board producer and procedures*, 94th Congress, 1st Session.
Viscusi, W.K. 1983. *Risk by choice: Regulating health and safety in the workplace*. Cambridge, MA: Harvard University Press.
Wilson, J.Q. (ed.). 1979. *The politics of regulation*. New York: Basic Books.

Regulatory Responses to the Financial Crisis: An Interim Assessment

Howard Davies

Abstract

The global financial crisis that started in 2007 exposed the failure of governments and legislative bodies to regulate the banking industry adequately at a national and global level. Post-crisis reforms and structural changes have ensued: the switch from the G7 to the G20 for example and an attempt to strengthen the centre of the international financial system. The EU now has to decide whether a fiscal and banking union is the logical next step to complete economic integration and to stave off future crises in the Eurozone. In the UK, the Bank of England has assumed greater responsibility for regulation. Whether such responses to this crisis will safeguard against future catastrophe is unknown, but there are already signs that the response so far will prove to be inadequate.

Keywords

Banking crises; Bank regulation; Financial policies; Financial crisis; Financial Stability Forum; Bank of England; ECB

JEL Classifications

G2; G21; G32; G38

"It's important not to waste a good crisis". That witty phrase has graced the first paragraph of many a speech on the latest financial crisis, so many, indeed, that it has now passed from witticism to cliché- another casualty of the turbulence of the last four years. Nonetheless, it remains a wise observation in relation to financial regulation. Most changes in regulation happen in response to a crisis, whether global, regional or national.

Crisis-Led Reform

We can look for proof of that contention by going as far back as the establishment of the SEC in the 1930s, following the Wall Street crash. More recently we owe the creation of the Financial Stability Forum to the Asian financial crisis of the late 1990s. US companies and their auditors can thank Enron for the Sarbanes Oxley Act. In Europe the beginnings of pan-European regulation, set out in the Lamfalussy report (1), can be found in the response to the dot com boom and bust. The move to single, integrated regulators, which has influenced many countries in recent years, in fact began as a response to severe banking problems in Scandinavia in the early 1990s. There are many other examples one could cite of regulatory change introduced hastily in crisis conditions.

The UK's last (but one) reform, the creation of the Financial Services Authority, might be thought to be a counter-example, as it was not implemented immediately after a collapse or a scandal. It was in part a delayed response to the Baring's debacle, when the Bank of England was perceived to have been asleep at the switch as Baring's pursued a risky trading strategy in

Singapore whichought about its collapse, but the prime motivation of the 1997 reform was the desire on the part of the new Labour government under Tony Blair to create an independent monetary authority. The switch of banking supervision out of the Bank of England was seen as a consequence of that reform, and the FSA was created by a side wind, so to speak. I know from personal experience that little thought was given by politicians to what the FSA might be, or what it would do.

That provides one clue as to why most reforms of financial regulation take place in response to a crisis. The blunt truth is that, for most politicians, most of the time, banking supervision does not feature among their top one hundred interests. Uncomfortable though it may be to admit it, especially for those of us whose lives have been consumed by preoccupations with loss given default rates, or the arcana of market transparency requirements, these are not usually topics of animated debate on the doorstep or in the wine bar – outside a few traders haunts in London EC4 or Lower Manhattan. Politicians reflect that reality. In calmer times they are reluctant to devote scarce legislative time to the structure or mandates of supervisory agencies. The subject bores them. Far more fun can be had with pork barrel spending Bills, or in going to war with a country without too many voters in one's own district. Only when markets go into spasm, and the public authorities have to step in with their cheque books at the ready, do legislators bend their minds to the issues. At that point it becomes abundantly clear that "something", must be done, and that "something" is usually a raft of legislation giving regulators new powers to secure the doors of all the empty stables in their jurisdiction, or structural reform, or a combination of the two.

There is a serious drawback to this way of proceeding. Policy-making in a crisis is fraught with hazards. Calm analysis, and evidence-based policy-making, are at a premium. In those circumstances it is easy to overreact, and to lose sight of the wood in focussing close attention on the trees. There are many examples around the globe of legislation which follows the doubtful sequence of "Load- Fire – Aim", not always a guarantee of success. In the UK the most celebrated example of this phenomenon is the Dangerous Dogs Act of 1991, passed by an outraged Parliament after a dog mauled a baby in its pram in South London. That was undoubtedly a very unhappy experience for the child's family, not to mention for the baby herself. But the resulting legislation brought Parliament into disrepute. The law says that certain breeds of dog must be muzzled at all times in public. I can honestly say that I have never seen one so accoutred. In this case, the law has definitely proved to be an ass. Sometimes, bad things happen, and no law will entirely prevent them.

Sadly, much financial regulation is in the Dangerous Dogs Act territory: a well meaning, but hasty overreaction to an unfortunate episode. That is the occupational hazard of crisis-induced reform.

The financial crisis which began in 2007 certainly demanded a political response. The losses were so large, and the economic consequences so severe that any politician who responded on the Rumsfeldian lines of "Hey, stuff happens", would have soon found herself looking for another line of business. So globally, and in individual countries, a response has been pieced together. We can now see the broad lines of that response, even though some of the detail remains to be worked through. How should we evaluate what has been done? Have the reforms made the world a safer place for future investors, or indeed taxpayers, who have been reluctant recapitalisers of the financial system? Or is the Law of Unintended Consequences coming into play?

One cannot attempt an answer to this question without offering a view on how far regulation was indeed at fault, and how significant regulatory failings were in the run-up to the crisis. In answering that question one should also try to distinguish between those failings which arose from weaknesses in the powers accessible to regulators, or in the structure of regulatory authorities, and how far they arose from errors made in the exercise of those powers. If "human error" was at issue, then there is in principle no need to change the law or the institutional structures. Instead we should change the people, and hope the new crew do better next time round.

Regulatory Failings Pre-1997

My own short answer to this question is that regulatory failings did play a part, but that those failings were by no means the only, or even necessarily the most significant factors. I have presented elsewhere a comprehensive taxonomy of the crisis (2). My principal argument there is that global imbalances, combined with relatively loose monetary policy, created the conditions in which leverage expanded rapidly. The monetary authorities on both sides of the Atlantic focussed attention on retail price inflation, and assumed that control of inflation was a sufficient condition to maintain financial stability. In that environment the incentive structures within financial firms pushed them to take on greater risks. In some cases senior management had a poor understanding of the risks they were taking on, blinded by the complexity of new and dangerous products. As a result, when asset prices began to fall, and a liquidity squeeze developed, a number of markets collapsed like packs of cards.

At that point, I argue "it became clear that financial regulators had not been tough enough, particularly in their approach to capital reserving, to constrain risktaking or to ensure that institutions were robust enough to cope with a period of severe stress". In my view it is wholly unrealistic to expect regulation to be the front-line defence against booms and busts. Monetary policy is a far more effective, though still imperfect weapon in that fight. But it is reasonable to expect regulators to act as speed bumps when the traffic is accelerating too rapidly. They did not perform that function. Most regulators themselves now accept that there was too little capital in the banking system, and especially that capital requirements in the trading books of the investment banks were far too light. The regime assumed the effectiveness of hedging strategies which proved of little value as previous price relationships broke down. It also assumed continuous liquidity, an assumption which proved dramatically false in 2008.Regulators, like the banks themselves, failed to identify the damage that could be done by a collapse of confidence in counterparties, in circumstances where many trades were highly complex over-the-counter deals which were extremely difficult to price, even in normal market conditions. Furthermore, capital calculations were essentially backward-looking, and arguably procyclical. If asset markets have been rising consistently for a long period, banks' losses on property lending will have been low. So the models, based on the extrapolation of past experience into the future, might well indicate that a bank needed very low reserves at the top of a boom, just at the time when large reserves would be most needed.

The intellectual framework within which regulators (and central banks) worked was one in which markets were assumed to be efficient, and prices to be set by a complex interplay of willing buyers and willing sellers. If new instruments created ever more sophisticated methods of trading risk, making markets more complete, they must be beneficial. It was not the regulator's place to substitute his judgement for that of the market, full of highly paid, (and perhaps highly sexed) bankers armed with PhDs from top universities around the globe.

For all these reasons, regulators were slow to identify dangerous trends and to warn against them. That failing was shared with the central bankers and the boards of the international financial institutions. The IMF was particularly weak in that respect, proclaiming, until just before the crisis hit, its belief that risk transfer innovations had made the financial system more robust, and bank failures less likely (3). While individual institutions warned against specific trends and imbalances – the Bank for International Settlements can probably claim the best record in the pre-crisis years – no entity pulled the pieces of the jigsaw together. Different committees and groups touched different parts of the pachyderm's anatomy, but none suspected that there was a full-size elephant in the room.

Structural Weaknesses

These weaknesses do point to some structural issues, though I would argue that they were less important than the confidence and mindset points. Two of them, however, stand out.

If we look at the pre-crisis global regulatory architecture we see an incontinent number of committees, boards and forums – a spider's web of interlocking relationships – with the Financial Stability Forum sitting awkwardly in the centre – but with no hierarchical relationships between them (4). So the FSF included the heads of the international standard-setters, the Basel Committee, IOSCO and so on, but had no authority to tell them what to do or when to do it. Each of them operated to their own leisurely timetable, dictated by the enthusiasms of the membership, and their willingness to devote to time and energy to international issues, rather than to domestic fire fighting.

So the Basel Committee spent over a decade producing the new capital rules known as Basel 2, even though serious flaws in the original Capital Accord had been identified. There was no superior body in authority over the Committee to enquire just when the new egg would be laid. So iteration after iteration, draft after draft were produced, of ever greater complexity, but no-one asked the big question of whether there was enough capital in the banking system overall. The exercise began with the lazy assumption that the quantum of capital was "about right" and the issue was simply how best to allocate it. By the time the crisis hit there was broad agreement on Basel 2, but the US had not resolved to implement it, and various versions were in existence around the globe, many of them relying excessively on banks' own internal models to determine risk.

The second obvious flaw in the global architecture was the lack of representation of the growing financial powers of the developing world in the principal standard-setting organisations. I drew attention to this problem in a valedictory lecture on leaving the UK's Financial Services Authority in 2003 (5). This was a version of the UN Security Council problem. The world is no longer what it was in 1945, yet the permanent members of the Council, each with a veto, remain as they were.

The membership of the financial bodies was mainly G7-based, at a time when the centre of the world's economic gravity was shifting rapidly to the East. The criticism was less valid in relation to organisations like IOSCO, who operated on broadly democratic principles for some of their committees. But even there the key Technical Committee was always chaired by a developed market regulator, and its membership was G7-dominated. The Basel Committee provided perhaps the most egregious example. In 2006 10 of its 13 members were from Europe, and the European Commission and the European Central Bank also attended. The most recent addition to the Committee, 5 years earlier, had been Spain.

I would not argue that this lack of representativeness can be seen as a prime cause of the regulatory failures described above, but there were signs that important countries, notably China, were becoming reluctant to be "price-takers", simply accepting standards set by others, on which they had not been consulted. That created the risk of uneven application of global standards, and we saw some evidence of that risk emerging.

There were structural flaws elsewhere, too, at regional and national level. It was already clear that the EU was living uncomfortably in a half-way house, between national and pan-European regulation. The Lamfalusssy Report (1) had identified inconsistencies in EU regulation, and proposed a network of Committees in an attempt to create greater coherence. But member states were reluctant to endow these committees with powers to impose common standards, so most formal authority still rested with national regulators.

The crisis starkly revealed the flaws in this approach. The Icelandic bank case was the most severe test. (Iceland is not, yet, a full member of the European Union, but it is part of the European Economic Area (EEA), which is the relevant jurisdiction for regulatory directives). According to EU law a bank authorised in any country of the EEA is entitled to take deposits in all other countries, without needing authorisation from the host regulator. When they began to run short of funds to fuel their aggressive expansion, Icelandic banks chose to seek retail deposits in the UK and the Netherlands, by the simple expedient of offering deposit rates slightly higher than those of the competition. When the crisis hit, and the three big Icelandic banks were revealed to be seriously overextended, they were unable to refund those deposits, and the Icelandic central bank was too small to be able to help. So British and Dutch taxpayers were the only effective source of compensation for depositors in a bank over which their own regulators had had no authority. They paid up, to the tune of several billion pounds. They were not legally obliged to do so, but politicians in both countries concluded that it would be unacceptable to allow depositors in Icelandic banks to suffer, especially at a time when governments in Europe were promising that no domestic depositors would lose money in the crisis.

This was clearly an unintended outcome. But different Europeans drew two opposing conclusions. Those inclined to favour greater European integration used the experience to argue that the system of mutual recognition on which the single financial market was originally constructed, was no longer viable, and that a system of pan-European regulation was clearly needed. Sceptics took the opposite view, maintaining that the real lesson was the need for host regulators to have the power to reject incomers from elsewhere in the EEA if they had doubts about their viability. That would begin to dismantle the Single Financial Market. Hedging his bets somewhat, the Chairman of the FSA noted that the episode clearly showed that we needed either “more Europe or less Europe (6)”, and that the status quo was not tenable.

The crisis also revealed structural problems or regulatory gaps in individual countries. The US was an obvious case in point. Critics have pointed in particular to the lack of regulation of the mortgage market, to the existence of a multiplicity of banking regulators creating scope for regulatory arbitrage, to dysfunctional disputes between the two securities regulators the SEC and CFTC, and to the lack of a body charged with oversight of systemic risk. (The last point has been made in the EU and in other individual countries). The consequent domestic US reforms are outside the scope of this essay. The Dodd-Frank Act made headway in some of these areas, but the safest conclusion for now may be that it is too early to say how effective those changes will prove to be.

In the UK there was an early challenge to the regulators in the form of the failure of Northern Rock, an almost exclusively domestic mortgage bank, formerly a mutual Building Society. The authorities’ initial response was hesitant and unsure, and for the first time for 150 years there was a fully fledged run on the bank, with queues of depositors outside branches trying to withdraw their funds. It was widely argued, notably by the Conservative opposition, that the fault lay in the reforms carried out by the Labour government in the late 1990s, and especially in the removal of banking supervision from the Bank of England. There was a convenient political dimension to this argument, of course, as it allowed the Opposition to pin blame for the crisis on Gordon Brown, but it certainly did seem that the so-called Tripartite system, involving the Treasury, the Bank of England and the FSA, had worked poorly, with different views taken by the different participants, leading to a sub-optimal outcome.

This catalogue of regulatory failure is depressing. It could be lengthened. There are other European countries, notably Germany, where tough questions have been asked about the oversight of regional banks, in particular. The Dutch Central Bank has been widely criticised for presiding over the almost total collapse of its banking system. But there is not scope here to review many individual countries in detail. So I will limit myself to asking whether the reforms agreed so far, at global level, in the EU as a whole, and in the UK have responded appropriately to the problems identified.

Post-crisis Reforms

Global

If we begin with the structural changes, the first and most rapidly agreed change was the switch from the G7 to the G20 as the basis for membership of the key financial oversight bodies. It was so obvious that an adequate response to the crisis needed cooperation from the large surplus countries (not least because they were engaged in recapitalisation of Western banks) that the convening of a G20 summit by President Obama in

December 2008 was accepted by all countries without demur. Changes in the membership of the FSF and the Basel Committee quickly followed, after the April 2009 London Summit.

There are those who argue that even this broader membership is not adequate. (We should note that in practice the G20 is not quite what it seems: a number of countries not formally members, like Spain, are routinely included, and regulators from the important financial centres of Hong Kong and Singapore, again not part of the G20 network, are also invited). In a report for the UN Secretary General (7), Joe Stiglitz and others have advocated a system built on more comprehensively global lines. But it seems unlikely that further expansion will be agreed in the near future.

Will this broader membership contribute to making the financial system safer? It is hard to say. We do not know what the important new countries want to achieve. So far, the signs are that China sees advantage in implementing tougher capital standards, and is committed to their enforcement globally. But some of the other benefits of expansion seem further away than ever. The Chinese have been determined to exclude from the agenda discussion of currency misalignments and global imbalances, which continue to be a feature of the international environment. So, for now, I would view the expansion of membership as an inevitable and overdue change, reflecting the new economic realities, but not one which will necessarily enhance the quality of regulation, or promote the co-ordination of macroeconomic policies which would help to avoid a recurrence of the catastrophic events of 2007–9.

Also at the London summit the G20 agreed to strengthen the centre of the system, by renaming the Financial Stability Forum the Financial Stability Board. What's in a name?, one might ask. Not necessarily a great deal, but the G20 finance ministers have looked to the FSB, as it now is, to present progress reports to successive summits on the reform agenda. That gives the Board some purchase on the standard-setters and others, and it is reasonable to believe that it has had some effect on the working practices of the Basel Committee, which produced a new capital regime, Basel 3, in little more than 10% of the time it took to gestate Basel 2.

But the FSB remains an informal body. There is no treaty basis for its existence. Its chair is a part-timer. For four years Mario Draghi was simultaneously Governor of the Bank of Italy; the current Chair, Mark Carney, is the Governor of the Bank of Canada. The secretariat is small, and depends heavily on the BIS. Its capacity for independent action is strictly limited. The commitment of some of its members, notably the US, is doubtful.

The Council on Global Financial Regulation, a group of former regulators, central bankers and academics (of which I am a member), has prepared a detailed critique (8) of the FSB in its new form, and advocated significant reforms to strengthen its position. While it is possible to argue that its output has been disappointing so far, the CGFR's view is that that is more the consequence of its uncertain status, than an inevitable consequence of its structure. It remains the only body which includes representatives of all the agencies needed to co-ordinate effective action at global level. Finance Ministries, regulators and central bankers are all at the (rather large) table, together with the International Financial Institutions and the key standard-setters. The personnel is right, therefore, but the institutional backing is still lacking.

The G20 summit in France in November 2011 broadly accepted this diagnosis and asked the FSB to review the options for giving itself a legal entity status, and strengthening its own staff. That work remains under way, in the summer of 2012, and appears to be moving slowly. One cannot therefore give this area of reform more than a modest grade so far.

We can be somewhat more optimistic about the changes under way as a result of the Basel Committee's supercharged work on Basel 3. They have produced a new framework, with far tougher requirements (9). Banks will in future be required to hold significantly larger capital reserves, and a larger proportion of those reserves must be in the form of tangible common equity. (A further reasonable criticism of Basel 2 was that it paid too little attention to the quality, as opposed to the quantity of capital. Banks were allowed to count a variety of hybrid convertible instruments towards their reserves, instruments which did not deliver usable capital in the crisis). The Committee have also proposed a new resolution regime, which aims to allow banks to be wound up without causing severe disruption to the wider economy. Systemic institutions must prepare "living wills", or "funeral plans" as they are sometimes known. Most have now done so, and the exercise has proved valuably. Contingency plans to sell off assets and wind down balance sheets have been prepared, and holdings of capital and liquidity have been enhanced. But, as the Lehman case demonstrated, there remain many obstacles to a rational cross-border insolvency regime, and progress in that area has been limited.

The framework also attempts to respond to the procyclicality problem by introducing a so-called "countercyclical buffer", an additional reserve which might be varied depending on regulators' view of the state of the business cycle, or of potential misalignments of asset prices. And on top of that, for the largest banks, there is a kind of "too big to fail" supplement imposed on institutions deemed to be systemically significant. The FSB was given the task of identifying a class of Global SIFIs (Systemically Important Financial Institutions) and complementary lists of regional and national SIFIs are also under preparation.

These reforms will undoubtedly make the banking system safer. But the behaviour of bank shares seems also to be telling us that they will markedly reduce its return on equity. That may be appropriate, as banks will in future look more like regulated utilities, with tight controls on capital and indeed on dividends. But what of the impact on the cost of bank borrowing, and thereby on investment in the economy more generally, and on economic growth and job creation? On that crucial question there is no consensus whatsoever. The Basel Committee released a study (10) which argued that the impact would be very modest indeed, and that growth would be less than half a per cent lower over five years. The OECD took a different view (11), estimating the impact at about twice that size. But economists at the Institute of International Finance, the trade association for the biggest international banks, argue that growth will be fully 3% lower over 5 years (12). If they are right (and you may think they would say that, wouldn't they) this would prove to be a very costly reform indeed. So the best we can say is that regulators have plugged the obvious gap, and done so quickly. But we do not know whether they have, to mix a metaphor uncomfortably, washed the baby of growth out at the same time. Only time will answer that question.

On the countercyclical question, while there is agreement on an additional capital buffer, we do not know how decisions on its implementation will be made. How do we assess when markets are out of line, or when credit growth is too rapid? As we saw, it was the failure to react pre-emptively to credit expansion which contributed as much as anything else to the inflation of the bubble which burst so dramatically in 2007. And who assesses which response is most appropriate? In principle, one can respond to excess credit growth by raising interest rates, or by lifting capital requirements by expanding the countercyclical buffer. But the first response is seen as the province of monetary policy-makers, while the second is a matter for regulators. These may seem arcane arguments at a time when our concerns are in the opposite direction, and when the Federal Reserve have promised to maintain short-term interest rates at close to zero for the foreseeable future, but one day the problem will arise again.

In principle the FSB could take a view, but so far members have been reluctant to see it stray into that territory. In Europe the European Systemic Risk Board, chaired by the President of the ECB, could do so, but interest rates remain the jealously guarded province of the ECB's Governing Council. In the US the new Financial Stability Oversight Council might opine, but once again control over interest rates lies elsewhere, with the Federal Open Market Committee. In the UK there is a new Financial Policy Committee, sitting alongside the Bank of England's Monetary Policy Committee, but with very different membership and procedures. It is hard to escape the conclusion that these structural reforms have not resolved the problem. We will, as before, depend on the judgment of the individuals in positions of influence- many of them the same people as before.

European

In Europe, too, it is difficult to be optimistic about the response so far to the "more or less Europe" question. Following in the footsteps of Alexandre Lamfalussy before him, Jacques de Larosiere was asked to address the problem. His report was characteristically thoughtful and careful (13). He drew attention to the problems which had merged and described the policy dilemma well. His predisposition, unsurprisingly, was to move in the direction of more Europe. To have recommended otherwise would have been to reverse the trend of fifty years of European integration. But he fell well short of advocating a single European regulator, or even an optional federal regime for pan-European institutions, on the American model. Some EU firms, notably Deutschebank, have been arguing for such an option for some years.

The essence of the change is that the three Lamfalussy committees have become "Authorities", so we now have a European Banking Authority (EBA), a European Securities and Markets Authority (ESMA) and a European Insurance and Occupational Pensions Authority (EIOPA). The three pillar structure no longer represents the way in which a majority of member states organise their domestic regulation, but perhaps that can be set aside. More importantly, they are located in three different cities: London, Paris and Frankfurt respectively, reflecting a purely political deal. That does not facilitate cross-authority co-ordination. And they are also barely "Authorities" in the normal sense of the term. Their powers are quite limited. ESMA has direct authority over credit rating agencies, but with that exception they operate through national authorities. They are charged with preparing, over time, a single European rule book, and they have the ability to arbitrate in the event of disagreements between states on the implementation of directives. The former could be a significant lever, if agreement can be reached. But we remain a long way from a federal system of regulation, and it is not even clear that the new arrangements would prevent a recurrence of the Icelandic bank problem.

Of course we must acknowledge that this dimension of European integration has taken a back seat during 2011 and 2012, as bigger issues relating to the future of the Eurozone have come to the fore. In that more existential debate the bigger issue of the nature of the European project is once again under consideration. Will the Eurozone move towards a fiscal union, as many believed would be the inevitable consequence of the single currency? Will there be a Eurozone finance ministry, as Jean-Claude Trichet, when President of the European Central Bank has proposed? Would that ministry issue Eurozone bonds, guaranteed collectively by all governments? Commentators increasingly see them as necessary to maintain the integrity of the euro itself, and the structure of financial regulation has also come back onto the political agenda.

The European Commission has proposed a "banking union", involving direct regulatory powers for the EBA, a mutually guaranteed deposit protection scheme across the Eurozone, and a lender of last resort for pan-European banks, which might be the ECB or the European Stability Mechanism. The ECB has put itself forward as the Euro's banking supervisor, correcting what it has seen as a weakness in the powers it was granted under the Maastricht Treaty. Both of these options will be difficult to establish, in practice, though the ECB model may be easier in legislative terms. The European summit at the end of June 2012 opted to make the ECB the centre of a new regulatory system.

My forecast would be that a genuinely pan-European system of regulation will (and should) eventually be set up, at least for major cross-border firms based in the Eurozone. That will pose an interesting challenge for the UK, which has set its face against any further transfer of powers to European regulators, and perhaps other noneurozone members. In the meantime, Europe is lodged in yet another halfway house, albeit one slightly closer to the federal model. The "more or less" question has been answered in the form favoured by polite houseguests offered a second helping of a dish they scarcely enjoyed: “a little bit more, thank-you, but not too much”.

UK

The UK’s reforms are still work in progress. In opposition the Conservative Party committed itself to “abolishing” the FSA. It is in the course of doing so, though in strict legal terms the FSA will remain in existence as the FCA- the Financial Conduct Authority. More than a letter has changed, however. In what is described as a “new approach to financial regulation” (14), the prudential functions of the FSA have been carved out into a new Prudential Regulatory Authority, which will be, as it were, a wholly-owned subsidiary of the Bank of England. It will retain a degree of separation, in part because over time the supervisors have come to be rather better paid than the remaining central bankers. They are more vulnerable to competing offers from financial firms. That makes full integration with the Bank of England problematic, but its Chief Executive will be a Deputy Governor of the Bank, and the governor will chair its Board.

In addition, as described above, there is a new Financial Policy Committee (FPC), located in the Bank of England, and also chaired by the Governor, whose role is to “contribute to the Bank’s financial stability objective by identifying, monitoring and taking action to remove or reduce systemic risks with a view to protecting and enhancing the resilience of the UK financial system” (9). The FPC can give instructions to the PRA and in theory to the FCA. It is not clear whether it can also offer advice to the MPC, or indeed vice versa.

It is too early to assess the impact of these reforms. City firms would give a very mixed verdict at present. The transactions costs are high and the FSA is losing staff rapidly. That is an inevitable consequence of any institutional change. It may be a price worth paying. But the political nature of the original impulse to reform gives grounds for concern, and more mature analysis of the crisis does not suggest that the Tripartite system itself was a major contributory factor. The absence of clear political leadership seems to have been far more important, and the government’s reluctance to acknowledge the need to take over Northern Rock, or to provide liquidity support for a private sector solution, made the handling of the problem far more messy than it need have been. It also appears, though there has been no review which allows us to know whether this conclusion is correct, that there were strongly held differences of view on the handling within the Tripartite Committee. In a valedictory interview as he left office in June 2012 the former Chief Executive of the FSA, Hector Sants, said as much to the BBC (15).

As its own Internal Audit report showed (16), the FSA’s supervision of Northern Rock had been deficient, but that looks to have been attributable to errors of judgment, not to structural flaws. Also, the reform has not in practice clarified accountabilities or responsibilities. There are now four, rather than three entities likely to be involved in crisis management: The Treasury, The Bank of England, the PRA and the FCA, with the FPC sitting between them. As before, and as is the case everywhere, the effectiveness of the arrangements will depend crucially on the skills and wisdom of the participants, rather than on the particular structure within which they work.

Conclusions

In the four years since the crisis erupted in the summer of 2007, much has been done to correct the regulatory flaws it revealed. For a time, it seemed that the political obstacles which had bedevilled earlier attempts at reform, globally regionally or nationally, would be blown away. So there was talk of a global body with genuine power to enforce regulations – a World Financial Authority as has been advocated in the past (17). In the US it seemed possible that a merger between the SEC and the CFTC could be engineered. In Europe a pan-European regulator seemed at one stage to be within reach.

Now, although the crisis is far from over, and markets remain extremely nervous, these grander ideas have disappeared from the agenda, and the art of the possible has again become the key skill for reformers. Tentative moves to strengthen the central nervous system of global finance have been made, but they fall well short of a revolution. In the US there have been only modest structural changes, but a barely-digestible wave of new legislation. In the EU we have the 'form' of European-wide regulation in the three new authorities, but not the substance. Another more fundamental structural reform is surely in prospect. In the UK we have once again shuffled the regulatory pack, and put the Bank on top of the pile, from which it had been dislodged a decade or so earlier. What goes around, comes around.

So has a good crisis been wasted? The wise commentator would opine sagely that it is too soon to tell. But, overall, it is hard to escape the conclusion that there has so far been less in the way of significant reform than meets the eye.

See Also

- ▶ Banking Crises
- ▶ Credit Crunch Chronology: April 2007–September 2009
- ▶ Euro Zone Crisis 2010
- ▶ Fall of AIG
- ▶ Run on Northern Rock

Bibliography

A global regulatory framework for more resilient banks and banking systems. Basel Committee on Banking Supervision. June 2011. www.bis.org

A new approach to financial regulation: Building a stronger system. HM Treasury. February 2011. www.hm-treasury.gov.uk

An assessment of the longer-term impact of stronger capital and liquidity requirements. BCBS and FSB. August 2010. www.bis.org

Davies, Howard. 2001. *The financial crisis: Who's to blame?* Polity Press.

Davies, Howard. Is the global regulatory system fit for purpose in the 21st century? Monetary Authority of Singapore Annual Lecture. 20 May 2003. www.fsa.gov.uk

Davies, Howard and David Green. 2010. *Global financial regulation: The essential guide*. 2nd ed. Polity Press.

Eatwell, John and Lance Taylor. 2000. *Global finance at risk*. Polity Press.

Global Financial Stability Report. April 2006. International Monetary Fund.

Hector Sants' interview with Robert Peston. 13 June 2012. www.bbc.co.uk

Practical measures for enhancing international financial regulatory co-ordination. Council on Global Financial Regulation. 14 April 2011. www.c-gfr.org

Remarks at Launch of Turner Review. 18 March 2009. www.fsa.gov.uk

Report of the Committee of Experts of the President of the United Nations General Assembly on the Reform of the International Monetary and Financial System. New York. 21 September 2009. www.un.org

Report of the Committee of Wise Men on the Regulation of Europe's Securities Markets. Brussels. 15 February 2001. www.ec.europa.eu

Report of the High-Level Group on Financial Supervision in the EU. Chaired by Jacques de Larosiere. Brussels. 25 February 2009. www.ec.europa.eu

Slovik, P., and B. Cournede. The macroeconomic impact of Basel 3. OECD Economic Department Working Paper No. 844. 14 February 2011. www.oecd-ilibrary.org

Suttle, Philip. The macroeconomic implications of Basel 3. IIF Spring Membership meeting. New Delhi. 3 March 2011. www.iif.org

The FSA's internal audit review of its supervision of Northern Rock and the FSA's management response. 26 March 2008. www.fsa.gov.uk

Reid, Margaret Gilpin (1896–1991)

Mary Jean Bowman

Keywords

Friedman, M.; Household production; Life stages; Permanent income; Permanent income hypothesis; Reid, M. G.; Transitory income

JEL Classifications

B31

Margaret Reid, a leading scholar in analysis of the economics of consumer behaviour, was made a distinguished fellow of the American Economic Association in 1980. She was Professor of Economics at Iowa State College (1930–43), the University of Illinois (1948–51), and the University of Chicago (1951–61).

R

A realistic theorist, Reid always looked behind data to processes that generate structural relationships. Her 1934 book on household production anticipated by three decades analyses built on the allocation of time, and she was the first (1947) to use wage-equivalent time measures of household work.

Already in Iowa she had questioned attempts to improve resource allocations by farm women that disregarded the nature of income effects. She went on to criticize assessments of the war-time cost-of-living index that neglected effects of changing incomes on the quality of goods traded, and she became the 'directing' member of the technical committee responsible for a report to the President's Commission on the Cost of Living (1945). Later on she challenged conventional treatments of income elasticities of consumption in general and of housing expenditures in particular (1952; 1962).

The concepts of 'permanent' and 'transitory' income were early a part of Reid's thinking (1952; 1953). Friedman drew on Reid in his 1957 application of the permanent income hypothesis to short-term shifts in consumption and saving, and Modigliani built on her work in his treatment of 'life stages' (Modigliani and Ando 1960 and subsequently). In Reid's hands the concepts of 'permanent' and 'transitory' income evolved subtly and progressively in multiple facets of the analysis of consumer behaviour. After her retirement she probed interactions between health and income both over life cycles and across cohorts.

Selected Works

1934. *Economics of household production*. New York: Wiley.
1943. *Food for people*. New York: Wiley.
1945. (With associates.) Appendix IV. Prices and the cost of living in wartime – an appraisal of the Bureau of Labor Statistics Index of the Cost of Living in 1941–44. In *Report of the President's Committee on the cost of living*. Washington, DC: Government Printing Office.
1947. The economic contribution of homemakers. In *Women's opportunities and responsibilities*, ed. L.M. Young. Philadelphia: Annals of the American Academy of Political and Social Science.
1952. Effect of income concept upon expenditure curves of farm families. In *Conference on research in income and wealth, studies*, vol. 15, part IV. Philadelphia: NBER.
1953. Savings by family units in consecutive periods. In *Savings in the modern economy*, ed. W.W. Heller, F.M. Boddy and C.L. Nelson. Minneapolis: University of Minnesota Press.
1955. Food, liquor and tobacco. In *America's needs and resources: A new survey*, ed. J.F. Dewhurst. New York: Twentieth Century Fund.
1960. Comments on J. Crockett and I. Friend. A complete set of consumer demand relationships. In *Proceedings of the conference on consumption and savings*, vol. I, ed. I. Friend and R. Jones. Philadelphia: Wharton School of Finance and Commerce, University of Pennsylvania.
1962. *Housing and income*. Chicago: University of Chicago Press.

Bibliography

Friedman, M. 1957. *A theory of the consumption function*. Princeton: Princeton University Press.

Modigliani, F., and A. Ando. 1960. The 'permanent income' and the 'life cycle' hypothesis of saving behavior: Comparisons and tests. In *Proceedings of the conference on consumption and savings*, ed. I. Friend and R. Jones. Philadelphia: Wharton School of Finance and Commerce, University of Pennsylvania.

Relative Income Hypothesis

Richard F. Kosobud

The relative income hypothesis as expressed by its foremost exponent was an effort to reconcile conflicting evidence revealed by consumption functions fitted to long and short-period time series, and budget data; to bring social psychology into consumer theory; and to restore virtue to the act of saving (Duesenberry 1962). While proposed as a critique of the Keynesian consumption function, neither its formulation nor implications weakened the concept of deficient aggregate demand nor the grounds for stabilization policy. It was consequently well within the framework of the vast research programme stimulated by the *General Theory*.

Written twenty-five years after that book, Duesenberry's work gained widespread recognition. Twenty-five years later, it has fallen out of favour replaced by two new streams of research into saving, and consumption, behaviour: one based on a conception of the rational consumer dealing with the problem of consumption now or later by maximizing an intertemporal utility function in perfect or instantaneously cleared markets; and the other based on the consumer constrained on occasion by realized income generated in markets that do not clear – the familiar Keynesian consumption function now being placed on a rigorous micro-footing.

It is too soon to write an obituary for the relative income hypothesis and this survey is not so intended; the hypothesis provides a suggestive account of aspects of consumer behaviour and seems capable of further development and research stimulation. Limitations of the theory and difficulties in obtaining corroborative evidence, however, may hinder its return to prominence.

The fundamental psychological law to which Keynes appealed to give shape to his consumption function gave rise 'as a rule, to a greater proportion of income being saved as real income increases' (Keynes 1935, p. 97). Increased saving that did not get absorbed by increased investment posed a threat as it led to inadequate effective demand, and output and employment below potential. Saving might be a superior good in this view, but hardly a superior social virtue. That the ratio of savings to income rose with income was consistent with budget data available at that time; but, the long-run time series data subsequently prepared by Kuznets did not reveal a change in that ratio (Kuznets 1952, pp. 507–26). Contradictory evidence further accumulated in the postwar period as functions fitted to short and long-period data revealed different numerical results. Research hounds bayed after these hares.

In an innovative study utilizing four budget studies, Brady and Friedman compared family savings–income ratios to both absolute and relative income, the latter being the ratio of family income to the group mean. An improved fit was obtained in the latter case (Brady and Friedman 1947, p. 261). Modigliani studied the procyclical movement of the savings rate and related it to the ratio of the aggregate current to highest previous income (Modigliani 1949). The most comprehensive effort to devise an alternative hypothesis to encompass this evidence was made by the Duesenberry who wrote: 'that for any given relative income distribution, the percentage of income saved by a family will tend to be a unique, invariant, and increasing function of its percentile position in the income distribution' (Duesenberry 1949, p. 3). On this foundation an aggregate saving or consumption function was erected that reconciled discrepancies in functions fitted to time series and budget data. An additional, and distinct, assertion that the savings, or consumption ratio, depends in the short run on the deflated, per capital ratio of current to prior peak income was required to reconcile functions fitted to long-run and cyclical time series data.

Why should the consumer unit's position in the income distribution matter? Our attention is first directed to the complex and differentiated package of services most commodities yield which consumers in our status-conscious society convert into a commodity quality hierarchy. The frequency of contact which consumers have with superior goods in the hierarchy is closely related to the comparisons they make between their consumption and that of others. This is the 'demonstration effect'.

Equally important is the drive, instilled in every individual's mind by the socialization process, toward a higher standard of living including the consumption of more and better quality goods and services. 'In view of these considerations it seems quite possible that after some minimum income is reached, the frequency and strength of impulses to increase expenditures for one individual depend entirely on the ratio of his expenditures to the expenditures of those with whom he associates' (Duesenberry 1949, p. 32). Together with the demonstration effect, this drive toward emulation or the making of individious comparisons for which opportunities abound in our society, explain what makes the consumer tick.

In the light of this social psychological argument, the traditional assumption that the utility function of the individual is independent of the functions of others will not do. Duesenberry introduces an interdependent utility function whose arguments are divided by a weighted average of the consumption of other relevant individuals, the weights being those meaningful to the consumer. This system of individual utility functions must be solved simultaneously to obtain demand schedules which are made to depend on current income, current assets, future income, expected future interest rates, and most important, the current consumption of other people. The apparent contradiction between functions fitted to long-run and budget data may be reconciled by the following argument. If all incomes, now and expected, are increased by a factor k, the distribution of income being unchanged, the ratios in the consumption function both in range and domain are unchanged. In the new equilibrium the savings–income ratio is unchanged as everyone has just managed to keep up with the Jones's expenditures. The stagnationist's fear of an increasing savings–income ratio over the long run may be put to rest.

The idea of introducing interdependence into the utility function has not yet commanded general acceptance among economists, suggestive as the arguments are. It has proved difficult so far to deduce many implications from the function that could be subjected to testing. This compares unfavourably with the life cycle–permanent income hypothesis. The consumer unit is looking sidewards at the consumption of peers rather than toward the future, as in more recent research. A rigorous derivation of the implications of the budget constraint has yet to be carried out to reveal precisely how the consumer unit is to finance the invidious, apparently endless, and typically unsuccessful pursuit of material happiness.

There remain short run fluctuations in saving to be explained. A new element was introduced: it was a fundamental psychological postulate 'that it is harder for a family to reduce its expenditure from a high level than for a family to refrain from making high expenditures in the first place' (Duesenberry 1949, p. 85). This postulate was given the following shape:

$$s/y = a(y/yo) + b$$

where the per capita real variables, in order, are current savings, current disposable income, and highest disposable income ever attained. This yields a changing but calculable multiplier in simple models, and an average greater than the marginal propensity to consume, solidly within the Keynesian tradition. The habitual standard of living appealed to in this saving function was further refined by a proposal to include the previous highest consumption rather than income (Brown 1952). Lagged consumption has proved highly significant in a statistical sense in a variety of econometric studies; but, rather than habit it has been interpreted as embodying all the information available to the consumer at that time and representing, in a rational expectations point of view, the best forecast of the next period's consumption (Hall 1978).

The pace and sophistication of quantitative and econometric studies of the relative income hypothesis and related ideas have been such that only a few highlights can be mentioned. The evidence is not conclusive, but has not been kind to the present formulation of the relative income hypothesis. Duesenberry found interesting evidence in Negro and white communities that apparent discrepancies in the saving behaviour by race can be reconciled by plotting the per cent saved against the family's position in the income distribution, instead of absolute income. Differences in group savings–income ratios, however, given comparable income distributions, are better explained in terms of differences in permanent income or life resources, or, more precisely, differences in the ratio of transitory or unexpected income to measured income.

The cyclical saving function when tested against limited data appeared to perform better than a simple function based solely on current income plus trend (Duesenberry 1949, p. 51 and 82). Tobin compared the relative and absolute income hypotheses using both budget and time series data, however, and found that by modifying the latter with the introduction of a wealth or financial resource measures other than income, the latter did better or at least equalled the quantitative performance of the former (Tobin 1951, pp. 135–56).

M. Friedman and F. Modigliani, both with associates, opened up a new horizon in consumption research by applying the pure theory of consumer behaviour in its traditional form (maximization of a utility function independent of the functions of others) to the problems of devising a long-run consumption plan. The former addressed the problem of the infinitely lived consumer who could be expected to consume a large proportion of permanent income, defined as the return on human and non-human wealth (Friedman 1957). The latter addressed the problem of the finitely lived consumer who could be expected to consume at a relatively constant rate the life resources available to him or her (Modigliani 1986). The implications of the permanent income and life cycle hypotheses have attracted the major share of research attention and their forward looking character has facilitated the application of rational expectations methodology to their development with a further accumulation of supporting evidence (Hall 1978).

The relative income hypothesis remains as a contribution to the development of knowledge of consumer behaviour so important for the advancement of macroeconomics. Whether the limitations in theoretical development – relatively few testable implications; difficulties in exploring the meaning of the budget constraint; the incorporation of expectations in an optimal way – can be overcome, constitute open questions.

See Also

▶ Consumption Function
▶ Life Cycle Hypothesis

Bibliography

Brady, D.S., and R.D. Friedman. 1947. Savings and income distribution. In *Studies in income and wealth*, No. 9. New York: National Bureau of Economic Research.

Brown, T.M. 1952. Habit persistence and lags in consumer behaviour. *Econometrica* 20: 355–371.

Duesenberry, J.S. 1949. *Income, savings and the theory of consumer behavior*. Cambridge, MA: Harvard University Press.

Friedman, M. 1957. *A theory of the consumption function*. Princeton: Princeton University Press.

Hall, R.E. 1978. Stochastic implications of the life cycle–permanent income hypothesis: Theory and evidence. *Journal of Political Economy* 86(6): 971–987.

Keynes, J.M. 1936. *The general theory of employment, interest and money*. London: Macmillan.

Kuznets, S. 1952. Proportion of capital formation to national product. *American Economic Review: Papers and Proceedings* 42: 507–526.

Modigliani, F. 1949. Fluctuations in the saving-income ratio: A problem in economic forecasting. In *Studies in income and wealth*, No. 11. New York: National Bureau of Economic Research.

Modigliani, F. 1986. Life cycle, individual thrift, and the wealth of nations. *American Economic Review* 76(3): 297–313.

Tobin, J. 1951. Relative income, absolute income, and savings. In *Money, trade, and economic growth, in honor of John Henry Williams*, ed. H.L. Waitzman. New York: Macmillan.

Relativity, Principle of, in Political Economy

J. N. Keynes

The principle that the economic doctrines true for any given epoch are relative to the particular circumstances of that epoch, and cannot be regarded as permanent or true for all times, is an essential element in the teaching of the historical school of economists. The idea of the relativity of economic doctrines follows easily from the conception of economic life as exhibiting continuous organic growth and development, and this conception is itself the natural outcome of historical study.

Richard Jones and Friedrich List are to be regarded as important forerunners of the historical movement rather than as themselves typical representatives of the movement itself. What is most characteristic, however, in their teaching is the insistence upon relativity in two particular spheres; and a brief reference to their views will serve to illustrate what is meant by the principle of relativity in general. Jones specially insisted on the limited applicability of the Ricardian theory of rent as regards both place and time. A theory based on the assumptions of individual ownership and freedom of competition could not, he pointed out, apply to oriental states of society in which joint ownership is the rule and rents are regulated by custom, nor even to those instances nearer home in which land is held on a customary tenure, as in the métayer system. Similarly, as regards limitation in time, he showed that the Ricardian law could not hold good in a condition of affairs such as existed in medieval economy, when land was to a great extent held in common, and the relations between the owners and the tillers of the soil were not controlled by free competition. Turning to List, we find that his defence of protective duties is based on the recognition of relativity in the conditions of economic productivity in a community. The foundation of the argument is the position that all civilized communities of the temperate zone pass through successive economic stages, of which the last three are the stage of agriculture pure and simple, the stage of agriculture combined with manufactures in a nascent and slowly developing condition, and then the stage in which agriculture, manufactures, and commerce have all reached a high and well-balanced development. In the purely agricultural stage free trade with richer and more developed countries is, in List's view, a necessary condition of advance, and in the last stage of all free trade is also advantageous. On the other hand, the training and development required for passing from the second stage to the third can be acquired only by means of a carefully arranged protective system; and in the second stage such a system is, therefore, necessary for progress. The solution of the problem of protection *versus* free trade is thus regarded as relative to each particular people, and the stage of development which they have reached.

The principle of relativity in the sphere of economics was expressed in a more general form by Wilhelm Roscher. Applying to economic phenomena ideas which writers on jurisprudence had already applied to legal institutions and conceptions, he insisted on the necessity of always taking into consideration the varying character of economic habits and conditions; and, in particular, he pointed out the fallacy of criticizing economic institutions, regardless of a people's history, and the stage of social and industrial development to which they had attained. Karl Knies affirmed still more definitely the relativity of economic doctrines in opposition to what he termed the *absolutism of theory*, that is, the claim – explicitly put forward by some of the older writers, and tacitly assumed by others – to offer something that is true unconditionally and in the same way for all times, lands, and nationalities.

> In opposition to the absolutism of theory, the historical conception of political economy rests [says Knies] upon the fundamental principle that the theory of political economy, in whatever form we find it, is, like economic life itself, a product of historical development; that it grows and develops, in living connection with the whole social organism, out of conditions of time, space, and nationality; that it has the source of its arguments in historical life, and ought to give to its results the character of historical solutions; that the laws of political economy should not be set forth otherwise than as historical explanations and progressive manifestations of the truth; that they represent at each stage the generalizations of truths known up to a certain point of development, and neither in substance nor in form can be declared unconditionally complete; and that the absolutism of theory – even when it gains recognition at a certain period of historical development – itself exists only as the offspring of the time, and marks but a stage in the historical development of political economy (*Die politische Ökonomie von geschichtlichen Standpunkte*, pp. 24, 25). This extract may be regarded as expressing the general view as to the relativity of economic doctrines taken by the historical school distinctively so-called.

The relativity of current political economy is affirmed no less definitely, but from quite a different point of view, by Walter Bagehot, who regards it of importance expressly to limit the science to one particular kind of society, namely, ‘a society of grown-up competitive commerce’, such as we find in the most highly civilized modern communities. Political economy is, in other words, limited to ‘the theory of commerce, as commerce tends more and more to be when capital increases and competition grows’. It will be observed that whilst the object of the historical school is to concentrate attention on economic history and on the study of economic development as opposed to the study of economic relations in a given society, Bagehot’s object is just the reverse. He wishes to concentrate attention on current economic phenomena, and to avoid the distraction that must result from turning aside to the superficially corresponding but yet essentially different phenomena of earlier epochs.

In endeavouring to form an estimate of the importance to be attached to the relativity of economic doctrines we shall do well to have regard (1) to the distinction between economic theorems and economic precepts, and (2) as regards economic theorems, to the distinction between abstract and concrete economics.

Roscher in his affirmation of relativity is thinking mainly of economic institutions and economic policy, and in this sphere the principle of relativity may be laid down with less qualification than when we are dealing with economic laws in the more strictly scientific sense, that is, with statements of uniformities as distinguished from recommendations as to what ought or ought not to be done in practice. It is only by the aid of abstraction that any claim to universality can be made good, and in formulating an economic policy, we cannot profitably carry abstraction very far. In theoretical investigations hypothesis and abstraction are often indispensible; but when we apply our theory with the object of laying down rules of practice, it is desirable to have recourse to hypothesis but sparingly. It is indeed doubtful how far, in the examination and criticism of economic institutions and policies, we can advantageously carry our abstraction even to the stage of neglecting social considerations of the purely non-economic character. Both the social and the economic bearings of a given line of action will, however, vary with the circumstances of different nations and different ages. Hence a given economic policy can in general be recommended only for nations having particular social and economic surroundings, and having reached a certain stage of economic development. It may be possible to formulate as having universal validity certain negative precepts, namely, that certain lines of action cannot under any circumstances be advisable; but on the whole the principle of relativity may be broadly accepted so far as economic precepts are concerned.

Passing from economic precepts to the body of positive doctrine which constitutes more distinctively the science of political economy, attention must be paid to the fact that economic doctrines vary in the degree of abstraction which they involve. Without professing to be able to draw any hard and fast line, we may adopt the suggestion made by W.S. Jevons, and distinguish broadly between two stages of economic doctrine, which may be called the abstract and the concrete stage respectively. Concrete economics is not content with merely hypothetical results, but avowedly takes into account special conditions of time, place, and circumstance; and it follows immediately that the conclusions already arrived at with regard to the relativity of economic precepts apply equally to concrete economic theorems. For the more fully we have regard to special conditions of time, place, and circumstance, the more limited must be the applicability of our results. Many of the circumstances which exert an important influence on economic phenomena vary widely with the legal form of society and with national character and institutions; and even when the same forces are in operation there may be variation to an almost indefinite extent in the relative influence which they exert. The contrasts presented by medieval and modern societies, and by contemporary oriental and European societies, considered in their economic aspects, are such as cannot possibly be overlooked. Many of the chief economic phenomena, such as rent, profit, exchange, have their counterparts on each side of the comparison, but are singularly unlike in many of their characteristics; and over and above this, as societies progress, new economic phenomena, practically novel is character, spring into existence. Consider, for example, modern problems of credit and of international trade; or again, the relations between the modern factory operative and the modern capitalist employer. Less striking contrasts, but contrasts that ought not to be neglected, are observable when we consider different modern communities of the European type in respect of particular economic phenomena, such as the tenure of land, the mobility of labour, and so forth. In every case the extent of the divergence can be ascertained only by direct observation and comparison; and it may be remarked in passing that as regards medieval and modern societies, whilst there was formerly danger of the differences being insufficiently emphasized, there is perhaps at the present time more danger of their being exaggerated. The notion, for example, that during the middle ages the forces of competition were entirely inoperative, is far from being borne out by the facts.

The relativity of concrete economic doctrines having been admitted, a claim for universality may still be put forward so far as the more abstract principles of the science are concerned. These principles do not profess to set forth the full empirical reality. They are admittedly based on hypothesis and abstraction. They require therefore to be constantly qualified and limited, sometimes in one direction, sometimes in another before they can serve for the interpretation and explanation of actual economic phenomena. At the same time, some at least of them are universal in the sense that they pervade all economic reasoning. The law of the variation of utility with quantity of commodity, and the principle that every man so far as he is free to choose will choose the greater apparent good, may be given as examples of fundamental economic principles, which, in the words of Jevons, ‘are so widely true and applicable that they may be considered universally true as regards human nature’. There are many other principles, which, with due modifications, are applicable to economic phenomena under widely different conditions. Take, for instance, the law of substitution in the form that where different methods of production are available for obtaining a given result, the one that can do the work the most cheaply will in time supersede the others, or the doctrine that facilities of transport tend to level values in different places, while facilities of preservation tend to level values at different times. Compare, again, the Ricardian law of rent as ordinarily stated, with the principle of economic rent in its most abstract and generalized form. The Ricardian law, so far as it claims to determine the actual payments made by the cultivators of the soil, is a relative doctrine, that is to say, it is based on assumptions which, as regards both time and place, hold good over a limited range only. The theory of economic rent in its most generalized form, however, merely affirms that where different portions of the total amount of any commodity of uniform quality supplied to the same market are produced at different costs, those portions which are raised at the smaller costs will yield a differential profit; and there is now no similar limitation to its applicability. This principle may even be said to hold good in a socialistic community, for the differential profit does not cease to exist either by being ignored or by being municipalized or nationalized. To take a further illustration, there is a good deal of abstract reasoning in regard to the laws of supply and demand that has a very wide application indeed. These laws work themselves out differently under different conditions, and in particular there are differences in the rapidity with which they operate. Their operation may, however, be detected beneath the surface even in states of society where custom exerts the most rigid sway. In all these cases and others similar the principles involved and the modes of investigation employed have a significance and importance which it would be misleading to call merely relative; and hence as regards the more abstract portions of economic doctrine the principle of relativity cannot be accepted.

The relativity of concrete economic truths, together with the universality of fundamental economic principles, might be illustrated by reference to the writings of the classical English economists. The historical school have rightly taught us that the works of these economists can be fully understood and appreciated only if they are studied in close connection with the economic history of the times when they wrote. Frequently the assumptions on which their reasonings are based have a special relation to the actual circumstances of their time; or, even if this is not the case, the form in which their doctrines are cast, or the emphasis laid upon particular points, will often be found to be specially related to the economic conditions in the midst of which they wrote. It is, however, going much too far to regard their whole teaching as limited throughout by the character of relativity which belongs to some of it. Much of what they wrote will be valuable for all time, not merely because of the light which their doctrines throw on the phenomena of particular periods, but because the principles underlying their best work are not confined in their applicability to any narrow or limited sphere.

In connection with the general subject of relativity in political economy, a word or two may be added with regard to the relativity of economic definitions. Partly on account of the familiarity of much of its subject-matter, and partly for reasons connected with the growth of the science, political economy is for the most part limited in its nomenclature to terms already in common use. In different departments of economic enquiry, however, lines of distinction may need to be drawn at rather different points, and hence it is sometimes difficult to avoid the multiplication of technical terms, unless we are content to use the same terms in slightly varying senses in different connections. Thus, from the point of view of production it may be convenient to give a definition of wealth, not in all respects identical with the definition that is appropriate from the standpoint of distribution. Again, with special reference to its measurement, there may be advantages in defining wealth differently from the cosmopolitan, national, and individual points of view respectively. This procedure, that is to say, the frank adoption of the principle of relativity in framing economic definitions, has considerable weight of authority in its favour; but it is clear that, in so far as it is adopted, special precautions are necessary to avoid confusion. Further, economic definitions may be relative, not only to different points of view or different departments of study, but also to different stages of industrial development. Thus, in relation to the complex conditions of modern trade and industry, such terms as market and money may need different definitions from those that are appropriate in relation to more primitive conditions. Whilst, however, many economic definitions may be allowed to possess a relative or progressive character, this relativity cannot be extended to the ultimate analysis of the fundamental conceptions of the science. If these conceptions assume a some-what different character in different connections, we shall still find something generic or universal in each one of them. Hence in the case of economic definitions as well as in that of economic doctrines, the admission of the principle of relativity must not be absolute or unqualified.

Religion and Economic Development

Sriya Iyer

Abstract

The role of religion in economic development warrants a nuanced perspective that integrates economic theory with an understanding of socio-political structures, appreciating the econometric issues that arise in quantifying religious processes. Existing research focuses on religious structures and organizations, state religions, faith-based welfare programmes, the regulation of religion, and the impact of religion on measures of well-being such as income and education. Viewing religion as spiritual capital, with the attendant role played by religious network externalities in fostering economic development, is vital for development policy. Contemporary research in religion and economic development is flourishing, encompassing all these diverse concerns.

Keywords

Aquinas; St Thomas; Club goods; Economic demography; Education; Fertility; Health care; Human capital; Human capital; Insurance; Islamic economic institutions; Poverty alleviation; Protestantism; Religion and economic development; Religion; Economics of;

R

Religious capital; Religious networks; Schumpeter, J.; Self-selection; Smith, A.; Social capital; Social interaction (empirics); Social norms; Tawney, R.; Technology; Terrorism; Trust; Weber, M.; Well-being; Work ethic

JEL Classifications

O1; O4; Z12; Z13

> The number of micro-level social anthropological studies is continually growing. Many of these concentrate on what to the economist may appear odd aspects of society such as ritual and religion ... and to which he pays little or no attention. For instance, an understanding of the complex of Hindu religious beliefs as they operate at village level ... is directly relevant to the problem of developing India's economy. This is but one of numerous examples which can be quoted to support the claim that development economists work in the dark unless they acquaint themselves with the relevant socio-political literature. (Epstein 1973, p. 6)

How times have changed since Scarlett Epstein first lamented economists' general neglect of the role of religion in the study of economic development. She need not have been quite so fearful: contemporary economics has seen the light, as it were, increasingly demanding a perspective on religion in order better to understand how it interacts with economic decision making. The increasing resilience of religion in both developed and developing countries, influencing globally both political will and popular debate, has been observed by scholars investigating the economics of religion (Iannaccone 1998; Stark and Finke 2001; Glaeser 2005). Recent studies have investigated how religion affects growth (Guiso et al. 2003; North and Gwin 2004; Noland 2005; Barro and McCleary 2003; Glahe and Vorhies 1989) with emphasis on particular religious traditions such as Islam, Hinduism or Catholicism (Kuran 2004; Sen 2004; Fields 2003). Other studies have focused on the impact of religion on fertility (Lehrer 2004; McQuillan 2004). Still others examine the impact of religion on political outcomes (Glaeser et al. 2005) and the role of religious organizations as insurance (Dehejia et al. 2005). Other studies examine how the causality may run the other way, from economic development to religion (Berman 2000; Botticini and Eckstein 2005; Goody 2003).

Several theories have been advanced to account for the links between religion and development. First, there are theories that typify the 'rational choice' approach to religion and development. This approach considers the resilience of religion as a rational economic response to changes in the political, ecological and economic environments in which religions operate. In addition, a range of other structural theories encompass family socialization, social networks and a belief in other-worldly or supernatural elements. However, regardless of the scholastic tradition from which one approaches the study of religion, examining the interactions between religion and development poses significant challenges: first, to understand the endogenous interactions between religion and economic growth; second, to examine the techniques and methods needed to quantify these interactions; and third, to evaluate the impact of religion on development policy more widely.

Early Writings

The economic concern with religion and development is not new, nor is it restricted to scholars of the 21st century. The writings of Thomas Aquinas, notably the *De Regno (De Regimine Principum) ad Regem Cypri*, written in 1267, dealt extensively with religion and public finance. Indeed, some scholars have considered the ideas in this work, as in Aquinas's *Summa Theologica* (1265–72), strikingly relevant for poverty reduction today; their themes of the 'universal common good' and 'global civil society' have implications for current debates about globalization and human development (Linden 2003). The links between religion and development also feature in Joseph Schumpeter's *History of Economic Analysis* (1954). Jacques Le Goff authored *La Naissance du Purgatoire* (1981), which argued that purgatory was a necessary religious innovation for medieval capitalist development. However, it was in 1904 that Max Weber put forward his famous theory of the Protestant ethic and the spirit of capitalism, arguing that economic development in northern Europe could be explained by developments that were associated with Protestantism – the concern with savings, entrepreneurial activity, the frugality which Puritanism demanded, and the literacy needed to read the scriptures. The essence of Weber's thesis was that nascent capitalism emerged in the 16th century in Europe on account of the Protestant ethic which arose from the Reformation. Ascetic Protestantism encouraged diligence, discipline, self-denial and thrift. Both Lutheran and Calvinist doctrines urged adherents robustly to undertake their 'calling'. Spiritual grace from religion was attained by demonstrating temporal success in one's calling. The Protestant ethic thus involved the diligent undertaking of one's calling as a religious obligation, which promoted a work ethic that increased savings, capital accumulation, entrepreneurial activity, and investment, all of which in turn fostered economic development. Many scholars have criticized Weber's thesis, typified in the writings of Tawney (1926) and Gorski (2005). Tawney was concerned with reverse causality: how religion affected development, and in turn how economic and social changes themselves acted on religious beliefs. In his words, '"The capitalist spirit" is as old as history, and was not, as has sometimes been said, the offspring of Puritanism' (1926, p. 225). Tawney argued that Puritanism both helped mould the social order and in turn was moulded by it. Gorski (2005) focuses more on whether Weber's thesis stands up to closer historical scrutiny, highlighting other aspects of the Reformation that contributed to economic development such as Protestant migration, reforms to landholding, fewer religious holidays, and insurgencies, all of which influenced labour supply and the actions of government in Protestant countries.

The Economic View of Religion

Against this backdrop, recent academic interest linking religion and development has centred on the economics of religion. Studies in the economics of religion have focused on applying the tools of modern economic analysis to the analysis of religious institutions, faith-based welfare programmes and the economic regulation of the church (Oslington 2003). Three principal themes emerge: first, identifying what determines religion and religiosity; second, examining how religion and religiosity may be described as social capital; and third, understanding the micro and macro consequences of religiosity.

Adam Smith (1776) made reference to the church in the *Wealth of Nations*; and recent work by economists such as Becker and Iannaccone has been very important for the development of this field. The broadly socio-economic view of religion, which expounds the rational choice approach, is set out in the work of Azzi and Ehrenberg (1975), Iannaccone (1998), Stark et al. (1996), and Stark and Finke (2000). The focus here has been both on the supply side (the structures of religious organizations) and on the demand side (the preferences of consumers in religious economies). The micro view explains religious activity as the outcome of rational choice, with utility derived both in the individual's lifetime and in the afterlife. For example, if we think of religion as a club good, then many practices are used by religions to screen potential free riders and to ensure better monitoring of the existing faithful (Iannaccone 1992). Religion also influences individual welfare through the externalities occasioned by social behaviour (Becker and Murphy 2000). Religious forces are important as they change the environment in which individuals operate, directly affecting individuals' choices and behaviour by changing the utilities of goods. Moreover, greater trust fostered by the religious environment can encourage repeated interactions, leading to more cooperative behaviour within networks.

It is in this way that the second theme – religion as social capital – becomes important. Three aspects are emphasized here: social networks, social norms, and sanctions to penalize deviations from norms. Corresponding to this emphasis, economists of religion have been examining 'spiritual capital' – or religious capital – which embodies the norms, networks and sanctions exercised by groups that are organized on the basis of religion and religious networks.

Finally, the macro and micro consequences of religiosity have been examined. For example, there are a number of channels through which religious capital might affect economic growth. Religious capital affects output by changing the manner in which technology and human capital are used. Religious capital exerts a positive impact on human capital by increasing education. For example, particularly in many less developed countries, religious networks are important not only for the religious services they provide but also for their non-religious services, specifically with respect to health and education. Moreover, as religious institutions provide this insurance function, these networks determine the extent to which education is taken up (Borooah and Iyer 2005). In developed countries, too, this would have implications for religious market structure and the growth of residential neighbourhoods that may be based upon faith-based activities (Gruber 2005). So understanding the economic consequences of religion is of central concern.

The Empirics of Religion and Development

Most empirical economic studies of religion and development attempt to solve classic decompositions of the form $Y_i - Y_j = \sum\beta(X_i - X_j)$ where the idea is to examine the various factors (X) that affect measures of religious attendance or behaviour (Y) across individuals (i, j), or more widely across countries, or alternatively in varied historical time periods, thence to arrive at conclusions based on the effects suggested by the parameters (β) estimated.

Empirical studies of religion and development across countries have investigated religious movements, examining particularly sect behaviour, with an emphasis on contrasting the 'European experience of religious monopoly' with the 'American case of religious cacophony' (Warner 1993, p. 1081), drawing implications for the issue of whether regulation of religious organizations is necessary. This concern manifests itself in a plethora of research projects, especially on religion in the United States (Marty 1986–1996; Finke and Stark 1988; Warner 1993). In cross-country studies, economists have also revisited Weber's hypothesis. Barro and McCleary (2003) assess the effect of religious participation and beliefs on a country's rate of economic progress. Using international survey data for 59 countries drawn from the World Values Survey and the International Social Sciences Program conducted between 1981 and 1999, these authors find that greater diversity of religions is associated with higher church attendance and stronger religious beliefs. For a given level of church attendance, increases in some religious beliefs – notably belief in heaven, hell and an afterlife – tends to increase economic growth.

Other studies have focused more on particular religions in varied historical time periods. For example, very useful insights have been gained by focusing on Islam and on Judaism. For Islam, there have been detailed investigations into financial systems in the Middle East including *zakat* (alms for charity) and the manner in which Islamic banks have been using a financing method equivalent to the rate of interest to overcome adverse selection and information problems. There has also been more detailed investigation into Islamic law and financial activity historically with implications for poverty reduction in the Middle East (Kuran 2004). There is research that has examined Jewish occupational selection using historical data from the eighth and ninth centuries onward to explain the selection of Jews into urban, skilled occupations prompted by educational and religious reform in earlier centuries (Botticini and Eckstein 2005). Data are also being used to elucidate the role of religion in explaining historical differences in education among Hindus and Muslims in India (Borooah and Iyer 2005).

A primary focus of current studies of religion and development is on explaining differences across individuals. For example, using data from the General Social Survey and the US Census, Gruber (2005) investigates religious market structure by estimating the effects of religious participation on economic measures of well-being, and concluded that residing in an area with more co-religionists improves well-being through the impact of increased religious participation. This particular study is also valuable from the methodological point of view, as it addresses a common problem in empirical studies of religion and development – the persistent endogeneity of religion to economic measures of well-being – and consequently the common econometric problem of how best to identify religion effects. While this particular study successfully uses ethnic heritage to provide an exogenous source of variation, and is thereby able to draw out cleanly the effects of religious participation on the variables of interest, econometrically the potential endogeneity of most religion variables is possibly the single most significant limitation of incorporating religion into empirical work in economics. This is mirrored in the many efforts to identify the effects of religion which generally have not been able to deal with self-selection issues easily.

To this end, fields such as economic demography have much to offer the study of religion and development. For example, recent research in economics has made a start towards examining the religious and economic reasons behind fertility differences between religious groups, especially in developing countries (Iyer 2002). The economics of religion has also elucidated the study of politics, both local and international: Glaeser (2005) presents an economic model of religious group behaviour and the so-called 'political economy of hatred'. The economic approach to religion has been evaluating whether religion and politics are mutually exclusive. Glaeser et al. (2005) link religion with strategic extremism – the issues and platforms espoused by political parties, and the manner in which private information matters for this. Other studies have focused on terrorism and display a more general preoccupation with understanding views and attitudes in the Muslim world (Gentzkow and Shapiro 2004).

Drawing a perspective from all these classes of studies, it strikes one that emerging economies are experiencing appreciable modern economic growth, yet this is coterminous with the increasing resilience of religious institutions. And it is this dichotomy between the sacred and the secular which epitomises the puzzle of the relationship between religion and economic development. It seems reasonable to address this puzzle by combining quantitative analysis of sample data with nuanced qualitative evaluations of the textual theology of religion, linking these to the manner in which individuals and institutions interpret religion at a local level. As well, an appreciation of the approach of the interdisciplinary economist would permit a more informed understanding of all these concerns. Economists will enthusiastically study religion and economic development in the future, and they will do so with ascetic assiduity – researching data with all the intensity of religious fervour in order to provide thoughtful prophecy for development policy.

See Also

▶ Islamic Economic Institutions
▶ Religion, Economics of
▶ Social Capital
▶ Social Interactions (Empirics)
▶ Weber, Max (1864–1920)

Bibliography

Aquinas, St Thomas. 1265–1272. *Summa theologica*. Online. Available at http://www.ccel.org/ccel/aquinas/summa.html. Accessed 22 June 2006.

Aquinas, St Thomas. 1267. *De Regno (De Regimine Principum) ad Regem Cypri*. In *Aquinas: Political writings*, ed. and trans. R.W. Dyson, Cambridge: Cambridge University Press, 2002.

Azzi, C., and R. Ehrenberg. 1975. Household allocation of time and church attendance. *Journal of Political Economy* 83: 27–56.

Barro, R.J., and R. McCleary. 2003. Religion and economic growth across countries. *American Sociological Review* 68: 760–781.

Becker, G., and K. Murphy. 2000. *Social economics: Market behavior in a social environment*. Cambridge: Harvard University Press.

Berman, E. 2000. Sect, subsidy and sacrifice: An economist's view of ultra-orthodox Jews. *Quarterly Journal of Economics* 115: 905–953.

Borooah, V., and S. Iyer. 2005. *Vidya, veda* and *varna*: The influence of religion and caste on education in rural India. *Journal of Development Studies* 41: 1369–1404.

Botticini, M., and Z. Eckstein. 2005. Jewish occupational selection: Education, restrictions, or minorities? *Journal of Economic History* 65: 922–948.

Dehejia, R., T. DeLeire, and E. Luttmer. 2005. Insuring consumption and happiness through religious organizations. Working paper no. 11576. Cambridge: NBER.

R

Epstein, T. 1973. *South India: Yesterday, today and tomorrow*. London: Macmillan.
Fields, B. 2003. *The Catholic ethic and global capitalism*. Aldershot/Burlington: Ashgate.
Finke, R., and R. Stark. 1988. Religious economies and sacred canopies: Religious mobilization in American cities, 1906. *American Sociological Review* 53: 41–49.
Gentzkow, M., and J. Shapiro. 2004. Media, education and anti-Americanism in the Muslim world. *Journal of Economic Perspectives* 18(3): 117–133.
Glaeser, E. 2005. The political economy of hatred. *Quarterly Journal of Economics* 120: 45–86.
Glaeser, E., G. Ponzetto, and J. Shapiro. 2005. Strategic extremism: Why Republicans and Democrats divide on religious values. *Quarterly Journal of Economics* 120: 1283–1330.
Glahe, F., and F. Vorhies. 1989. Religion, liberty and economic development: An empirical investigation. *Public Choice* 62: 201–215.
Goody, J. 2003. Religion and development: Some comparative considerations. *Development* 46(4): 64–67.
Gorski, P. 2005. The little divergence: The Protestant Reformation and economic hegemony in early modern Europe. In *The Protestant ethic turns 100: Essays on the centenary of the Weber Thesis*, ed. W. Swatos and L. Kaelber. Boulder/London: Paradigm Publishers.
Gruber, J. 2005. Religious market structure, religious participation, and outcomes: Is religion good for you? *Advances in Economic Analysis & Policy* 5(1), article 5.
Guiso, L., P. Sapienza, and L. Zingales. 2003. People's opium? Religion and economic attitudes. *Journal of Monetary Economics* 50: 225–282.
Iannaccone, L. 1992. Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy* 100: 271–291.
Iannaccone, L. 1998. Introduction to the economics of religion. *Journal of Economic Literature* 36: 1465–1495.
Iannaccone, L., R. Finke, and R. Stark. 1997. Deregulating religion: The economics of church and state. *Economic Inquiry* 35: 350–364.
Iyer, S. 2002. *Demography and religion in India*. Delhi: Oxford University Press.
Kuran, T. 2004. Why the Middle East is economically underdeveloped: Historical mechanisms of institutional stagnation. *Journal of Economic Perspectives* 18(3): 71–90.
Lehrer, E. 2004. Religion as a determinant of economic and demographic behaviour in the United States. *Population and Development Review* 30: 707–726.
Le Goff, J. 1981. *La Naissance du Purgatoire*. Paris: Gallimard.
Linden, I. 2003. *A new map of the world*. London: Darton, Longman and Todd.
Marty, M. 1986–1996. *Modern American religion*. Vol. 3. Chicago: University of Chicago Press.
McQuillan, K. 2004. When does religion influence fertility? *Population and Development Review* 30: 25–56.
Noland, M. 2005. Religion and economic performance. *World Development* 33: 1215–1232.
North, C., and C. Gwin. 2004. Religious freedom and the unintended consequences of state religion. *Southern Economic Journal* 71: 103–117.
Oslington, P., ed. 2003. *Economics and Religion*, The international library of critical writings in economics 167. Vol. 1 and 2. Cheltenham: Edward Elgar.
Schumpeter, J. 1954. *History of economic analysis*. London: Oxford University Press.
Sen, A. 2004. Democracy and secularism in India. In *India's emerging economy: Performance and prospects in the 1990s and beyond*, ed. K. Basu. Cambridge/London: MIT Press.
Smith, A. 1776. In *An inquiry into the nature and causes of the Wealth of Nations*, 5th ed., ed. E. Cannan. London: Methuen and Co. Ltd. , 1904.
Stark, R., and R. Finke. 2000. *Acts of faith: Explaining the human side of religion*. Berkeley: University of California Press.
Stark, R., and R. Finke. 2001. Beyond church and sect: Dynamics and stability in religious economies. In *Sacred markets and sacred canopies: Essays on religious markets and religious pluralism*, ed. T. Jelen. Lanham: Rowman and Littlefield.
Stark, R., L. Iannaccone, and R. Finke. 1996. Religion, science, and rationality. *American Economic Review* 86: 433–437.
Tawney, R. 1926. *Religion and the rise of capitalism*, 1990. London: Penguin Books.
Warner, R. 1993. Work in progress toward a new paradigm for the sociological study of religion in the United States. *American Journal of Sociology* 98: 1044–1093.
Weber, M. 1904. *The Protestant ethic and the spirit of capitalism*, trans. T. Parsons. London: Routledge, 1992.

Religion, Economics of

Laurence R. Iannaccone and Eli Berman

Abstract

Adam Smith invented the economics of religion, famously arguing for church-state separation on efficiency grounds since state religions become inefficient monopoly providers of religious services and because competition for monopoly status is often violent. Smith also developed theories of religious sects and sectarian violence. Modern applications of theory and data generally support Adam Smith's conjectures. Recent work also explores: religious activity as a consumer choice, the demand for spiritual services, religious human capital and religious social capital, club models of sects – benign and violent – and the macroeconomic consequences of beliefs and religiosity.

Keywords

Addiction; Boulding, K.; Capitalism; Charitable Donations; Clubs; Fertility; Firm, Theory of; Free-Rider Problem; Human Capital; Intertemporal Utility; Marriage and Divorce; Mutual Aid; Rational Choice; Rationality; Religion, Economics of; Religious Capital; Religious Economics; Rent Seeking; Sect; Social Capital; Social Cohesion; Social Norms; Stable Preferences; Terrorism; Weber, M.; Women's Work and Wages

JEL Classifications

Z12

Adam Smith laid the foundations for the economic study of religion in *The Wealth of Nations* (Smith 1776, pp. 788–814). He argued that self-interest motivates the clergy; that market forces constrain churches just as they constrain secular firms; that competition improves the quality of religious services provided; and that government regulation distorts the provision of religion, reducing quality and promoting conflict. He also outlined a theory of sectarianism, a theory of religious violence and civility, and a general theory of Church and State.

After this inspired start the economics of religion lay dormant and nearly dead for two centuries. It is now enjoying a rebirth, animated by new data, methods and theory. Economists and other social scientists have harnessed rational choice models and modern empirical tools to study secularization, pluralism, church growth, religious extremism, conversion, fertility, Church–State relations, and more. The field now claims hundreds of papers, scores of contributors, an annual conference and international association (the Association for the Study of Religion, Economics, and Culture), university research centres, and even an AEA subject code (Z12). (New university centres are at Harvard, George Mason University, the University of Southern California and in Canberra, Australia.)

Current research on religion and economics falls into three related subfields: economic theories of religion, studies of religion's economic consequences, and religious assessments of economic policy. Adam Smith's critique of state-supported religion in the *Wealth of Nations* exemplifies the first subfield; Max Weber's 'Protestant ethic' conjecture the second. Together these two subfields constitute *the economics of religion* – the subject of this article. Our goal is to introduce readers to the distinctive economic ideas and models that have enhanced the social-scientific study of religious beliefs, behaviour, and institutions. (For a more complete review of the literature prior to 1998, see Iannaccone 1998.)

This article makes no attempt to survey the field of religious economics, both because the latter tends to be religion-specific and because it is far from the mainstream of economic research. Religious economics seeks to evaluate economic behaviour and institutions in the light of sacred precepts. Mahmoud El-Gamal's recent book, *Islamic Finance* (El-Gamal 2006), is a good example, examining whether current practices in banks that follow Islamic law actually serve the objectives of those laws. The literature on religious economics is large, diverse, and as old as religion itself – including, for example, the many biblical injunctions concerning property, slavery, wages, tithing, interest, wealth and poverty. With the help of economists and philosophers, contemporary clerics continue to debate the merits of income inequality, tax laws, private property, deficit spending, monetary policy, income redistribution, workers rights, interest rates, banking laws, entrepreneurship, government regulation, international trade, debt relief, unionization, entitlement programmes and much more. For representative readings in religious economics, see Oslington (2003) and the *Journal of Markets and Morality* (published by the

R

Acton Institute for the Study of Religion and Liberty).

Social scientists once viewed religion as a dying vestige of our primitive and pre-scientific past. Modern research and contemporary events have destroyed this simplistic view of human history. The rise of radical Islam, the revival of religious practice in much of the former Soviet Union, the explosive growth of Protestantism in Latin America and sub-Saharan Africa, and the contribution of religion to identity (and often to conflict) all over the world testify to the continuing vitality of religion. And although religious belief and activity have declined in many economically advanced countries since the 1960s, the corresponding US data display remarkable stability, whether one focuses on rates of attendance, contributions, membership, or belief. Indeed, religiosity has become one of the strongest predictors of voting patterns and political orientation in America (Glaeser et al. 2006).

We cannot say why economics ignored religion for so long. The brilliant and iconoclastic economist Kenneth Boulding discussed economic features of religion long before the modern revival, but his insights seem to have gone largely unnoticed by economists or sociologists. (Boulding's essays on religion and economics from the 1950s appear in Boulding 1970.) The other social sciences have subfields dedicated to the study of religion and most have sought to understand the connection between religious and economic trends – the most famous and influential generalization being Max Weber's 'Protestant ethic' thesis. (Weber studied economic history and was well-acquainted with Smith's *Wealth of Nations*. His essay, 'The Protestant Sects and the Spirit of Capitalism', describes how denominational membership enhanced the reputation and business prospects of Americans around about the ways in which sect membership benefits a poor person: see Smith 1776.) It seems likely, however, that most economists saw religion as too far removed from the realm of rational choice and market behaviour. We encourage the reader to revisit the issue of religion and rationality after reading this essay.

Economics, Sociology, and Rational Choice

Nearly all economic theories rely on the twin assumptions of *rational choice* and *stable preferences*. In the realm of religion, this means choosing which religion, if any, to accept and how extensively to participate in it. These optimal choices need not be permanent. Economic models do a good job explaining differences in religious activity, both over time and across individuals. In keeping with the assumption of stable preference, however, these explanations rarely invoke varied norms, tastes or beliefs. A good economic story explains behaviour in terms of optimal responses to varying circumstances – such as prices, incomes, skills, experiences, technologies or resource constraints.

Although the previous paragraph merely extends modern economic orthodoxy to the realm of religion, it borders on sociological heresy. The commitment of economists to rational choice and stable preferences must be understood as relative, not absolute. Since the late 1970s, economists have devoted a great deal of attention to modelling preference formation. Formal models of religious capital formation (Iannaccone 1984, 1990) are, in fact, directly linked to Becker's (1996) subsequent work on rational addiction and taste change. Recent work in the fields of behavioural, experimental and evolutionary economics underscores the extent to which choice systematically deviates from rationality; and social norms, social networks and imperfect information constrain choices further still. But it would be wrong to conclude that economists and sociologists now embrace a common 'world view' – as is readily apparent when one contrasts the papers presented by economists and sociologists at the joint annual meetings of the *Association for the Study of Religion, Economics, and Culture* and the *Society for the [Social] Scientific Study of Religion*. Most sociologists remain very sceptical about the value of formal models, rational choice theory and methodological individualism – a legacy passed down from the founders of the field, who promoted sociology as a corrective to errors and omissions of economics (Swedberg 1990). Add the influence of Weber (1920, 1963), who made 'rationality' central to his analysis of religion while using the word in ways foreign to most contemporary economists, and 'doctrinal' debate is unavoidable. But the overall response to economic forays into religion has been surprisingly ecumenical, with several leading sociologists of religion going so far as to characterize economic theory and market models as 'the new paradigm' for religious research (Stark and Finke 2000; Warner 1993; Young 1997).

Households and Consumer Choice

Economists finally returned to the study of religion in the 1970s, inspired by Gary Becker's path-breaking work on economics of the family. The first papers modelled church attendance and religious contributions as a special form of household production – one that involved trade-offs between time and money inputs, secular versus religious outputs, and present versus afterlife utility (Azzi and Ehrenberg 1975; Ehrenberg 1977). Formally, households maximize an intertemporal utility function $U = U(Z_1, \ldots, Z_n, A)$, where Z_t denotes secular consumption activities in period t, and A is consumption activity in an afterlife (of possibly infinite duration). In each period (of this life) households can spend their time, T, and goods, X, on either secular consumption or religious activities, $Z_t = Z(T_{Z_t}, X_{Z_t}), R_t = R(T_{R_t}, X_{R_t})$. Religious activities over a lifetime create afterlife consumption, $A = A(R_1, \ldots, R_n)$. Combined with a standard lifetime budget constraint, and on the assumption that the marginal product of religious activity does not decrease with age, the Azzi–Ehrenberg model predicts that religious activity increases with age, *ceteris paribus*. The model also predicts that households with high value of time (high wages) will substitute goods for time in producing religious activity.

As Azzi and Ehrenberg (and many others) have shown, religious activity does tend to increase with age. But it is not at all clear that the Azzi–Ehrenberg model captures the principal cause of this age effect. Ulbrich and Wallace (1983) found that activity increases with age even among those who do *not* believe in the afterlife. And Iannaccone (1984) showed that even in the absence of afterlife expectations, the rational accumulation of religious human capital (that is, rational religious 'addiction') could simultaneously account for the observed age effect as well as observed patterns of religious conversion, intermarriage, and marital stability (see also Lehrer and Chiswick 1993; Neuman 1986).

Predictions concerning religious substitution are on much stronger ground. Substitution of goods for time is observed across individuals, households and denominations. Although we cannot directly observe most religious commodities, we can observe the inputs used to produce them. The principal time and money inputs – attendance and contributions – are routinely measured in surveys. More specialized studies provide detailed information on time (such as time devoted to religious services, private prayer and worship, religious charity, and many other religious activities) and money (such as expenditures for special attire, transport, religious books and paraphernalia, sacrificial offerings, and contributions used to finance staff, services and charitable activities of religious organizations). Several studies, including Ehrenberg (1977), Iannaccone (1990), Hungerman (2005) and Gruber (2004), have found that attendance and donations are substitutes – and the recent work demonstrates that substitution remains strong even after one controls for endogeneity bias.

Both in theory and in fact, substitution induces different methods of religious organization and worship across different socio-economic strata. High-income congregations tend to hold shorter services, make heavy use of professional staff and inhabit more elaborate facilities. Longer services, volunteer workers, rented meeting halls and pot-luck dinners are typical of poorer congregations. We observe these differences within denominations and even within congregations (as members improve their socio-economic status), but the differences are especially stark *across* the denominations of a religious tradition, such as Reform Judaism versus Orthodox Judaism or Episcopalians versus Southern Baptists. Many Episcopalian or Presbyterian congregations have plenty of money to cover salaries and operating expenses but remain hard-pressed to recruit volunteers for their choirs, youth programmes, committees and other traditional programmes. For such denominations prosperity has proved a mixed blessing.

Economic trends forced adaptation and none more so than the growth of women's wages and workforce participation. As women have moved into the labour force and overall family earnings have grown, congregations have had to purchase and Nemeth's (Luidens and Nemeth 1994) study of expenditure trends in Presbyterian denominations, which found that their (fourfold) increase in real per-capita giving from the 1940s to the 1990s was spent primarily on local congregational services previously supplied by volunteers.

Religion, Magic and Uncertainty

Contemporary theories of rational religious belief begin with just a few assumptions about human nature and the human condition – in essence scarcity, rationality, and the capacity to conceive of supernatural beings or forces (Iannaccone and Berman 2006; cf. Stark and Bainbridge 1987). From these, they derive a universal demand for supernaturalism and a universal distinction between magic (emphasizing control of impersonal supernatural forces) and religion (emphasizing interaction with supernatural beings). Specialized suppliers arise naturally in both realms, but markets for magic and religion operate quite differently. It is relatively easy to test (and disprove) a magician's ability to control supernatural forces, but much harder to falsify a priest's claims concerning God. In practice, only religion can sustain long-term relationships, high levels of commitment, and moral communities. As Emile Durkheim (1915, p. 42) famously observed, 'there is no church of magic'.

As we shall see in Section 6, a strong religion can induce its members to foreswear all other suppliers of supernatural goods and services. But exclusivity is not a 'natural' outcome. Given the tremendous uncertainty that surrounds the supernatural, rational consumers are inclined to patronize many different suppliers – investing, so to speak, in diversified *portfolios* of supernatural commodities (Iannaccone 1995). Diversification over different supernatural products and suppliers is pervasive in the (non-communal) market for magic, including the so-called 'New Age' movement. It also prevails in most polytheistic settings, including in the Greco-Roman world, and it remains common in Asian religious traditions. Judaism, Christianity and Islam display a much greater capacity to sustain exclusivity, but (as we shall discuss in Section 6) only within communal settings that promote collective action, strong social ties, and large investments in religious capital.

Religious Capital

James Coleman's (1988, p. 97) concept of 'social capital' helps connect rational choice theory to sociological analysis. Iannaccone's (1984, 1990) concept of 'religious capital' offers an analogous bridge from rational choice theory to the sociology of religion. Both concepts are inspired by human capital theory (Becker 1964; Schultz 1961), and both emphasize relationships rather than purely individual capacities.

Let S_{R_t} denote the stock of relationships, sensitivities and skills that alter a person's real or perceived benefits from religious activity at time t. Religious commodity production thus depends on current inputs of time and money and the current stock of religious capital, $R_t = R(T_{R_t}, X_{R_t}, S_{R_t})$. The SR variable can encompass a range of concepts, including religious habits, spiritual capital and social capital. Indeed, the mathematical models and empirical analyses remain essentially the same whether one frames the model in terms of the formation of religious 'preferences' or the accumulation of (unobservable) religious 'capital'. In either case, however, religious experience has two key features. First, past experience alters the value of current religious activities and thereby affects rates of religious participation: $S_{R_t} = F(T_{R_t}, X_{R_t}, S_{R_{t-1}})$. Second, most religious experience is 'context specific', yielding maximal benefits within the context of specific relationships, congregations, denominations and traditions. Religious capital remains a distinctive form of social and human capital because religions claim to promote relationships with supernatural beings. This enables religious institutions to maintain exceptionally high levels of commitment, but not without collective production, exclusivity and sacrifice. The bundle appeals less to people with better secular opportunities; hence we observe a 'church-to-sect' *spectrum* of denominations within most religious traditions. For details, see Iannaccone (1995) and Iannaccone and Berman (2006).

Capital models yield predictions that are well supported by evidence, including: (*a*) children tend to choose the same or similar religious denominations as did their parents; (*b*) conversion (like career choice) tends to occur early in adulthood, leaving time to accumulate religion-specific capital; (*c*) interfaith marriage is less likely when religious capital accumulation is high; (*d*) shared-faith marriages lead to higher rates of religious participation (due to complementarities in household production, and not mere sorting of more religious partners into shared-faith marriages); and (*e*) shared-faith marriages have lower rates of divorce and higher rates of fertility (Iannaccone 1990; Lehrer and Chiswick 1993; Waite and Lehrer 2003).

Religion also contributes to extended relationships, social networks and shared norms. Indeed, Coleman's (1988) seminal article on social capital concerned the impact of (Catholic) religious schools. Empirical studies find that nearly half of all associational memberships, personal philanthropy and volunteering in the United States is church-related, leading Putnam (2000) to conclude that '[f]aith communities ... are arguably the single most important repository of social capital in America'. Yet social capital research has yet to give much attention to religion – see, for example, the literature review by Sobel (2002). There remain tremendous opportunities for policy-relevant research on religion's contribution to cooperation (Sosis and Ruffle 2003), social multipliers (Becker and Murphy 2000), threshold effects (Granovetter 1978), public preferences (Kuran 1995), and much more.

Measuring the Effects of Religious Capital

Numerous empirical studies suggest that religious belief and participation yield a wide range of benefits, including mental and physical health, longevity, reduced substance abuse and marital stability (see Koenig et al. 2001, for an extensive review of the relevant research). The statistical results must, however, be viewed with caution. We lack good instruments for religion on both the supply and demand sides, and most research examines only contemporary American data. Problems of spurious correlation and unobserved heterogeneity may afflict many published studies, as Heaton (2006) notes in his re-analysis of data on religion and crime. On the other hand, the positive association between religion and health has held up despite many different efforts to root out spuriousness, and Freeman's (1986) careful data analysis provides compelling evidence that church attendance really does lead to higher employment rates, higher school attendance, less crime and lower alcohol consumption and drug use among Black males in the United States.

There are many plausible reasons why religiosity might promote beneficial outcomes. As Adam Smith emphasized in his *Theory of Moral Sentiments* (Smith 1759), faith in an omniscient deity can solve otherwise intractable problems of self- and social control. Since religion is the quintessential credence good, religious institutions tend to be relatively efficient producers of moral restraint. And there can be no doubt that communities of faith do provide many concrete services while seeking to instil faith in the young, maintain faith among adults and constrain deviant behaviour. The potential benefits from these mechanisms are underscored by the (not undisputed) evidence that religious constraints on sexual conduct have reduced or limited AIDS among Muslims in central Africa and Christians in Uganda (Green 2003).

Club Models of Religion

Club models have made major contributions to our understanding of 'cults', 'sects' and religious extremism. They also account for characteristics of religion that seem inconsistent with rational choice and risk-aversion – including the success of groups that demand exclusivity, sacrifice and stigma.

Club models start with the fundamental fact that religious 'commodities' are more compelling and gratifying when they are produced and consumed *in groups*. Effective congregations require highly committed members, not mere customers. In this respect, effective congregations are more like families than firms. This suggests that models in which the *i*th member's religious satisfaction has the form $R_i = F(T_{R_i}, X_{R_i}, Q)$, where Q is an index of the religious inputs of all the *other* group members.

As economists well know, shirking and free-riding constantly threaten collective action, especially in large groups. Paying people to attend church, accept church doctrine or support fellow members fails to solve the problem because a member's commitment and inputs to the group are difficult to observe, and payment rewards the wrong motivations. But the problems *can* be mitigated by seemingly gratuitous costs – the sacrifice and stigma characteristic of deviant religious. Sacrifice and stigmas enhance utility by screening out people who lack commitment and boosting involvement among those who remain in the group. Such groups manifest many distinctive characteristics that empirical researchers have long associated with 'sectarian' religions, including distinctive diet, dress or sexual conduct; physical separation from mainstream society; painful or costly rites; rules that limit social contact with non-members; and prohibitions restricting normal economic or recreational activities. (For more on the modern theory of church and sect, see Iannaccone 1988, 1991.)

Sect theory also accounts for people's willingness to forgo religious diversification despite the obvious risk associated with most religious assurances. Sectarian religions can maintain levels of commitment and involvement that compensate for the increased risk associated with exclusivity. Corresponding constraints can almost never be sustained in standard, secular markets (nor in the impersonal market for magic) because exclusivity does not enhance the production non-collective goods and services (Iannaccone 1995).

The club model has received wide acceptance, in part because it fits the data so well. Both cross-sectional surveys and case studies find substantially higher levels of mutual aid and social cohesion in more sectarian religious communities.

Iannaccone's work on (mostly Christian) sects has been extended to radical religious Jews (Berman 2000) and Muslims (Berman and Stepanyan 2003; Chen 2004). Despite some lingering debate over the extent of free-rider problems in mainline churches or the actual level of costs imposed by contemporary conservative churches, the basic model remains the natural starting point for studies of high-cost groups. The club model works well not only for religious groups routinely called sects, cults and fundamentalists, but also for communes, gangs, radical militias and terrorist organizations. The basic insight is that an organization designed to exclude free-riders and limit free-riding will be well equipped to exclude defectors and limit defection, the Achilles' heel of militias and terrorists. Thus religious sects prove to be especially effective at terrorism, militia activity and suicide terrorism (Berman 2003; Berman and Laitin 2005; Berman and Stepanyan 2003).

Churches as Firms

Many religious organizations are legally designated as firms, and many more look surprisingly firm-like. Around the time economists became interested in religious households, several sociologists of religion began thinking of churches as firms, re-examining old data sources with new theories of rational exchange, entrepreneurship and market competition (Finke and Stark 1988; Stark and Bainbridge 1985). Finke and Stark (1992) trace the explosive growth of Methodist and Baptist churches in 19th century America to superior marketing, organization and clergy incentives. By the 1990s, these economic and sociological streams of scholarship together included studies of sectarianism, denominational vitality, 'franchising' of religious brands, religious extremism, doctrinal innovation, Church and State, religious markets, non-Western faiths, religious history, and more. Ekelund et al. (1996) analyse numerous features of medieval Catholicism in terms of its monopoly status. Drawing upon standard theories of monopoly, rent seeking and transaction costs, they offer economic explanations for interest rate restrictions, marriage laws, the Crusades, the organization of monasteries, indulgences, and the doctrines of heaven, hell and purgatory (see also Ekleund, Hébert, and Tollison, Ekelund et al. 2006). Work on churches as firms continues to grow rapidly, in part because firms are easier to model than clubs, but also because the theory of the firm is so rich in predictions and data.

Religious Markets and Government Intervention

Whether we think of them as clubs or as firms, individual denominations collectively constitute a *religious market* as long as they provide services that are substitutes. The theories of religion described above predict the existence of different market segments: exclusive ‘sects’ that operate like clubs, inclusive ‘churches’ sustained by a core of professionals which are more firm-like, and markets for ‘magic’ organized around simple exchanges between practitioners and clients.

Almost all economists and sociologists of religion accept the notion that religion in America constitutes a vast competitive market, overflowing with ‘products’ that range from New Age paraphernalia to orthodox liturgies. Scholars likewise accept that market success requires entrepreneurship, innovation, and sensitivity to the demands of consumers. As a result, themes that rarely surfaced prior to Finke and Stark’s *Churching of America* (Finke and Stark 1992) now parade as common sense. Even the harshest critics of rational choice theory (such as Bruce 1999), emphasize the centrality of religious choice in today’s world.

The most informative studies closely study how markets actually work. Market-oriented research must carefully address numerous issues, including product attributes, marketing strategies, incentive structures, exchange relationships, consumer characteristics, and Church–State relationships. Andrew Chesnut’s (2003) study of rapidly growing religious movements in Latin America illustrates this point by showing how specific religions offer distinctive products that directly address the health- and family-oriented concerns of poor and middle-class women. Anthony Gill (1998) shows that Catholic bishops are much more likely to side with the poor in Latin American countries where Protestant growth threatens the Church’s historic monopoly.

Adam Smith (1776, pp. 788–814) argued that established religions face the same incentive problems that plague other state-sponsored monopolies: lack of competition generates a low quality product.

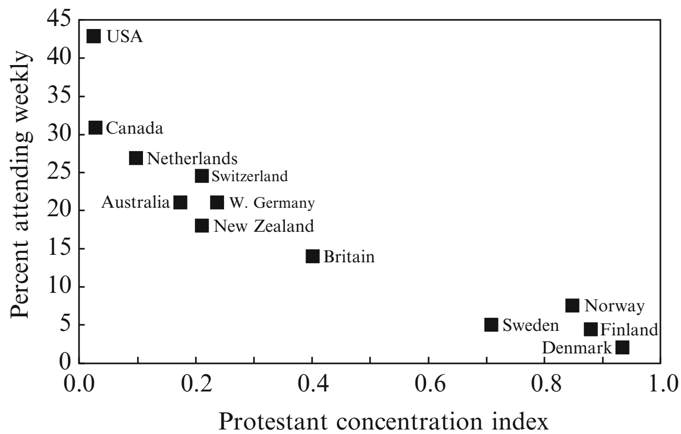
> The teachers of [religion] .. ., in the same manner as other teachers, may either depend altogether for their subsistence upon the voluntary contributions of their hearers; or they may derive it from some other fund to which the law of their country many entitle them.. .. Their exertion, their zeal and industry, are likely to be much greater in the former situation than the latter. In this respect the teachers of new religions have always had a considerable advantage in attacking those ancient and established systems of which the clergy, reposing themselves upon their benefices, had neglected to keep up the fervour of the faith and devotion in the great body of the people.

Iannaccone tested Smith’s conjecture with modern data. Fig. 1 illustrates that within predominantly Protestant countries, church attendance declines sharply as the religious market becomes more concentrated. (The Herfindahl-style ‘Protestant Concentration Index’ proxies state support for particular religions and has the form $H = \sum_i S_i^2$, where S_i is the population share of the ith Protestant denomination.) All other surveyed measures of religiosity, including belief in God, fall with concentration as well. The data, and Smith’s theory, strongly suggest that America’s ‘religious exceptionalism’ is largely a product of religious laissez-faire. North and Gwin (2004) report similar results using a much larger number of countries and more direct measures of Church–State relationships.

Several studies have found positive correlations between local levels of religious diversity and religious activity *within* the USA, including an

R

Religion, Economics of, Fig. 1 Market concentration and church attendance (*Source*: Gallup polls. See Iannaccone (1991) for details)



especially well-crafted study by Finke et al. (1996). But much of this work suffers from specification problems that inevitably arise when rates of religious *membership* (as opposed to more direct measures of belief and participation) are regressed onto (membership-based) measures of religious diversity (Voas et al. 2002). Nor is it clear that concentration *should* signal the presence of inefficient religious 'monopoly' across cities or states given the nation's minimal barriers to religious entry and innovation. There is, however, strong historical evidence that religious competition raises religious participation (note especially the work of Finke and Stark 1992, and Olds 1994, who show that post-colonial disestablishment led to the rapid growth in overall church membership rates, clergy demand, and the – primarily Baptist and Methodist – non-established denominations, while the major denominations that had enjoyed the support of colonial governments – primarily Episcopal, Presbyterian, and Congregational) rapidly *lost* market share).

Macroeconomic Consequences of Religion

Weber's *The Protestant Ethic and the Spirit of Capitalism* (Weber 1920) famously claimed that the Calvinist doctrine of predestination triggered a mental revolution within Protestantism that gave rise to modern capitalism. This remains the most influential single conjecture on the macroeconomic effects of religion, which is unfortunate since nearly all subsequent empirical research has shown it to be *false*.

Almost all the capitalist institutions that Weber emphasized actually preceded the Reformation (Stark 2005; Tawney 1998). Across and within European countries economic development was uncorrelated with religion (Samuelsson 1993; Delacroix and Nielsen 2001). The second country to industrialize was Belgium, which is Catholic. Although Germany and the Netherlands were early developers and majority Protestant, the fastest growth within those countries was among Catholic families of the Rhineland and Amsterdam, which were majority Catholic.

Despite much work by historians and sociologists, there is no consensus concerning the macroeconomic impact of Protestantism, Christianity, monotheism, or religiosity in general. Economists have recently entered this field of enquiry with cross-national studies of survey and census data. Barro and McCleary's (2003) cross-national analysis of survey and census data suggests that belief in hell boosts economic growth whereas frequency of church attendance retards growth (perhaps because the former induces honest and industry whereas the latter reduced time spent working). Using cross-national data from the World Value Surveys, Guiso et al. (2003) find that religious beliefs in general, and Christian beliefs in particular, are positively associated with economic attitudes conducive to higher per capita income and growth. Many other economists have begun doing similar studies, but data problems abound. In addition to standard econometric difficulties, there are scarcely any cross-national religious surveys that predate the 1980s; we cannot validate most responses concerning religion; and the meaning of religious participation and belief varies dramatically across cultures. There is better evidence of links between religious and socio-economic variables at the level of individuals and groups than for countries or cultures. For example, average family size and socio-economic status differ quite substantially across different religious groups in both rich and poor nations (Chiswick 1983; Iyer 2002). Historical studies do suggest strong relationships between religious and economic institutions, most notably those of medieval Europe. Ekelund et al. (1996) interpret many distinctive features of medieval Catholicism as forms of rent seeking, and there is no doubt that the Church was by far the most important economic institution in medieval Europe. Richardson (2005) offers strong evidence that the doctrine of purgatory gained rapid acceptance because it served to link religious and economic activities (within guilds) in a way that solved commitment problems that arose because of the social disruptions induced by the Black Death. Timur Kuran (2004) makes a compelling case that specific Islamic legal institutions contributed significantly to the economic decline in Muslim countries relative to those of Europe over the past 500 years.

Recent attempts to promote development in poor and post-Communist countries affirm the importance of ethical norms and moral precepts, many of which have religious foundations (Hayek 1988, pp. 135–40). Communism may be the most striking example of an economically and socially destructive religion, albeit a religion without traditional deities. In this sense, the strongest evidence for Weberian-style theory may be negative: some powerful systems of belief *do* retard economic progress.

Religious Militancy

American economists have tended to ignore religion as a subject of public policy, in part because the 'establishment' and 'free exercise' clauses of the First Amendment radically limit the religious role of government. Within this constitutionally mandated environment of religious laissez-faire (which initially constrained the federal government, but later extended to the states), Americans have maintained extraordinarily high rates of religious activity, diversity and tolerance. But elsewhere, religion remains a major factor in wars, civil unrest and ethnic conflict.

Adam Smith recognized that a detached and lazy clergy was just one cost associated with the marriage of Church and State. When government favours a particular religion in return for its support of the state, the favoured group inevitably demands the suppression of its competitors, and all other groups resist suppression and fight to capture favoured status. It is no coincidence that the USA has remained remarkably free of religious partisanship and militancy while other nations burn with religious conflict. Policies analogous to those embodied in the First Amendment's free exercise and establishment clauses may be key components of the so-called 'war on terror' (Iannaccone and Berman 2006).

Conclusion

The economics of religion has animated research on secularization, pluralism, church growth, religious extremism, religious markets, the consequences of religion, and more. Forecasting the future of the field is a task best left to prophets. Yet promising areas include the study of non-Western religions, religious militancy, religion and demography, the relationship between religious decline and the growth of the welfare state, and the role of religion in the formation of preferences and social capital. Insights from experimental economics, behavioural economics, game theory, industrial organization, and the economics of information and uncertainty have scarcely been explored. And if the past is any indication of the future, economists still have much to learn from religious historians, sociologists, anthropologists, and other scholars after 200 years of wandering in the secular wilderness.

See Also

- ► Smith, Adam (1723–1790)
- ► Social Capital

Bibliography

Azzi, C., and R. Ehrenberg. 1975. Household allocation of time and church attendance. *Journal of Political Economy* 84: 27–56.

Barro, R.J., and R.M. McCleary. 2003. Religion and economic growth across countries. *American Sociological Review* 68: 760.

Becker, G.S. 1964. *Human capital: A theoretical and empirical analysis*. 1st ed. New York: Columbia University Press for the NBER.

Becker, G.S. 1996. *Accounting for tastes*. Cambridge, MA: Harvard University Press.

Becker, G.S., and K.M. Murphy. 2000. *Social economics: Market behavior in a social environment*. Cambridge, MA/London: Belknap Press of Harvard University Press.

Berman, E. 2000. Sect, subsidy and sacrifice: An economist's view of Ultra-Orthodox Jews. *Quarterly Journal of Economics* 115: 905–953.

Berman, E. 2003. *Hamas, Taliban and the Jewish underground: An economist's view of radical religious militias*, Working paper No. 10004. Cambridge, MA: NBER.

Berman, E., and D.D. Laitin. 2005. *Hard targets: Theory and evidence on suicide attacks*, Working Paper No. 11740. Cambridge, MA: NBER.

Berman, E. and Stepanyan, A. 2003. Fertility and education in radical Islamic sects: Evidence from Asia and Africa. Mimeo, UC San Diego.

Boulding, K.E. 1970. *Beyond economics: essays on society, religion, and economics*. Ann Arbor: Ann Arbor Paperbacks.

Bruce, S. 1999. *Choice and religion: A critique of rational choice theory*. Oxford/New York: Oxford University Press.

Chen, D. 2004. *Club goods and group identity: Evidence from Islamic resurgence during the Indonesian financial crisis*. Working paper, University of Chicago.

Chesnut, R.A. 2003. *Competitive spirits: Latin America's new religious economy*. New York: Oxford University Press.

Chiswick, B.R. 1983. The earnings and human capital of American Jews. *Journal of Human Resources* 18: 313–335.

Coleman, J.S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94-(Supplement):S95–S120.

Delacroix, J., and F. Nielsen. 2001. The beloved myth: Protestantism and the rise of capitalism in nineteenth century Europe. *Social Forces* 80: 509–553.

Durkheim, E. 1915. *The elementary forms of the religious life*. Trans. K.E. Fields. New York: The Free Press, 1995.

Ehrenberg, R.G. 1977. Household allocation of time and religiosity: Replication and extension. *Journal of Political Economy* 85: 415–423.

Ekelund, R.B., R.F. Hébert, and R.D. Tollison. 2006. *The marketplace of Christianity*. Cambridge, MA: MIT Press.

Ekelund, R.B., R.F. Hébert, R.D. Tollison, G.M. Anderson, and A.B.. Davidson. 1996. *Sacred trust: The medieval Church as an economic firm*. New York: Oxford University Press.

El-Gamal, M.A. 2006. *Islamic finance: Law, economics and practice*. Cambridge: Cambridge University Press.

Finke, R., A.M. Guest, and R. Stark. 1996. Mobilizing religious markets: Pluralism and religious participation in the empire state, 1850–1865. *American Sociological Review* 61: 203–218.

Finke, R., and R. Stark. 1988. Religious economies and sacred canopies: Religious mobilization in American cities, 1906. *American Sociological Review* 53: 41–49.

Finke, R., and R. Stark. 1992. *The churching of America, 1776–1990: Winners and losers in our religious economy*. New Brunswick: Rutgers University Press.

Freeman, R.B. 1986. Who escapes? The relation of church-going and other background factors to the socio-economic performance of black male youths from inner-city poverty tracts. In *The black Youth employment crisis*, ed. R.B. Freeman and H.J. Holzer. Chicago: University of Chicago Press.

Gill, A. 1998. *Rendering unto Caesar: The Catholic Church and the State in Latin America*. Chicago: University of Chicago Press.

Glaeser, E.L., B.A. Ward, E.L. Glaeser, and B.A. Ward. 2006. Myths and realities of American political geography. *Journal of Economic Perspectives* 20(2): 119–144.

Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* 83: 1420–1443.

Green, E.C. 2003. *Rethinking aids prevention: Learning from successes in developing Countries*. Westport: Praeger.

Gruber, J. 2004. *Pay or pray? The impact of charitable subsidies on religiosity*, Working Paper No. 10374. Cambridge, MA: NBER.

Guiso, L., P. Sapienza, and L. Zingales. 2003. People's opium? Religion and economic attitudes. *Journal of Monetary Economics* 50: 225–282.

Hayek, F.A. 1988. *The fatal conceit: The errors of socialism*. Chicago: University of Chicago Press.

Heaton, P. 2006. Does religion really reduce crime? *Journal of Law and Economics* 49: 147–172.

Hungerman, D. 2005. Are church and state substitutes? Evidence from the 1996 Welfare Reform. *Journal of Public Economics* 89: 2245–2267.

Iannaccone, L.R. 1984. Consumption capital and habit formation with an application to religious participation. Ph.D. thesis, University of Chicago.

Iannaccone, L.R. 1988. A formal model of church and sect. *American Journal of Sociology* 9 (Supplement): 241–68.

Iannaccone, L.R. 1990. Religious participation: A human capital approach. *Journal for the Scientific Study of Religion* 29: 297–314.

Iannaccone, L.R. 1991. The consequences of religious market regulation: Adam Smith and the economics of religion. *Rationality and Society* 3: 156–177.

Iannaccone, L.R. 1992. Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy* 100: 271–292.

Iannaccone, L.R. 1995. Risk, rationality, and religious portfolios. *Economic Inquiry* 38: 285–295.

Iannaccone, L.R. 1998. An introduction to the economics of religion. *Journal of Economic Literature* 36: 1465–1495.

Iannaccone, L.R., and E. Berman. 2006. Religious extremism: The good, the bad, and the deadly. *Public Choice* 128: 109–129.

Iyer, S. 2002. *Demography and religion in India*. Oxford/New York: Oxford University Press.

Koenig, H.G., M.E. McCullough, and D.B. Larson. 2001. *Handbook of religion and health*. Oxford/New York: Oxford University Press.

Kuran, T. 1995. *Private truths, public lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.

Kuran, T. 2004. *Islam and mammon: The economic predicaments of islamism*. Princeton: Princeton University Press.

Lehrer, E.L., and C.U. Chiswick. 1993. Religion as a determinant of marital stability. *Demography* 30: 385–404.

Luidens, D., and R. Nemeth. 1994. Congregational vs. denominational giving: An analysis of giving patterns in the Presbyterian church in the United States and the reformed church in America. *Review of Religious Research* 36: 111–122.

Neuman, S. 1986. Religious observance within a human capital framework: Theory and application. *Applied Economics* 18: 1193–1202.

North, C.M., and C.R. Gwin. 2004. Religious freedom and the unintended consequences of the establishment of religion. *Southern Economic Journal* 71: 103–117.

Olds, K. 1994. Privatizing the church: Disestablishment in Connecticut and Massachusetts. *Journal of Political Economy* 102: 277–297.

Oslington, P., ed. 2003. *Economics and religion*. Cheltenham: Edward Elgar.

Putnam, R.D. 2000. *Bowling alone: The collapse and revival of American community.* New York: Simon and Schuster.

Richardson, G. 2005. Craft guilds and Christianity in late-medieval England: A rational-choice analysis. *Rationality and Society* 17: 139–189.

Samuelsson, K. 1993. *Religion and economic action: The Protestant ethic, the rise of capitalism, and the abuses of scholarship*. Toronto: University of Toronto Press.

Schultz, T.W. 1961. Investment in human capital. *American Economic Review* 51: 1–17.

Smith, A. 1759. In *The theory of moral sentiments*, ed. D.D. Raphael and A.L. Macfie, 1984. Indianapolis: Liberty Fund.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*, 1981. Indianapolis: Liberty Classics.

Sobel, J. 2002. Can we trust social capital? *Journal of Economic Literature* 40: 139–154.

Sosis, R.H., and B.J. Ruffle. 2003. Religious ritual and cooperation: Testing for a relationship on Israeli religious and secular kibbutzim. *Current Anthropology* 44: 713–722.

Stark, R. 2005. *The victory of reason: How Christianity led to freedom, capitalism, and western success*. New York: Random House.

Stark, R., and W.S. Bainbridge. 1985. *The future of religion*. Berkeley/Los Angeles: University of California Press.

Stark, R., and W.S. Bainbridge. 1987. *A theory of religion*. Bern: Peter Lang Publishing.

Stark, R., and R. Finke. 2000. *Acts of faith: Explaining the human side of religion*. Berkeley/Los Angeles: University of California Press.

Swedberg, R. 1990. *Economics and sociology: Redefining their boundaries: Conversations with economists and sociologists*. Princeton: Princeton University Press.

Tawney, R.H. 1926. *Religion and the rise of capitalism*. Reprinted with new introduction by Adam B. Seligman. New Brunswick/London: Transaction, 1998.

Ulbrich, H., and M. Wallace. 1983. Church attendance, age, and belief in the afterlife: Some additional evidence. *Atlantic Economic Journal* 11: 44–51.

Voas, D., D.V.A. Olson, and A. Crockett. 2002. Religious pluralism and participation: Why previous research is wrong. *American Sociological Review* 67: 212–230.

Waite, L.J., and E.L. Lehrer. 2003. The benefits from marriage and religion in the United States: A comparative analysis. *Population and Development Review* 29: 255–275.

Warner, R.S. 1993. Work in progress toward a new paradigm in the sociology of religion. *American Journal of Sociology* 98: 1044–1093.

Weber, M. 1920. *The protestant ethic and the spirit of capitalism*. Trans. S. Kalberg, 3rd ed. Los Angeles: Roxbury, 2002.

Weber, M. 1963. *The sociology of religion*. Trans. E. Fischoff. Boston: Beacon Press, 2002.

Young, L.A. 1997. *Rational choice theory and religion: Summary and assessment*. New York: Routledge.

Renewable Resources

Colin W. Clark

A resource stock may be termed ‘renewable’ if constant periodic removals from it can be indefinitely prolonged. A renewable resource may be further classified as depletable or nondepletable, according to whether or not its productivity is affected by the level of exploitation. Biological resources such as fish, bird and animal populations, forests, grasslands and agricultural soils, provide examples of the depletable type, while surface water resources, solar and geothermal energy may be classified as nondepletable.

Economic Analysis

In spite of the absolute dependence of all economic systems upon renewable resources, no detailed economic analysis of the economics of renewal resources as such was attempted until the mid-20th century. Renewable resources were simply subsumed under the concept of economic rent of land, defined by Ricardo as ‘that portion of the produce of the earth which is paid to the landlord for the use of the original and indestructible powers of the soil’ (Ricardo 1817). But expansion of human populations and technological development throughout the 19th century gradually resulted in the depletion, sometimes to the point of extinction, of once superabundant renewable resource stocks. (A famous example, the passenger pigeon of Eastern North America, once the New World’s most abundant bird species and a resource of economic significance in colonial America, had passed into extinction by 1914.) Such development made it clear that the original powers of the soil were in fact far from indestructible. Popular concern with resource issues led to the ‘conservation movement’ of the early 20th century, resulting in legislation devoted toward the preservation of agricultural, forest and wildlife resources.

Theoretical analysis of the role of renewable resources in economics was hindered by the inevitable temporal dimension of resource exploitation, necessitating the use of dynamic models and the calculus of variations (see Hotelling 1931). Works devoted to verbal analysis of ‘the economics of conservation’, such as those of Ciriacy-Wantrup (1952) and Scott (1955), set the stage for subsequent comprehensive theoretical treatment of resource economics by variational techniques. Finally by the 1970s, a major expansion of public interest in the ‘environment’, and in the ‘limits to growth’ (Meadows et al. 1972), combined with such resource-associated events as the OPEC cartelization of petroleum production and the collapse of major marine fisheries, led the economics profession to take a serious interest in resource and environmental issues – if merely in some instances to defuse the public hysteria. Theoretical developments in constrained dynamic optimization had meanwhile greatly improved the requisite mathematical techniques.

A generalized model of resource exploitation by private or public resource owners may be expressed as follows:

$$\frac{dx}{dt} = G(x) - h(t), \quad t \geqslant 0 \tag{1}$$

$$x(0) = x_0, \quad x(t) \geqslant 0, \quad h(t) \geqslant 0 \tag{2}$$

$$\pi = \pi[x(t), h(t), t] \tag{3}$$

$$V(x_0) = \max_{[h(t)]} \int_0^\infty \alpha(t)\pi \, dt \tag{4}$$

in which $x(t)$ denotes the size (‘state’) of the resource stock at time t, $G(x)$ is the natural rate of replenishment, $h(t)$ is the rate of removals, or ‘harvest’, of the resource, π denotes the net flow of economic benefits at time t and $V(x_0)$ is the optimized present value of net benefits, relative to the discount factor

$$\alpha(t) = \exp\left[-\int_0^t r(s) ds\right] \tag{5}$$

where $r(s)$ is the instantaneous rate of discount at time s.

The specification $G(x) \equiv 0$ provides a general exhaustible resource model. A renewable resource model is obtained by allowing G to depend on the resource stock x, with $G(x) = 0$ for some $\bar{x} > 0$ and $G(x) > 0$ for $0 < x < \bar{x}$. For the case of a biological resource stock one would assume that $G(0) = 0$: a nondepletable resource could be modelled (not very well) by assuming $G(0) > 0$. For the latter two cases, $\bar{x}$ represents the natural, or environmental 'carrying capacity' for the given stock.

A popular, widely accepted objective of renewable resource management is the so-called 'maximum sustained yield' (MSY), characterized simply by the equation

$$h_{\mathrm{MSY}} = \max_x G(x). \tag{6}$$

According to this principle, any renewable resource stock should be maintained at the level $x = x_{\mathrm{MSY}}$ at which its exploitable productivity $G(x)$ is a maximum. Perpetuation of MSY has indeed been considered as the sacred and sole trust of many resource management agencies, seldom with any cognizance of the economic implications of such a policy. Not infrequently the resource industry itself seems to exhibit a preference for some quite different objective.

The solution of the optimization problem of equation (4) is characterized by the following necessary conditions:

$$\pi_h = \lambda(t) \tag{7}$$

$$G_x + \frac{\pi_x}{\pi_h} = r(t) - \frac{\dot{\lambda}(t)}{\lambda(t)} \tag{8}$$

where $\lambda(t)$ denotes the current 'shadow price' (formerly, user cost) of the resource stock, equal to the marginal value of the resources stock $x(t)$, and where subscripts designate partial derivatives and overdot the time derivative.

If π does not depend explicitly on time t, and if $r(t) = r$ is constant, equation (8) possesses an equilibrium solution $x = x^*$ determined by

$$G_x + \pi_x/\pi_h = r \tag{9}$$

$$h = G(x). \tag{10}$$

Equation (9) is recognizable as the standard marginal productivity rule of optimal capital accumulation, in which marginal productivity G_x is equated to the discount rate r. The correction term π_x/π_h arises from the fact that x and $G(x)$ are specified in physical units, rather than as asset and flow values, respectively (see Clark 1976, ch. 3).

In the event that $\pi_x \equiv 0$ (costs and benefits independent of stock level x) and $r = 0$, equation (9) becomes $G_x = 0$, namely $x = x_{\mathrm{MSY}}$, the MSY solution. In general, under the reasonable assumptions that $G_{xx} < 0$ and $\pi_x > 0$, $\pi_h > 0$, we see from equation (9) that (i) discounting tends to decrease x^*, whereas (ii) the dependence of π on x tends to increase x^*, relative to the MSY solution. The numerical significance of these effects can only be assessed by estimating model parameters for particular cases.

The Effects of Discounting

A theme that runs through much of the conservation literature pertains to the effect of discounting, or time-preference, on the conservation of resource stocks. Pigou, for example, says that

> There is widespread agreement that the state should protect the interests of the future in some degree against the effects of irrational discounting, and of our preference for ourselves over our descendants. The whole movement for 'conservation' in the United States is based upon this conviction. (Pigou 1920, p. 29, as quoted by Scott, 1955).

This raises the question of whether the social rate of time preference differs, or ought to differ, from the market rate.

For the case of renewable resources, equation (9) suggests that the effect of discounting will be large when G_x is small. Since G_x represents marginal growth rate of the resource stock, we conclude that discounting will be especially important for resource stocks having low growth rates (although the effect of π_x/π_h must also be considered).

Biological resources exhibit a wide range of growth rates. Some species, called '*r*-selected' by ecologists, are highly fecund: populations consisting of numerous small individuals expand rapidly to take advantage of environmental opportunities. At the opposite end of the growth spectrum are large, slow growing '*K*-selected' species, which are also often highly valued and easily exploited by modern techniques. The latter type, which includes whales and other marine mammals, forests, desert grazing lands, and the like, are particularly subject to severe over-exploitation, both under common-property conditions of exploitation and under private profit maximization by firms employing market rates of discount.

User Conflicts

The paradigm, assumed above, of an isolated renewable resource stock exploited by a sole owner is seriously unrealistic for most actual renewable resource industries. Imperfection of ownership rights, multiple uses and users, and a wide variety of externalities, are the rule rather than the exception in fisheries, forestry, wildlife, water resources, and such like.

Most commercial fisheries, for example, are still exploited under common-property conditions, although the introduction of 200-mile fishing zones in the late 1970s has at least placed the majority of marine fishery resources under national jurisdiction. Wildlife, water, and recreational resources are often also utilized as common-property resources. The historical trend, however, is towards progressive allocation of resource-use rights to individuals, if not via outright sole ownership than vai user permits, quotas and fees. As a general rule, the delineation of resource sub-allocations can be expected to increase along with the economic importance of the resource, but the process is inevitably confounded by political and legislative components.

Over-exploitation of renewable resource stocks and over-expansion of harvest capacity are frequent occurrences in resource industries utilizing non-owned, or commonly owned resources. In the absence of property rights, each exploiter tends to ignore the effects that his own removals will have on the total resource stock and its future production. Thus in a common-property resource industry, the rate of harvest will increase (unless restricted by regulation) to a level at which the marginal exploiter receives zero net revenue:

$$\pi_h = 0. \tag{11}$$

The industry thus behaves as if the shadow price λ of the resource were zero [see equation (7)]. By imposing a removals tax $\tau = \lambda$ per unit of harvested resource, the management authority (should one exist) can in principle force the competitive exploiting industry to utilize the socially optimal rate of exploitation. Resource rents then accrue to the management authority, rather than being dissipated through over-exploitation or over-capacity.

Renewable resource industries impose a variety of externalities upon other resource users. Logging of forests may affect surface water retention and flow. Public demand for parks and wilderness areas may lead to conflicts with resource industries such as forestry, agriculture and hydro-electric power. Pesticides employed to protect forests or crops may damage fish and wildlife populations.

An important long-term externality resulting from the alteration or destruction of natural habitats by resource industries is the progressive loss of genetic material, which may ultimately limit the diversity of domestic crops and animals, and reduce the supply of naturally derived pharmaceutical and industrial compounds (Oldfield 1984).

Most externalities of his kind increase in economic importance with the intensity of resource exploitation. Consequently the socially optimal exploitation policy will often involve less intensive exploitation than would be practised by private resource owners. Much of the rationale for the establishment of government management authorities doubtlessly stems from these considerations. The fact that the external costs of resource exploitation are often much longer lasting than the internal benefits adds to the need for timely government regulation.

Resource Management

Numerous management agencies have been established to regulate the exploitation of renewable resources such as water supplies, marine and freshwater fish stocks, wildlife populations and forests. Such agencies face many difficulties, including particularly the allocation of a limited supply in the presence of excessive demand, enforcement of regulations, and the problem of dealing with major uncertainties regarding resource inventories, ecosystem dynamics and environmental factors. It is also becoming increasingly recognized that the traditional resource management objective of maximum sustained yield is often not adequate to deal with resource conflicts, multiple uses and externalities of resource use.

Both fiscal and quantitative instruments are frequently employed by resource management agencies. Fees and taxes levied on resource users reduce excess demand for the resource, while collecting resource rent for the public purse. In many localities where resource-based industries dominate the economy, such charges can constitute a major component of state revenue, although a dominant resource industry may have sufficient political influence to prevent the full capture of rents by government.

The degree to which resource taxes can be used in practice as proxies for shadow prices is severely limited by the complexity and uncertainty of both biological and economic systems. Consequently direct regulation is the usual rule, at least for resource industries based on publicly owned resource stocks. Regulation may ultimately pertain to almost every aspect of exploitation, including time, place, amount and methods of harvest, as well as details of species, size, sex etc. permitted to be taken.

It might of course be argued that the need for such complex systems of regulation would be obtained if ownership were to be transferred entirely to private hands, but this may be neither feasible nor desirable in cases where resource stocks are not readily appropriated, or where significant externalities must be controlled. But it is certainly true that regulations can have perverse economic consequences. An example common by the 1970s was the tendency to regulate commercial fisheries by means of total annual catch quotas. Such non-allocated quotas force individual fishermen into a competitive 'scramble' wherein each attempts to catch as many fish as possible prior to the closure of the fishery. The consequences include unnecessary expansion of fishing capacity in terms of number, size and horsepower of vessels, reduction in the quality of fish and highly uneven rates of delivery of fish to processors and markets.

A potential method for overcoming these problems is the use of allocated quotas: if such quotas are transferable, the price of quotas will play a similar role to a tax on catches. Quota allocations are also of potential value for the regulation of other resources such as water resources and public grazing lands. Monitoring and enforcement are of course essential to the success of any allocated quota system. The quotas must also be flexible, to allow for natural fluctuations in resource abundance.

Fluctuations and Uncertainties

Renewable resource industries face significant uncertainties regarding both supply and demand. Unpredictable environmental fluctuations can have large-scale effects on the production and availability of renewable resource stocks, which can in some cases have nationwide or worldwide economic consequences. Unexpected decreases in resource abundance become especially serious when exploitation has reached high levels, with industries or even entire segments of the economy dependent upon the resource. While developed nations may possess institutions to mollify the worst effects of such natural fluctuations, the less developed nations often face economic disaster in times of drought, flood, insect or crop pathogen plagues, or fishery collapses.

Temporary periods of low resource availability can be extremely unpleasant in themselves. But they can also result in severe over-exploitation as the dependent industry continues to harvest the resource in desperation. In extreme cases ultimate recovery may become impossible owing to irreversible destruction of breeding stocks, or of soil productivity. The more infrequent are the bad years, the more likely ultimate disaster may become, as communities grow to rely upon the resource and discount the possibility of a decline.

Resource managers face many kinds of uncertainty beyond that pertaining to the scope and timing of natural fluctuations (Mangel 1985). The long-term response of depletable resource stocks to exploitation is often difficult if not impossible to predict quantitatively. Even current inventories of resource stocks such as marine populations may be highly uncertain – current estimates of whale stocks in the Antarctic, for example, range over two orders of magnitude. Discerning trends from such inaccurate data often borders on the impossible, but improving the accuracy of the data base is often unacceptably expensive.

In response to such gross levels of uncertainty the risk-averse public resource manager tends to prefer a conservative exploitation policy which minimizes the probability of depletion.

The exploiting industry, however, often takes the opposite view, preferring certain current revenues to uncertain future benefits. Since uncertainty increases with the planning horizon, an additional bias towards depletion of renewable resource stocks is observed.

See Also

▶ Bioeconomics
▶ Exhaustible Resources
▶ Fisheries
▶ Forests

Bibliography

Ciriacy-Wantrup, S.V. 1952. *Resource conservation: Economics and policies*. Berkeley: University of California Press. 3rd ed., 1968.

Clark, C.W. 1976. *Mathematical bioeconomics: The optimal management of renewable resources*. New York: Wiley–Interscience.

Clark, C.W. 1985. *Bioeconomic modelling and fisheries management*. New York: Wiley–Interscience.

Duerr, W.A. 1979. *Forest resource management*. Philadelphia: Saunders.

Hotelling, H. 1931. The economics of exhaustible resources. *Journal of Political Economy* 39: 137–175.

Mangel, M. 1985. *Decision and control in uncertain resource systems*. New York: Academic.

Meadows, D.H., D.L. Meadows, J. Randers, and W.W. Behrens. 1972. *The limits to growth*. New York: Universe Books.

Oldfield, M.L. 1984. *The value of conserving genetic resources*. Washington, DC: US Department of the Interior.

Pigou, A.C. 1920. *The economics of welfare*. London: Macmillan.

Ricardo, D. 1817. In *Principles of political economy and taxation*, ed. R.M. Hartwell. Harmondsworth: Penguin.

Scott, A.D. 1955. *Natural resources: The economics of conservation*. Toronto: University of Toronto Press. 2nd ed., Toronto: McClelland & Stewart, 1973.

Rent

Armen A. Alchian

Abstract

'Rent' is the payment for use of a resource, whether it be land, labour, equipment, ideas, or even money. The term is often restricted to payment for use of land or equipment. 'Economic rent' is payment for use of any resource whose supply is fixed. Rent serves a social purpose because market levels of rent indicate which uses of fixed resources are the highest valued, and direct such resources to those uses. 'Monopoly rent' is paid to producers in markets that are artificially restricted; it may be dissipated by 'rent seekers' who compete for monopoly status.

Keywords

Composite quasi-rent; Differential rents; Economic rent; Land; Land tax; Marshall, A.; Monopoly rents; Producer surplus; Quasi-rent; Rent; Rent seeking; Ricardian rents; Ricardo, D

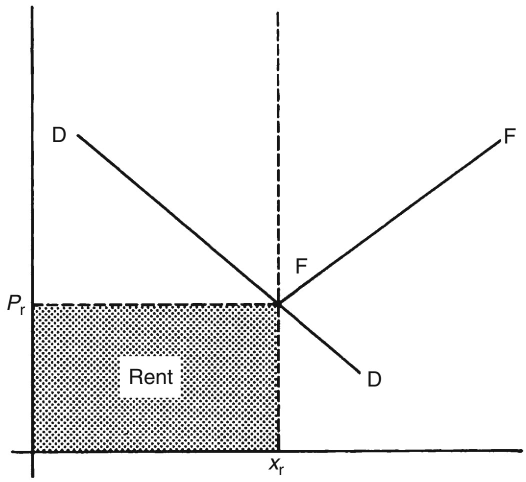
JEL Classifications
A1

'Rent' is the payment for use of a resource, whether it be land, labour, equipment, ideas, or even money. Typically the rent for labour is called 'wages'; the payment for land and equipment is often called 'rent'; the payment for use of an idea is called a 'royalty'; and the payment for use of money is called 'interest'. In economic theory, the payment for a resource where the availability of the resource is insensitive to the size of the payment received for its use is named 'economic rent' or 'quasi-rent' depending on whether the insensitivity to price is permanent or temporary.

To early economists, 'rent' meant payments for use of land; Ricardo, in particular, called it the payment for the 'uses of the original and indestructible powers of the soil' (Ricardo 1821, p. 33). Subsequently, in recognition that a distinctive feature of what was called 'land' was its presumed indestructibility (i.e. insensitivity of amount supplied to its price), the adjective 'economic' was applied to the word 'rent' for any resource the supply of which is indestructible (maintainable for ever at no cost) and non-augmentable, and hence invariant to its price. In the jargon of economics, the quantity of present and future available supply is completely inelastic with respect to price, a situation graphically represented by a vertical supply line in the usual 'Marshallian' price-quantity graphs.

Economic Rent

The concept of 'economic rent' is graphically depicted by the standard demand and supply lines in Fig. 1 with a vertical supply curve (quantity supplied invariant to price) at the amount X_r. At all prices the supply is constant. The entire return to the resource is an 'economic rent'. If the aggregate quantity of such resources may in the future be increased by production of more indestructible units of the resource in response to a higher price (but the amount available at any moment is fixed regardless of the rent for its services), the supply line at the current moment is vertical. The supply curve for future amounts slopes upward from the existing amount, as depicted by the line FF in Fig. 1. The long run rent would be P_r and the equilibrium stock would be X_r: at that equilibrium stock the 'market supply' (in Marshall's terminology) would be a vertical line. Thus, the supply of indestructible units would have depended on past anticipated prices about the present prices, but the supply of current units would be insensitive to the current price or rent. The return could be called 'economic rent', except that no convention has been developed with respect to the terminology for this situation of indestructible but augmentable resources.



Rent, Fig. 1

Quasi-rent

Closely related to 'economic rent' is 'quasi-rent', a term apparently initiated by Alfred Marshall (Marshall 1920, pp. 74, 424–6). Because virtually every existing resource is unresponsive to a change in price for at least some very small length of time, the return to every resource is like an 'economic rent' for at least a short interval of time. In time, the supplied amount will be altered, either by production or non-replacement of current items. Yet, the fact that the amount available is not instantly affected by price led to the term 'quasi-rent', which denotes a return, variations in which do not affect the current amount supplied to the demander but do affect the supply in the future.

If a rental (payments) stream to an existing resource is not sufficient to recover the costs incurred in its production the durability of that existing resource will nevertheless enable the resource to continue to provide services, at least for some limited time. In other words, because of the resource's durability it will continue for some interval to yield services even at a rent insufficient to recover its cost of production, but sufficient for current costs of use including interest on its salvage value (which is its highest value in some other use). Any excess over those current costs is a 'quasi-rent'.

Quasi-rent resembles an 'economic rent' in that it exceeds the amount required for its current use, albeit temporarily – except that a flow of rents that did not cover all 'quasi-rent' would preserve it for only a finite future interval, after which the resource would be diminished until not worth more than its salvage value. If the resource received a payment exceeding all the initially anticipated and the realized costs of production and operation, it will have achieved a profit, that is, more than pure interest on the resource's investment cost. The question exists as to whether 'quasi-rent' means just that portion of the rent in excess of the minimum operating costs over the remaining life of the asset, or all the excess, including profits, if any. Convention seems still to be missing. Marshall seems to have excluded interest on the investment as well as any profits from what he called quasi-rents, so that any excess over variable costs of operation were partitioned into quasi-rents, interest on investment and profits (Marshall 1920, pp. 412, 421, 622).

Composite Quasi-rent

'Composite quasi-rent' was another important, but subsequently ignored, concept coined by Marshall (Marshall 1920, p. 626). When two separately owned resources are so specific to each other that their joint rent exceeds the sum of what each could receive if not used together, then that joint rent to the pair was called 'composite quasi-rent'. The two resources presumably already had been made specific to each other (worth more together than separately) by some specializing interrelated investments. Marshall cited the example of a mill and a water power site, presumably a mill built next to a dam to serve the mill, each possibly separately owned. One or both of the parties could attempt to hold up or extract a portion of the other party's expropriable quasi-rent. It is interesting to quote Marshall about this situation:

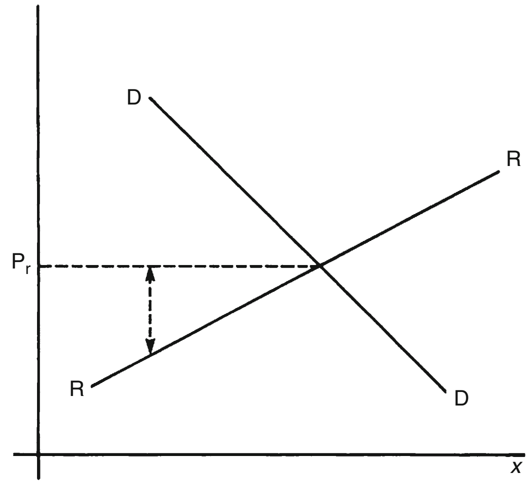
> The mill would probably not be put up till an agreement had been made for the supply of water power for a term of years; but at the end of that term similar difficulties would arise as to the division of the aggregate producer's surplus afforded by the water power and the site with the mill on it. For instance, at Pittsburg when manufacturers had just put up furnaces to be worked by natural gas instead of coal, the price of the gas was suddenly doubled. And the history of mines affords many instances of difficulties of this kind with neighbouring landowners as to rights of way, etc., and with the owners of neighbouring cottages, railways and docks. (Marshall 1920, p. 454)

A reason for attributing importance to the concept of 'composite quasi-rent' is now apparent. If it arises with resources that have been made specific to each other in the sense that the service value of each depends on the other's presence, the joint value of composite quasi-rent might become the object of attempted expropriation by one of the parties, especially by the one owning the resource with controllable flow of high alternative use value. To avoid or reduce the possibility of this behaviour, a variety of preventative arrangements, contractual or otherwise, can be used prior to making the investments in resources of which at least one will become specific to the other. These include, among a host of possibilities: joint ownership, creation of a firm to own both, hostages and bonding, reciprocal dealing, governmental regulation, and use of insurers to monitor uses of interspecific assets. This is not the place to discuss these arrangements, beyond asserting that without the concept of 'quasi-rent' and especially 'expropriable quasi-rent' – which Marshall called 'composite quasirent' – a vast variety of institutional arrangements would otherwise be inexplicable as a means of increasing the effectiveness of economic activity.

Though Marshall briefly mentioned similar problems between employers and employees, *I* have not found any subsequent exposition by him about the precautionary contractual arrangements and institutions that attempt to avoid this problem, which has become a focus of substantial important research on what is called, variously, 'opportunism, shirking, expropriable quasi-rents, principal–agent conflicts, monitoring, problems of measuring performance, asymmetric information, etc.'.

Ricardian Rent

The rents accruing to different units of some otherwise homogeneous resource may differ and result in differences of rent over the next most valued use, differences that are called 'Ricardian rents'. This occurs where the individual units, all regarded as of the same 'type' in other uses, are actually different with respect to some significant factor for its use *here,* though this factor, which is pertinent *here,* is irrelevant in any other uses. Examples of such factors can be location, special fertility, or talent that is disregarded in the other potential uses. For some questions, the inaccurate 'homogenization' can be a convenient simplification, but for explaining each unit's actual rents, it can lead to confusion and misunderstanding. The service value, hence rents, for the use of the services here may differ, though equal in every relevant respect elsewhere. Whether the specific use uniqueness is created by natural talent or sheer accident, the special differences in use value here imply differences in payments, often called 'Ricardian rents' to distinguish them from differences in rents (prices) obtained because of monopolizing or unnatural restrictions on any potential competitors, which may lead to higher rents, called 'monopoly rents' for the protected resources.



Rent, Fig. 2

Differential Rents

'Differential rents' are another category representing rent differences in a sort of reverse homogeneity. Units of resource that are equal with respect to their value in use *here* differ among themselves in their values of use elsewhere. This can be represented graphically as in Fig. 2. The differential rents of successive units are represented by the differences between the price line and the curve RR, which arrays the units from those with the lowest alternative use values to the highest, a curve labelled RR. The arrayed units are not homogeneous for uses elsewhere, so even if identical for use here, calling them successive units of the same good is misleading. They are not totally homogeneous; if they were, each unit would have the same as any other unit's use value and rent elsewhere. A curve like RR is equivalent to Marshall's particular expenses curve, which arrayed units according to each individual unit's cost of production, or use value elsewhere, from lowest to highest (Marshall 1920, p. 810n). The difference between price or rent here and the value on the RR curve is called 'producers' surplus' or 'differential rent'. In sum, 'Ricardian rents' indicate differences in rents to units that are equal in their best alternative use values, but different in their rent value here, while 'differential rents' are the premia to units that are the same value here but different in their best alternative use values.

It is worth digressing to note that an upward rising true supply curve, which reflects increasing marginal costs of production, is different from the RR curve. In the true supply curve the area between the supply curve and the price line does not represent any of the above mentioned rents nor 'producers' surplus' (as it does with the RR curve). It is the portion of earnings of the supplier that exceed the variable costs and are applicable to cover the costs (possibly past investment costs) that are invariant to the rate of output. That area does not represent any excess of rental or sale value of units produced over their full costs, since only the variable costs are under the marginal cost curve. It represents the classic distribution of income to capital, if, for example, labour is presumed to be a variable input and capital a fixed input.

High Rents a Result, Not Cause, of High Prices

An earlier unfortunate analytic confusion occurred in the common misimpression that high rents of land made its products more expensive. Thus the high rent of land in New York was and is still often believed to make the cost of living, or the cost of doing business, higher in New York. Or higher rent for some agricultural land is believed to increase the cost of growing corn on that land. Proper attention to the meaning of 'demand' and 'costs' would have helped avoid that confusion. Demand here for some unit for resource is the highest value use of that resource if used *here.* The cost of using it here is the highest valued forsaken alternative act elsewhere. For any resource the cost of its use *here* is its best value elsewhere, that is, its demand elsewhere. Land rent is high for 'this' use because the land's value in some other use is high. The reason the rent is high here and can be paid is that its use value *here* is bid by competitors for its use here into the offered rent and exceeds the value in some other use. The product of the land can get higher price here; that is why the rent is bid up so high, even though the particular winning bidder then believes a high price of the products must be obtained because the rent was high, rather than the reverse. As with every marketable resource, its highest value use here determines its rent, rather than the reverse. It was the implication of this kind of analysis that Marshall attempted to summarize in the famous aphorism, which he attributed to Ricardo (1817): 'Rent does not enter into [Money] Cost of production' (Marshall 1890, p. 482).

Probably the source of the confusion in believing that high rents of land caused high prices for products produced on expensive land is that an individual user of that expensive resource has to be able to charge a higher price for the product, if the rent is to be covered. Bidders for that land compete for the right to the land that can yield a service worth so much – though to any individual successful bidder that rent has to be paid regardless of how well the successful bidder may be at actually achieving the highest valued use of the land. Hence it may appear to an individual bidder that the rent determines the price that must be charged, rather than, as is the correct interpretation, the achievable high valued use enables the high bid for the land for the person best able to detect and achieve that highest valued use.

Function of Rent

Some people were aware of this bidding for the 'land' and concluded that the rent served no social purpose, since the land would exist anyway. But the high receipt resulting from competitive bidding for its uses serves a useful purpose. It reveals which uses are the highest valued and directs the land to that use. In principle, a 100 per cent tax on the land rent would not alter its supply (assuming initially that 'land' is the name of whatever has a fixed indestructible supply). This would be correct if in this case the 'owner' of the land had any incentive left to heed the highest bidder where the highest bid determines the rent. The assertion assumes that somehow the highest valued use can be known and that amount of tax be levied without genuine bona fide competitive bids for its use, a dubious if not plainly false proposition.

Monopoly Rent

Let the word 'monopoly' denote any seller whose wealth potential is increased by restrictions on other potential competitors, restrictions that are artificial or contrived in not being naturally inevitable. Laws prohibiting others from selling white wine, or opening restaurants, or engaging in legal practice are examples. It should be immediately emphasized that this does not imply nor is it to be inferred that all such restrictions are demonstrably undesirable. Nevertheless, the increased wealth potential is a 'monopoly rent'. Whether it is realized by the monopolist as an increase in wealth depends upon the costs of competing for the imposition of such restrictions. Competition for 'monopoly rents' may transfer them to, for example, politicians who impose the restrictions, and in turn may be dissipated by competition among politicians seeking to be in a position to grant such favours. The 'monopoly rents' may be dissipated (by what is often called 'rent-seeking' competition for such monopoly status of rights to grant it) into competitive payments for resources that enable people to achieve status to grant such restrictions. Those who initially successfully and cheaply obtained such 'monopoly' status may obtain a wealth increase, just as successful innovators obtain a profit stream before it is eliminated by competition from would-be imitators.

See Also

- Malthus, Thomas Robert (1766–1834)
- Marshall, Alfred (1842–1924)
- Ricardo, David (1772–1823)
- Thünen, Johann Heinrich von (1783–1850)
- West, Edward (1782–1828)

Bibliography

Marshall, A. 1890. *Principles of economics*. 1st ed. London: Macmillan.
Marshall, A. 1920. *Principles of economics*. 8th ed. London: Macmillan. reprinted, 1946.
Ricardo, D. 1821. *Principles of political economy and taxation*. 3rd ed, 1965. London: Dent Dutton.

Rent Control

Stephen Malpezzi

Abstract

Rent control generically describes a range of regulations governing rents, as well as related contract features such as security of tenure and required maintenance. There is debate in the literature about the efficacy of controls based on (1) whether the housing market is best modelled as a competitive market, or one where landlords have market power; and (2) whether regulators have sufficient information and appropriate mechanisms to improve imperfect market outcomes. Many empirical studies find that rent controls score badly as redistributive systems. Many basic questions, especially regarding dynamic effects on the supply of housing, have yet to be credibly answered.

Keywords

Asymmetrical information; Housing markets; Housing supply; Loss aversion; Market power; Property taxation; Rent control; Tax incidence

JEL Classifications

R38

Rent controls of one kind or another affect roughly 40 per cent of the world's urban dwellers. Rent control is usually thought of as a policy applied to private markets, but publicly provided housing (for example, much urban housing in Russia and in China) is also subject to controls. In addition to regulations governing rents, controls often address additional contract features such as security of tenure and required maintenance. Actual rent control regimes vary enormously in their design and in their effects.

History

Rent controls are often instituted in response to a major economic or political shock which limits the responsiveness of the housing market. Controls were introduced in the Second World War in Europe, North America, and, under European colonial influence, much of the developing world as well. Most jurisdictions in the United States and Canada removed controls in the post-war years; however, controls of varying degrees of stringency were maintained in much of Europe and the developing world. Poorer countries tend to have more stringent regimes, though enforcement patterns vary at least as much as *de jure* codes.

Exactly why controls exist, or at least are retained after wartime or similar emergencies are clearly over, is still debated. An obvious point of political economy is that there are more tenants than landlords; but there is little correlation between the fraction of a country's population renting and the stringency of controls, according to Malpezzi and Ball (1991). On the other hand the relatively small number of US cities with rent control tend to have large renter populations, notably New York. Fischel (2001) presents several interesting conjectures about the political economy of controls, notably that homeowners might ally with landlords to oppose controls because they fear negative spillovers from reduced maintenance of stringently controlled buildings, as well as shifting property tax burdens. The strong opposition to relaxation of controls in New York, while nearby uncontrolled jurisdictions see little agitation for imposition, might be analysed in Kahneman and Tversky's (1979) loss aversion framework. A clear understanding of the political economy of controls awaits future research.

Features

One key feature is whether regulations set the level of rents, or control increases in rent. Others include how controlled rents are adjusted for changes in costs (with cost pass-through provisions, or adjustments for inflation); how close the adjustment is to changes in market conditions; how it is applied to different classes of units; or whether rents are effectively frozen over time. Other key provisions which vary from place to place include breadth of coverage, how initial rent levels are set, treatment of new construction, whether rents are reset for new tenants, and tenure security provisions. Rent control's effects can vary markedly depending on these specifics, and on market conditions, as well as enforcement practices.

Theory

Rent control can be analysed as an implicit tax on housing capital. In the simplest case, where imposition of controls reduces the price of an existing stock of rental housing, the tax is borne by landlords for the benefit of tenants. Over time, as the market adjusts to controls, the incidence of the 'tax' becomes more complicated.

Much of the debate in the literature about the efficacy of controls stems from maintained assumptions about the nature of the housing market, and the regulator, in turn. The first question is: is the housing market best modelled as a competitive market, or one where landlords have market power, for example from information asymmetries? If the former, then clearly rent control reduces the efficiency of the rental market, although the magnitude of such effects can be debated, and distributional arguments remain. If the latter, a second question readily follows: does the regulator have sufficient knowledge, and an appropriately designed set of regulations, to improve on the market outcome? Arnott (1995) ably reviews the contrast between competitive and 'market power' theoretical approaches, and also discusses why it is so difficult to resolve these issues empirically.

Whatever one's priors about market power, there are many alternative adjustment mechanisms which can arise in a notionally controlled market. Four of the adjustments can be embodied in rent control laws: (*a*) indexing (keeping real rents constant), (*b*) reassessment for new tenants, (*c*) differential pricing of new and existing units, and (*d*) differential pricing for upgraded units. Three are market responses which many would generally consider undesirable outcomes, namely, (*e*) outright evasion, (*f*) side payments such as key money, (*g*) adoption by tenants of maintenance expenditures, and (*h*) accelerated depreciation and abandonment, (*i*) distortions in consumption, not only in the composite housing services but also crowding, length of stay, mobility and tenure choice.

Key questions are: What are the efficiency losses from controls? Are the benefits to some tenants worth the costs? Do they redistribute income as intended? Several broad approaches have been taken in the empirical literature to answer these questions.

Static Analysis

One of the first published studies of the costs and benefits of rent control is Olsen's. Using data from New York City in 1968, Olsen (1972) found the average controlled rent for an apartment was $999 a year (for comparison, the average income was $6,229). Olsen first estimated how much the controlled units would rent for in the absence of controls. The average estimated uncontrolled rent for controlled units was $1,405, implying a subsidy (static cost to landlords) of $406. Olsen next estimated how much households in controlled units would spend in the uncontrolled market, given their income and family size. The average estimated market expenditure for the controlled households was $l,470, indicating that they consumed slightly less housing than they would have in the free market. Olsen then computed the economic benefit of rent control to each surveyed controlled tenant using a simple consumer surplus model. Olsen's estimate of the average net benefit is $213, little more than half the gross subsidy of $406.

Examining the distribution of these benefits among controlled households, Olsen found the annual benefit decreased by about one cent for every dollar of additional income, $9 a year of head's age, and $69 per additional household member. Rent control in New York City in 1968 appears to redistribute income, but very weakly, and in no way proportional to its cost.

A number of other studies have been carried out along these lines (Malpezzi and Ball 1991, review several). For example, in Cairo, Egypt, monthly rents for a typical unit are less than 40 per cent of estimated market rents. But 'key money' (illegal upfront payments to landlords) and other side payments make up about a third of the difference.

Dynamic Analysis

Murray et al. (1991) is an early study of rent control dynamics. A simulation model was used to predict the time path of rents and the quantity of housing services given alternative control regimes. The magnitude of the effects varied substantially with details of the regime. In general, Murray et al. find that dynamic losses can be substantial; in fact they outweigh static consumer's surplus losses by as much as a factor of 18. Generally tenant benefits are were substantially less than landlord costs; the transfer efficiency in three representative cases ranged from 65 per cent to 83 per cent.

Another potential dynamic effect of controls, with possible spillovers to labour markets, is reduced household mobility. Several studies, such as Munch and Svarer (2002) and Simmons-Moseley and Malpezzi (2006), find that household mobility is inversely related to the estimated net benefits received from a control regime.

Given their potential importance, dynamic effects of controls are understudied. For example, no one has yet credibly analysed the effects of controls on the aggregate supply of housing. Reviews of the theoretical literature by Arnott (1995), and of the empirical literature by Turner and Malpezzi (2003), point out that empirical work lags theory in this area. Malpezzi and Ball (1991) did find that countries with stricter rent control regimes invested less in housing, in the aggregate; but while the analysis accounted for income and demographics on the demand side, other potential constraints on housing supply (for example, land use constraints, financial constraints) were not well specified. Since these may well be correlated with the strength of controls, these results cannot be viewed as the final word. Given the myriad ways real world regimes work, and the variety of possible ways around controls (legal and illegal) the size of the net aggregate effect on supply remains unknown.

Distributional Issues

Such evidence as exists casts doubt on controls' effectiveness as income transfer mechanisms. In Kumasi and Rio, benefits were found to be somewhat 'progressive' in the common sense of the term (larger benefits to poorer households). On the other hand, in Cairo and Bangalore, no relationship was found between the benefits gained from reduced rent and household income, because rent control is not well targeted to low-income groups (Malpezzi and Ball 1991). In fact, research on New York controls by Glaeser and Luttmer (2003) suggests that previous research largely *underestimates* the misallocation of housing under controls, and that, because of excess demand for controlled units, benefits are more or less randomly distributed.

Another questionable assumption behind redistribution as a rationale for controls is the notion that landlords are rich and tenants are poor. In Cairo, Kumasi and Bangalore, the income of tenants and landlords was compared; and, while the landlords' median income was higher in all three, there was significant overlap. In Cairo, for example, about 25 per cent of tenants had incomes that were higher than the landlord median, and about 25 per cent of landlords had incomes lower than the tenant median. There is no guarantee the transfers will occur only from high-income landlords to low-income tenants.

Most careful empirical studies find that at least some tenants are, on balance, worse off under controls because of constraints on housing consumption. And in markets with significant uncontrolled sectors, rent controls can drive up the price of uncontrolled housing, an important unintended consequence further complicating the incidence of its costs.

See Also

► Housing Policy in the United States
► Housing Supply

Bibliography

Arnott, R. 1995. Time for revisionism on rent control. *Journal of Economic Perspectives* 9 (1): 99–120.
Fischel, W.A. 2001. *The homevoter hypothesis: How home values influence local government taxation, school finance, and land-use policies*. Cambridge, MA: Harvard University Press.
Glaeser, E.L., and E.F.P. Luttmer. 2003. The misallocation of housing under rent control. *American Economic Review* 93: 1027–1046.
Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.
Malpezzi, S., and G. Ball 1991. *Rent control in developing countries*, Discussion paper, no. 129. Washington, DC: World Bank.
Munch, J.R., and M. Svarer. 2002. Rent control and tenancy duration. *Journal of Urban Economics* 52: 542–560.
Murray, M.P., C.P. Rydell, C.L. Barnett, C.E. Hillstead, and K. Neels. 1991. Analyzing rent control: The case of Los Angeles. *Economic Inquiry* 29: 601–625.
Olsen, E.O. 1972. An econometric analysis of rent control. *Journal of Political Economy* 80: 1081–1100.
Simmons-Mosley, T.X., and S. Malpezzi. 2006. Household mobility in New York City's regulated rental housing market. *Journal of Housing Economics* 15: 38–62.
Turner, B., and S. Malpezzi. 2003. A review of empirical evidence on the costs and benefits of rent control. *Swedish Economic Policy Review* 10: 11–56.

Rent Seeking

Gordon Tullock

Abstract
'Rent seeking' refers to the investment of resources in efforts to create monopolies. Such investments impose a social cost (which may outweigh the benefit to the monopolist) because they are unproductive. That cost is greater than the mere cost of lobbying by special interests for privilege when the privilege is

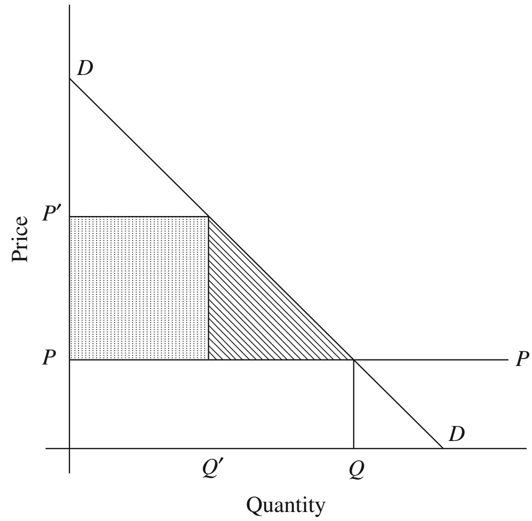
conferred in a way that is economically inefficient but politically feasible (which is often true of regulations). Research on rent seeking has demonstrated that the true social costs of promoting special interests thus greatly exceed the deadweight costs of the distortions introduced into the economy.

Keywords

Consumer surplus; Corruption; Deadweight loss; Krueger, A. O.; Lobbying; Monopoly; Morgan, J. P.; Political influence; Pseudo-equilibrium; Rent; Rent seeking; Social cost; Special interests; Tullock, G.

JEL Classifications

D46



Rent Seeking, Fig. 1

The term 'rent-seeking' was introduced by Ann O. Krueger (1974), but the relevant theory had already been developed by Gordon Tullock (1967). The basic and very simple idea is best explained by reference to Fig. 1. On the horizontal axis we have as usual the quantity of some commodity sold, on the vertical axis its price. Under competitive conditions the cost would be the line labelled PP and that would also be its price. Given a demand curve, DD, quantity Q would be sold at that price. If a monopoly were organized, it would sell Q' units at a price of P'.

The traditional theory of monopoly argued that the net loss to society is shown by the shaded triangle, which represents the consumer surplus that would have been derived from the purchase of those units between Q' and Q, that are now neither purchased nor produced. The dotted rectangle, on the other hand, has traditionally been regarded simply as a transfer from the consumers to the monopolist. Since they are all members of the same society, there is no net social loss from this transfer.

This argument tends to annoy students of elementary economics (because they don't like monopolists), but until the development of the work on rent seeking it was nevertheless thought to be correct by most economists. Its basic problem, however, is that it assumes that the monopoly is created in a costless manner, perhaps by an act of God, whereas in fact real resources are used to create monopolies.

Most discussion of rent seeking has tended to concentrate on those monopolies that are government created or protected, probably because these are observed to be the commonest and strongest. It should be kept in mind, however, that purely private monopolies are possible – indeed, some actually exist. Concentration on government-created monopolies (or restrictions of various sorts that increase certain peoples' income) is probably reasonable, granted the contemporary frequency of such activities. Nevertheless, as we point out below there are certain significant areas where private rent seeking causes net social loss.

In the initial work both of Tullock and Krueger it was assumed that profitseeking businessmen would be willing to use resources in an effort to obtain a monopoly, whether it was privately or government sponsored, up to the point where the last dollar so invested exactly counterbalanced the improved probability of obtaining the monopoly. From this it was deduced that the entire dotted rectangle (Fig. 1) would be exhausted. Although this assumption is open to question (see Tullock 1980), for the time being we will continue to assume that in effect there is no transfer from purchasers to the monopolist, but simply a social loss which comes from the fact that resources have been invested in unproductive activity, i.e. the negatively productive activity of creating a trade restriction of some sort. Theoretical reasons exist for believing that this assumption probably does not fit perfectly anywhere, but it is just as likely to overestimate as to underestimate the social cost; it will be discussed more thoroughly below.

To quote an aphorism frequently used in rent seeking: 'the activity of creating monopolies is a competitive industry.' For this reason it is anticipated that quite a number of people at any given time are putting at least some resources into an effort to secure a monopoly, only some of whom are successful. The situation is like a lottery, in which many people buy lottery tickets, a few win a very large amount of money and the rest lose, perhaps large or small amounts, depending on how much they have committed. In almost all existing lotteries, of course, the total investment of resources by the gamblers is considerably greater than the total payoff, whereas here it is still assumed that total resources committed to rent-seeking equal the total monopoly profits.

Thus the activity of creating monopolies could both absorb very large resources, particularly those resources that take the form of exceptionally talented individuals who devote their attention to this difficult and highly rewarded activity, and lead to considerable redistribution of wealth in the community. Suppose that ten different lobbyists go to Washington representing ten different associations, and each spends one million dollars over the course of a couple of years in the hope of influencing Congress to provide them with a monopoly. Only one of the lobbyists is successful and the monopoly turns out to have a present discounted value of ten million dollars. There is a substantial redistribution of resources from the unsuccessful lobbyists to the successful.

This substantial redistribution has occurred simultaneously with a considerable waste of resources in general, both because these highly intelligent people could otherwise be doing something of higher productivity and because the economy's use of resources has been further distorted by the creation of the monopoly. Further, although so far the discussion has been primarily about monopoly, actually very many possible interventions in the market process raise the same problem. A simple maximum or minimum price may have very large redistributive effects and the people who thus benefit may put considerable resources into receiving them. Of course there are many situations in which one lobbyist is pushing for a particular restriction and another lobbyist is pushing against it. The second activity is sometimes called 'rent avoidance', but it is costly and of course would not exist if there were not also rent seeking activity.

Another area is simple direct transfers. A tax on A for the purpose of paying B will lead to lobbying activity for the tax on the part of B and against it on the part of A. The total of these two lobbying activities could very well equal the total amount transferred (or prevented from being transferred), although one or other of these entrepreneurs will of course gain if his lobby is successful. Assume that A puts in $50 for lobbying to get $100 from B and B puts in $50 lobbying against that. Regardless of the outcome, one party will gain $50 from his lobbying. Society has lost $100.

Of course it is not true that everyone in society is in an equally good position to seek rents. Some kinds of interest are more readily organized than others and we would anticipate that they would win. There are however very many such interests and anyone who spends any time in Washington quickly realizes that there is a major industry engaged in just this kind of activity.

Actual social cost however is clearly very much greater than the mere cost of the various lobbying organizations in Washington. In particular it is normally necessary for the rent-seeking group to undertake directly productive activities in a way that is markedly inefficient, because it is necessary to introduce a certain element of deception into the process. In 1937, when the US Civil Aeronautics Board was organized, it would not have been politically feasible to put a direct tax on purchasers of airline tickets and use it to pay off the stockholders of the airline companies. Regulation, which has a similar effect but at a very much higher cost to the users of airlines per dollar of profit to the owners, was however, politically possible. The necessity of using inefficient methods of transferring funds to the potential beneficiary, because the efficient methods would be just too open and above board, is often one of the major costs of rent seeking. The rent avoidance lobbyist would have had too easy a time if the proposal had been a tax on uses of airlines for the benefit of the stockholders.

Note that in this case the argument against rent seeking turns out also to be an argument against political corruption. Suppose you are in a society which has an exchange control system and that it is possible to buy foreign currency by bribing an official in the exchange control office. This is the kind of situation dealt with by Krueger (1974), who was able to obtain a measure of the total social cost in Turkey and India where the amounts of the necessary bribes were well known; the cost varied from 7–15 per cent of the total volume of transactions.

Traditionally economists have tended to view this kind of bribery as in itself desirable, because it gets around an undesirable regulation. However, it leads to rent seeking. In this case the rent seeking does not come from the users of the permits but from the competition to get into the position where you can receive the bribe. Throughout the underdeveloped world, large numbers of people take fairly elaborate educational programmes which have no real practical value for their future life and engage in long periods of complicated political manoeuvring in hope that they will be appointed, let us say, a customs inspect in Bombay. Since these young men have a free career choice presumably the expected returns from this career are the same as in any other. The difference is that a doctor, say, begins earning money immediately on completing medical school whereas the young man who has studied economics and is now trying to obtain appointment as customs inspector will have a considerable period of time in which he is not appointed at all. Indeed, there will probably be enough such candidates that he has only perhaps one chance in five of being so appointed. The total cost of the rent seeking is the inappropriate education and the political manoeuvring of the five people of whom only one is appointed.

So far we have assumed that the total cost of rent seeking is the present discounted value of the income stream represented by the dotted rectangle in Fig. 1. This assumes a special form for the function which ‘produces’ the monopoly or other privilege. It must be linear, with each dollar invested having exactly the same payoff in probability of achieving the monopoly as the previous dollar (Tullock 1980). Most functions do not have this form, instead they are either increasing or decreasing cost functions.

If the organizing of private monopolies, or of influencing the government into giving you public monopolies, is subject to diseconomies of scale, then total investment in rent seeking will be less than the total value of the rents derived even if we assumed a completely competitive market with completely free entry. When there are economies of scale the situation is even more unusual. Either there is no equilibrium at all or there is a pseudo-equilibrium, in which total investment to obtain the rents is greater than the rents themselves. This is called a pseudoequilibrium, because although it meets all the mathematical requirements for an equilibrium, it is obviously absurd to assume that people would, to take a single example, pay \$75.00 for a 50–50 chance of \$100.

Obviously, what is needed is empirical research, and an effort to measure the production functions appropriate to rent seeking. So far, however, no one has been able to develop a very good way of making such measurements. It seems likely that it would be easier to measure the costs of generating political influence than of private monopolies, if only because many of the expenditures used to influence the government appear in accounts in various places. The costs of private monopolies on the other hand, tend to be much more readily concealed. This does not mean that they do not exist.

The reader has no doubt been wondering what is wrong with rents and why we concern ourselves deeply with rent seeking. The answer to this is that the term itself is an unfortunate one. Obviously, we have nothing against rents when they are generated by, let us say, discovering a cure for cancer and then patenting it. Nor do we object to popular entertainers like Michael Jackson earning immense rents on a rather unusual collection of natural attributes together with a lot of effort on his part to build up his human capital. On the other hand, we do object to the manufacturer of automobiles increasing the rent on his property, and his employees increasing the rent on their union memberships, by organizing a quota against imported cars. All of these things are economic rents, but strictly speaking the term 'rent seeking' applies only to the latter. Its meaning might be expanded to seeking rents from activities which are themselves detrimental. The man seeking a cure for cancer is engaged in an activity which clearly is not detrimental to society. Thus we may observe immediately that activities aimed at deriving rents cover a continuum, but that the term 'rent seeking' is only used for part of that continuum.

The analysis of 'rent seeking' has been one of the most stimulating fields of economic theory in recent years. The realization that the explanation of the social cost of monopoly which was contained in almost every elementary text in economics was wrong, or at the very least seriously incomplete, came as quite a surprise. Revision of a very large part of economic theory in order to take this error into account is necessary. And history also needs to be revised. That J.P. Morgan was an organizer of cartels and monopolies during most of his life is well known, as is the fact that he received very large fees for this, fees which were part of the rent seeking cost of generating these monopolies. It is possible to argue that as a stabilizing factor in the banking system, Morgan more than repaid to the United States the social cost of his monopolistic activities in industry. But that there was a very large rent seeking cost is obvious. This cost is in addition to the deadweight cost of the monopolies.

To date, research on rent seeking has to a considerable extent changed our way of looking at things. We now talk of a great deal of government activity as rent seeking on the part of somebody or other. It was known that special interest existed, but we have traditionally tended to underestimate its cost greatly because we looked only at the deadweight costs of the distortion introduced into the economy. The realization that the actual cost is much greater socially, that the large-scale lobbying industry is truthfully a major social cost, is new although presumably, at all times, anyone who thought about the matter must have realized that these highly talented people could produce more in some other activity.

See Also

- ▶ Bribery
- ▶ Directly Unproductive Profit-Seeking (DUP) Activities

Bibliography

Buchanan, J., R. Tollison, and G. Tullock, eds. 1980. *Toward a theory of the rent-seeking society*. College Station: Texas A and M University Press.

Krueger, A.O. 1974. The political economy of the rent-seeking society. *American Economic Review* 64: 291–303.

Tullock, G. 1967. The welfare cost of tariffs, monopolies, and theft. *Western Economic Journal (now Economic Inquiry)* 5: 224–232.

Tullock, G. 1980. Efficient rent seeking. In eds. Buchanan, Tollison and Tullock, 91–112.

Rentier

J. A. Kregel

An individual who lives on interest income (rent) received in compensation for the loan of property held in the form of money is called a *rentier,* not to be confused with landowners who receive rent paid for the loan of property in land. The original French word signifies the holder of *rentes,* a form of French government debt that was widely diffused in the 19th century. It served to distinguish this form of income from the definition of interest, used by some economists such as Walras (1926, p. 223), as payment made by the entrepreneur to the *capitaliste* for the use of his capital. Today the term refers generically to the owner of any debt obligation, public or private, paying periodic, annual or semi-annual, usually fixed amounts of interest over a long term.

As the origin of the word suggests, the emergence of rentiers accompanied the accumulation of the economic surplus in money form, and the finance of state expenditure (usually for bellicose purposes) by long-term debt. Modern securities markets originated in the sale and exchange of government 'stocks' in London and Amsterdam to merchants who could find no more remunerative employment for their capital in commercial activities. The rentier is therefore a typical feature of the money economy emerging from early capitalist development.

Thus Ricardo's (1821, p. 335) early 19th-century position that landlords' interests were inimical to industrial expansion was replaced in the 20th century by criticism of the rentier as a brake on the dynamics of capitalist accumulation. Individuals who derived income from neither labour nor productive capital investment were viewed as parasites living off the effort of the labourer and the entrepreneur-capitalist.

Economists' views of the rentier largely depend on their theory of interest. The rentier may be considered as enjoying the consumption he had postponed by saving out of income earned at an earlier time, or he may be considered as appropriating through interest the current output of labour and capital, thereby reducing the surplus available for expansion. For Schumpeter (1934, p. 175), rentiers could only exist if capitalist entrepreneurs were sufficiently dynamic to generate the profits required to pay interest, while for Marshall (1920, p. 232) investment was possible only when individuals were willing to 'wait' to consume their income and wealth.

Rentiers as a class have also been criticized for actively discouraging economic and social change in defence of their vested interests in private property rights and money contracts. As well as defending the right to interest income and accumulated wealth, the rentier has to defend the purchasing power of his interest income and the capital value of his wealth. Inflation is thus the first enemy of the rentier living on fixed interest payments, for it reduces the real purchasing power of current income. As a class, rentiers will thus favour conservative government policies to balance budgets and produce deflationary conditions, even at the expense of economic growth and high levels of employment. An example may be found in rentier support in the interwar years for re-establishing prewar gold parities in order to restore asset values decimated by the postwar inflations (Rolfe and Burtle 1975). On the other hand, monetary policy to control inflation will be considered undesirable, for increasing interest rates reduces the current market value of fixed interest debt. Thus the rentier favours balanced budgets, and stable monetary policy and exchange rates to assure security of real income and capital value.

It may appear paradoxical that rentiers favour balanced government budgets, which would eliminate the creation of part of the assets which produce their incomes. However, this simply reflects different frameworks used to analyse the economic significance of the rentier. Ex post, existing rentiers who have already purchased debt and are receiving interests incomes which they seek to protect will have different interests from current savers who are seeking the highest possible future income and who thus look favourably on high interest rates and a plentiful supply of government debt produced by government deficits. A similar argument applies to the issue of private debt. This conflict between current holders and purchasers of debt reflects that between economists in favour of conservative and progressive government economic policies for growth and development.

This conflict of interests has had an important influence on both economic history and economic theory. Keynes, in his *Tract on Monetary Reform* (1923), had already noted the importance of the impact of monetary and fiscal policy on rentiers' standards of living for the political stability of the then emerging modern Europe. This is a theme which he also subsequently incorporated into his *General Theory* (1936, p. 376) in the guise of the concept of liquidity preference and his recommendation for the gradual 'euthanasia' of the rentier by means of expansionary fiscal policy and low rates of interest.

Keynes's theory did not rely on the earlier arguments concerning the inimical effects on growth and employment of the use made by rentiers of their income (for personal consumption) and wealth (to finance public consumption), but rather on the advantages that rentiers would find in holding liquid assets rather than in financing employment creating investment in periods when they felt threatened by uncertainty over the future value of their income and capital. In such conditions employment creating investment would have to compete with the rentiers' preference for liquidity, creating rates of interest far in excess of what entrepreneurs could pay from the expected earnings of productive investment. Further, rentier preferences might be so strong as to render the monetary authority powerless to reduce interest rates to stimulate activity. Keynes thus advocated a policy of direct intervention through the 'socialization of investment', accompanied by low, stable rates of interest which would eventually eliminate the power of rentiers to hinder policies for full employment.

In the 1950s and 1960s expansionary policies and low interest rates reduced the share of interest in total incomes; preservation of individual savings was secured by insurance companies, pension funds and investment trusts. In the 1970s, however, the oil crisis created large quantities of wealth for petroleum producers, many of whom were private individuals who behaved as typical rentiers in investing their funds. The use of restrictive monetary policy to combat the subsequent inflation led to sharply positive real rates of interest, an increase in the share of interest in total incomes and resuscitation of the rentier. Economists are also divided on the implications of this phenomenon, some relating it to the emergence of a new conservatism in economic policy, while others see it as having created the willingness to lend, upon which the investment necessary for the economic recovery and the reconstruction of the industrialized economies is thought to depend.

See Also

► Interest and Profit
► Keynesianism

Bibliography

Keynes, J.M. 1923. *A tract on monetary reform*. London: Macmillan.

Keynes, J.M. *1936. The general theory of employment, interest and money. London*: *Macmillan*.

Marshall, A. 1920. *Principles of economics*. 8th ed. London: Macmillan.

Ricardo, D. 1821. *On the principles of political economy and taxation.* Vol. I of *The works and correspondence of David Ricardo,* ed. P. Sraffa with M. Dobb, Cambridge: Cambridge University Press for the Royal Economic Society, 1951.

Rolfe, S.E., and J.L. Burtle. 1975. *The great wheel: The world monetary system*. New York: McGraw-Hill.

Schumpeter, J. 1934. *The theory of economic development*. New York: Oxford University Press 1961.

Walras, L. 1926. *Elements of pure economics.* Trans. and ed. W. Jaffé, London: Allen & Unwin for the Royal Economic Society, 1954.

Reparations

Kim Oosterlinck

Abstract

Reparations for damage caused, paid by the loser following wars, have been known since Antiquity, although much of the literature focuses on the First World War. There has been much debate, both politically and among economists, on the appropriate basis on which to pay.

Keywords

Reparations; Transfer problem; War

JEL Classifications

F30; F31; F33; F34; F41; F51; F53; F56; N14; N40

From time immemorial, wars have led to international financial transfers. During Antiquity, the practice of imposing a payment of a tribute on the vanquished was already common practice (Livy V, pp. 48–9). War indemnities were common during the Middle-Ages. From a legal point of view, few authors questioned the practice, and victory was seen as a sufficient motivation to extract tribute. With the development of theories on international law in the 16th and 17th centuries, the idea of 'repairing' the torts done by the war came to be discussed.

During the 19th century various legal arguments were invoked to justify war indemnities. The amounts to be paid could, for example, be meant to cover war expenditures incurred by the victor, to make good losses as a result of the war, or to guarantee the victor's future safety. Most treaties remained silent or at best vague regarding the exact computation of the amounts due, and the final figures were often the result of long negotiations between the parties. To secure payments, occupation of part of the defeated country was common. France had to cope with a German occupation after the Second Treaty of Paris (1815) and as a consequence of the Treaty of Frankfurt (1871) ending the Franco-Prussian war. For both wars, France would eventually pay in full (White 2001). The amounts (close to FF1.9 billion for the Napoleonic war and equal to FF5 billion for the Franco-Prussian one) were at the time viewed as exorbitant, and produced heated debates and controversies in France. Devising the means to pay for the reparations also proved to be an arduous task, which has been analysed in detail by economic historians. Loans would eventually provide the bulk of the payment, even though, in 1871, economists such as Louis Wolowsky and bankers such as Henri Germain suggested instead creating a progressive income tax.

The importance of war indemnities before the First World War should not be downplayed. For example, the 230 million silver taels extracted from China after the Sino-Japanese War played a preeminent role in Japan's decision to join the gold standard (Metzler 2006, p. 3). The First World War would however act as a real turning point regarding the literature on reparations. Several elements contributed to this change. First, the intensity of the conflict and its huge material and human costs rendered the question of reparations of foremost importance. Second, the Treaty of Versailles (art. 232) directly referred to the notion of reparations. Since the payments were only meant to 'repair' the damages done, economists would attempt to evaluate the exact size of this damage. This would raise two questions: which were the losses for which Germany was to pay, and how could one evaluate these? Third, the phrasing of the Treaty of Versailles opened the door to further discussions regarding the amounts due. Indeed, it left the determination of the amounts to be paid to the Reparation Commission, an ad-hoc Inter-Allied Commission, which had to provide an estimate before 1 May 1921. Fourth, the Treaty (art. 232 and art. 234) alluded to the fact that Germany's capacity to pay had to be taken into account, which meant this capacity had somehow to be estimated. Eventually, the World War reparations would certainly not have been so central in the literature had John Maynard Keynes not attacked them vigorously in his *Economic Consequences of the Peace* (1919). Keynes's subsequent fame led the broader public to accept his point of view, even though it was firmly contested when first expressed (Bainville 1920; Ohlin 1929).

Keynes attacked the treaty on several fronts. First, he suggested a limited interpretation regarding the scope of the damages. During the discussions related to the amounts to be paid, France and Great Britain at first pleaded that Germany should be made to pay all the costs of the war. For the French finance minister, Louis-Lucien Klotz, it was clear that Germany would pay ('*le boche paiera*'). Eventually however, the position defended by the United States and Belgium would prevail, and the treaty (art 232) would state that reparations were to cover 'the damage done to the civilian population of the Allied and Associated Powers and to their property'. Keynes defended a narrow interpretation of this article, and considered that only direct damages caused by the war should be taken into account. To Keynes's consternation, costs such as pensions and separation allowances were included in the total. Keynes found the terms so harsh that he would qualify the Treaty of Versailles as a ‘Carthaginian Peace’, referring to the extremely hard conditions imposed upon defeated Carthage by Rome after the Second Punic War (218 to 201 BC) (Livy XXX, 16 and 37).

Keynes then proceeded to criticize the first tentative evaluations of the amounts that had been put forward during the Paris Peace Conference. According to him, French estimates regarding reconstruction costs were much too high. Once converted into pounds, they ranged from £2.6 to 5.63 billion, whereas Keynes’s own computation led him to suggest they should be rounded up to £2 billion. Mantoux (1946), probably one of the fiercest critics of Keynes’s estimates, showed that Keynes erroneously used the pre-war exchange rate to convert the French figures into pounds, therefore largely inflating these once converted. He further suggested that the amounts really spent for the reconstruction were actually quite close to the French estimates.

Eventually, the Reparation Commission concluded on 5 May 1921 that Germany would have to pay a total of 132 billion marks (£6.6 billion), and cover Belgium’s war debt, worth the equivalent of 5.6 billion marks. In practice, Germany would pay interest and amortization on two bonds series: the A one representing 12 billion marks and the B one worth 38 billion. The remaining 82 billion would be issued in a third series of bonds (the C one) only if Germany ever became prosperous enough to service them on top of the series A and B. As a matter of fact, most Allied politicians never believed this last issue would ever take place. Furthermore Germany was bound to pay a fixed annuity of 2 billion marks and a variable amount worth 26 per cent of its exports.

According to Keynes, Germany did not have the capacity to pay these amounts. He viewed the requested sums as so high that they would ‘reduce Germany to servitude for a generation, degrade the lives of million of human beings and deprive a whole nation of happiness’ (Keynes 1929, p. 17). In his eyes, two different issues were to be taken into account when one considered Germany’s capacity to pay: there was both a budgetary problem linked to German capacity to extract the required sums from its citizens and a transfer problem linked to the conversion of the obtained marks into hard currency.

According to Keynes the transfer problem was as important an issue as the budgetary problem (Keynes 1929). The macroeconomic impact of international transfers had already been addressed before the war. The traditional view was that the transfer-paying country would suffer deterioration in its terms of trade. In the case of the First World War, Keynes believed that creditor countries would not allow Germany to run a huge trade surplus and would thus force Germany to let its terms of trade deteriorate. This would in turn reduce Germany’s real income and in a sense add a further burden to the defeated country. Ohlin (1929) attacked this view on the grounds that income effects could induce the recipient countries to buy more German goods and thus improve its terms of trade. The transfer problem in the framework of reparations has since then led to a large literature, and has recently been analysed using recent macroeconomics approaches (Morrison 1992; White 2001; Devereux and Smith 2007 for example).

This transfer problem would partially be addressed in the framework of the Dawes plan. At the end of 1921, Germany started negotiations to get a moratorium on its debts. In exchange for this moratorium, France requested productive pledges, which Great Britain opposed. On 11 January 1923 French and Belgian troops entered the Ruhr to seize part of Germany’s mines and industrial production. German passive resistance to the occupation, its dramatic economic situation characterized by hyperinflation, and geopolitical changes all led to the resumption of negotiations. Eventually, the plan drafted by the former US Director of the Bureau of the Budget, Charles G. Dawes, was agreed upon in August 1924. The Ruhr was to be evacuated, and it was hoped that the transfer problem would be resolved by accepting that Germany pay the annuities either in gold marks or in its German monetary equivalent on a special account of the Reichsbank. The Dawes plan was meant to be a provisional settlement but remained in application for five years. It also opened the door to US investment in Germany, which quickly boomed and rendered the payment of reparations much easier for Germany (Klug 1990).

Reparations represented but part of the international debts stemming from the war. The sums lent by the United States to its European allies were huge, and had been a major point of contention since the end of the war. Allied European countries wanted them suppressed or reduced. In 1928, they pleaded to link their reimbursement to the reparations. This would lead to a revision of the Dawes plan by a team of experts led by Owen D. Young. The Young report was finalized on 7 June 1929 and adopted shortly afterwards. The amounts due by Germany were reduced to 121 billion marks and were to be administered by the Bank for International Settlements.

The Young plan was short lived. The great depression forced Germany to ask for a moratorium, which lead to negotiations and the Lausanne agreements which stated that Germany would pay 3 billion gold marks as final settlement for the reparations. The Lausanne agreements were never formally signed by all parties, but payments nonetheless stopped after 1932. Depending on the sources, the estimation of the amounts paid differs greatly, ranging from close to 21 billion marks by the Reparation Commission to close to 68 billion according to the pre-Second World War German government.

In the 1970s the Keynesian view became the subject of more and more critiques by historians and economists who had access to new archival material. As mentioned by Schuker (1988), however, 'not only did the Reich entirely avoid paying net reparations to its wartime opponents; it actually extracted the equivalent of reparations from the Allied powers, and principally from the United States'.

In view of the preeminence of the post-First World War German reparations in the literature, many economists still base their views on reparations on this single example, and conclude that reparations are in essence an inefficient mechanism. Contrasting with the German case, history however provides many examples when reparations were paid in full. Table 1 provides a comparison of war reparations for four historical episodes. Even though Germany's indemnities were higher in percentage of one year's GDP than the ones imposed on defeated France in 1819 and 1871, they still remained far below the indemnities paid by occupied France during the Second World War. Noteworthy, amounts were paid in full for all the episodes mentioned except the First World War.

Reparations, Table 1 A comparison of war reparations

	Indemnities (billions)	Percentage of one year's GDP	Share of debt service to GDP
France 1815–19	FF 1.65 to 1.95	18 to 21	1.2 to 1.4
France 1871	FF 5.0	25	0.7
Germany	DM 50	83	2.5
1923–31			
Vichy 1940–4	FF 479 (633)	111 (147)	2.6 (3.4)

Source: White (2001), Klug (1990) and Occhino et al. (2008). Note: the numbers in parentheses include occupation payments and looting

Even though the experience of the First World War appeared as negative to most, the Allied countries, anticipating victory, started discussing reparations even before the end of the Second World War. In the framework of the Yalta Conference, which took place on 11 February 1945, the Allied countries agreed to make Germany pay for the 'losses caused by her to the Allied nations in the course of the war'. The protocol further suggested that reparations in kind should be extracted from Germany (by requesting equipment, machine tools, ships, rolling stock, German investments abroad, and shares of industrial, transport and other enterprises in Germany, or by asking for annual deliveries of goods, or even by using German labour). Exact details were to be determined by a Commission of Damage to be seated in Moscow.

Following Germany's surrender, the Commission of Damage met to estimate the reparations. Heated debates, fuelled by geopolitical considerations, animated the discussions. Since the Soviet Union had suffered dramatically from the war, generous reparations would have allowed it to recover quickly, a prospect that did not please countries from the Western bloc. Despite the diverging views, the Commission managed to present a series of principles which were agreed upon at the Potsdam Conference on 2 August 1945. Most importantly, dismantling of German factories was to represent one of the main sources of reparations. The Cold War and Germany's partition would soon lead to a separation of responsibilities regarding the reparations: West Germany would pay reparations to all countries but Poland and the Soviet Union, which were to be covered by East Germany. More than 650 factories (or part of factories) were dismantled from West Germany and transferred for an estimated value of US$130 million, representing 25 per cent of the payment eventually made by West Germany. In East Germany, dismantling was soon abandoned in favor of the transfer of part of its industrial production. Estimates of the magnitude of the transfers are tentative at best, but tend to indicate that they were of large magnitude.

During the Cold War few conflicts led to reparations, most probably because each belligerent could most of the time count on the support of one of the superpowers and therefore avoid a painful settlement (D'Argent 2002). The practice did not however disappear, and as a consequence of the first Gulf War, a UN Security Council resolution forced Iraq to pay reparations to the victims of its aggression. The first Gulf War case would, on top of the 'traditional' questions, raise an additional one. Since payments were expected to come from oil sales, we could wonder, as does Morrison (1992), 'how to proceed when the guilty nation has the ability to affect world prices – even to the point where it may be able to reduce its reparations burden and inflict real income losses on those seeking compensation'.

See Also

► War and Economics
► World Wars, Economics of

Bibliography

Bainville, J. 1920. *Les Conséquences Politiques de la Paix*. Paris: Nouvelle librairie nationale.

D'Argent, P. 2002. *Les réparations de guerre en droit international public. La responsabilité internationale des Etats à l'épreuve de la guerre*. Bibliothèque de la Faculté de Droit de Louvain, 36. Brussels: Bruylant.

Devereux, M.B., and G.W. Smith. 2007. Transfer problem dynamics: Macroeconomics of the Franco-Prussian war indemnity. *Journal of Monetary Economics* 54: 2375–2398.

Keynes, J.M. 1919. *The economic consequences of the peace*. London: Macmillan.

Keynes, J.M. 1922. *A revision of the treaty*. London: Macmillan.

Keynes, J.M. 1929. The German transfer problem. *Economic Journal* 39: 1–7.

Klug, A. 1990. *The theory and practice of reparations and American loans to Germany, 1925–1929*. Princeton working papers in international economics, G-90-03, International Finance Section.

Livy. *Ab urbe condita* (The early history of Rome, books I–V, and The history of Rome from its foundation, books XXI–XXX: The war with Hannibal). London: Penguin Classics, 2002 and 1976.

Mantoux, E. 1946. *The Carthaginian peace or the economic consequences of Mr. Keynes*. London: Oxford University Press.

Metzler, M. 2006. *Lever of empire: The international gold standard and the crisis of liberalism in prewar Japan*. Berkeley/Los Angeles: University of California Press.

Morrison, R.J. 1992. Gulf war reparations: Iraq, OPEC, and the transfer problem. *American Journal of Economics and Sociology* 51: 385–399.

Occhino, F., K. Oosterlinck, and E. White. 2008. How much can a victor force the vanquished to pay? *Journal of Economic History* 68: 1–45.

Ohlin, B. 1929. The reparation problem: A discussion. *Economic Journal* 39: 172–182.

Schuker, S.A. 1988. 'American reparations' to Germany, 1919–33: Implications for the third-world debt crisis. Princeton Studies in International Finance no. 61.

White, E.N. 2001. Making the French pay: The cost and consequences of the Napoleonic reparations. *European Review of Economic History* 5: 337–365.

Repeated Games

Kandori Michihiro

Abstract

This article shows why self-interested agents manage to cooperate in a long-term relationship. When agents interact only once, they often have an incentive to deviate from cooperation. In a repeated interaction, however, any mutually beneficial outcome can be sustained in an equilibrium. This fact, known as the folk theorem, is explained under various information structures. This article also compares repeated games with other means to achieve efficiency, and briefly discusses the scope for potential applications.

Keywords
Antitrust enforcement; Bargaining; Cartels; Collusion; Contract theory; Cooperation and its evolution; Correlated equilibrium; Equilibrium selection; Finite horizons; Folk th; Imperfect monitoring; Infinite horizons; Informal contracts; International policy coordination; Long-term relationships; Mechanism design; Multiple equilibria; Perfect monitoring; Private monitoring; Public monitoring; Repeated games; Relational contracts; Subgame perfect equilibrium; Trigger strategy; Uniqueness of equilibrium

JEL Classifications
C7

Repeated games provide a formal and quite general framework to examine why selfinterested agents manage to cooperate in a long-term relationship.

Formally, repeated games refer to a class of models where the same set of agents repeatedly play the same game, called the ‘stage game’, over a long (typically, infinite) time horizon. In contrast to the situation where agents interact only once, *any* mutually beneficial outcome can be sustained as an equilibrium when agents interact repeatedly and frequently. A formal statement of this fact is known as the folk theorem.

Repeated Games and the General Theories of Efficiency

Thanks to the developments since the mid-1970s, economics now recognizes three general ways to achieve efficiency: (a) competition; (b) contracts; and (c) long-term relationships. For standardized goods and services, with a large number of potential buyers and sellers, promoting market competition is an effective way to achieve efficiency. This is formulated as the classic First and Second Welfare Theorems in general equilibrium theory. There are, however, other important resource allocation problems which do not involve standardized goods and services. Resource allocation within a firm or an organization is a prime example, as pointed out by Ronald Coase (1937), and examples abound in social and political interactions. In such cases, aligning individual incentives with social goals is essential for efficiency, and this can be achieved by means of *incentive schemes* (penalties or rewards). The incentive schemes, in turn, can be provided in two distinct ways: by a formal contract or by a long-term relationship. The penalties and rewards specified by a formal contract are enforced by the court, while in a long-term relationship the value of future interaction serves as the reward and penalty to discipline the agents’ current behaviour. The theory of contracts and mechanism design concern the former case, and the theory of repeated games deals with the latter. These theories provide general methods to achieve efficiency, and have become important building blocks of modern economic theory.

An Example: Collusion of Gas Stations and the Trigger Strategy

Consider two gas stations located right next to each other. They have identical and constant marginal cost c (the wholesale price of gasoline) and compete by publicly posting their prices. Suppose their joint profit is maximized when they both charge $p = 10$, whereby each receives a large profit π. Although this is the best outcome for them, they have an incentive to deviate. By slightly undercutting its price, each can steal all the customers from its opponent, and its profit (almost) doubles. The only price free from such profitable deviation is $p = c$, where their profit is equal to zero. In other words, the only Nash equilibrium in the price competition game is an

R

inefficient (for the gas stations) outcome where both charge $p = c$. This situation is the rule rather than the exception: the Nash equilibrium in the stage game, the only outcome that agents can credibly achieve in a one-shot interaction, is quite often inefficient for them. This is because agents seek only their private benefits, ignoring the benefits or costs of their actions for their rivals.

In reality, however, gas stations enjoy positive profits, even when there is another station nearby. An important reason may well be that their interaction is not one-shot. Formally, the situation is captured by a *repeated game,* where the two gas stations play the price competition game (the stage game) over an infinite time horizon $t = 0$, 1, 2,. . . . Consider the following repeated game strategy:

1. Start with the optimal price $p = 10$.
2. Stick to $p = 10$ as long as no player (including oneself) has ever deviated from $p = 10$.
3. Once anyone (including oneself) deviated, charge $p = c$ for ever.

This can be interpreted as an explicit or implicit agreement of the gas stations: charge the monopoly price $p = 10$, and any deviation triggers cut-throat price competition ($p = c$ with zero profit). Let us now check whether each player has any incentive to deviate from this strategy. Note that, if neither station deviates, each enjoys profit n every day. As we saw above, a player can (almost) double its stage payoff by slightly undercutting the agreed price $p = 10$. Hence the short-term gain from deviation is at most π. If one deviates, however, its future payoff is reduced from π to zero in each and every period in the future. Now assume that the players discount future profits by the *discount factor* $\delta \in (0, 1)$. The number δ measures the value of a dollar in the next period. The discounted future loss is $\delta\pi + \delta^2\pi + \cdots = \frac{\delta}{1-\delta}\pi$. If this is larger than the short-term gain from defection (π), no one wants to deviate from the collusive price $p = 10$. The condition is $\pi \leq \delta/(1 - \delta)\pi$, or equivalently, $1/2 \leq \delta$.

Next let us check whether the players have an incentive to carry out the threat (the cut-throat price competition $p = c$). Since $p = c$ is the Nash equilibrium of the stage game, charging $p = c$ in each period is a best reply if the opponent always does so. Hence, the players are choosing mutual best replies. In this sense, the threat of $p = c$ is credible or self-enforcing.

In summary, under the strategy defined above, players are choosing mutual best replies *after any history,* as long as $1/2 \leq \delta$. In other words, the strategy constitutes a *subgame perfect equilibrium* in the repeated game. Similarly, in a general game, any outcome which Pareto dominates the Nash equilibrium can be sustained by a strategy which reverts to the Nash equilibrium after a deviation. Such a strategy is called a *trigger strategy.*

Three Remarks: Multiple Equilibria, Credibility of Threat and Renegotiation, and Finite Versus Infinite Horizon

A couple of remarks are in order about the example. First, the trigger strategy profile is not the only equilibrium of the repeated game. The repetition of the stage game Nash equilibrium ($p = c$ for ever) is also a subgame perfect equilibrium. Are there any other equilibria? Can we characterize *all* equilibria in a repeated game? The latter question appears to be formidable at first sight, because there are an infinite number of repeated game strategies, and they can potentially be quite complex. We do have, however, some complete characterizations of all equilibria of a repeated game, such as folk theorems and self-generation conditions as will be discussed subsequently.

Second, one may question the credibility of the threat ($p = c$ for ever). In the above example, credibility was formalized as the subgame perfect equilibrium condition. According to this criterion, the threat $p = c$ is credible because a *unilateral* deviation by a *single* player is never profitable. The threat $p = c,$ however, may be upset by *renegotiation.* When players are called upon to carry out this grim threat after a deviation, they may well get together and agree to 'let bygones be bygones'. After all, when there is a better equilibrium in the repeated game (for example, the trigger strategy equilibrium), why do we expect the players to stick to the inefficient one (p = c)? This is the problem of *renegotiation proofness* in repeated games. The problem is trickier than it appears, however, and economists have not yet agreed on what is the right notion of renegotiation proofness for repeated games. The reader may get a sense of difficulty from the following observation. Suppose the players have successfully renegotiated away $p = c$ to play the trigger strategy equilibrium again. This is self-defeating, however, because the players now have an incentive to deviate, as they may well anticipate that the threat $p = c$ will be again subject to renegotiation and will not be carried out. For a comprehensive discussion of this topic (and also of a number of major technical results on repeated games), see an excellent survey by D. Pearce (1990).

Third, let me comment on the assumption of an *infinite* time horizon. Suppose that the gas stations are to be closed by the end of next year (due to a new zoning plan, for ex). This situation can be formulated as a *finitely* repeated game. On the last day of their business, the gas stations just play the stage game, and therefore they have no other choice but to play the stage game equilibrium p = c. In the penultimate day, they rationally anticipate that they will play $p = c$ *irrespective of* their current action. Hence they are effectively playing the stage game in the penultimate day, and again they choose $p = c$. By induction, the *only* equilibrium of the finitely repeated price competition is to charge $p = c$ in *every* period. The impossibility of cooperation holds no matter how long the time horizon is, and it is in sharp contrast to the infinite horizon case.

Although one may argue that players do not really live infinitely long (so that the finite horizon case is more realistic), there are some good reasons to consider the infinite horizon models. First, even though the time horizon is finite, if players do not know in advance exactly *when* the game ends, the situation can be formulated as an infinitely repeated game. Suppose that, with probability $r > 0$, the game ends at the end of any given period. This implies that, *with probability 1,* the game ends in a finite horizon. Note, however, that the expected discounted profit is equal to $\pi(0) + (1 - \mathrm{r})\,\delta\pi(1) + (1 - \mathrm{r})^2\,\delta^2\,\pi(2) + \ldots$, where $\pi(\mathrm{t})$ is the stage payoff in period t. This is identical to the payoff in an infinitely repeated game with discount factor $\delta' = (1 - r)\delta$. Second, the drastic 'discontinuity' between the finite and infinite horizon cases in the price competition example hinges on the uniqueness of equilibrium in the stage game. Benoit and Krishna (1985) show that, if each player has multiple equilibrium payoffs in the stage game, the long but finite horizon case enjoys the same scope for cooperation as the infinite horizon case (the folk theorem, discussed below, approximately holds for T-period repeated game, when $T \to \infty$).

The Repeated Game Model

Now let me present a general formulation of a repeated game. Consider an infinitely repeated game, where players i = 1, 2,..., N repeatedly play the same stage game over an infinite time horizon t = 0, 1, 2,.... In each period, player i takes some action $a_{\mathrm{i}} \in A_{\mathrm{i}}$, and her payoff in that period is given by a stage game payoff function $g_i(a)$, where $a = (a_1, \ldots, a_{\mathrm{N}})$ is the action profile in that period. The repeated game payoff is given by

$$\prod_i = \sum_{t=0}^{\infty} g_i(a(t))\,\delta^t,$$

where $a(t)$ denotes the action profile in period t and $\delta \in (0, 1)$ is the discount factor. It is often quite useful to look at the *average payoff* of the repeated game, which is defined to be $(1 - \delta)\Pi_i$. Note that, if one receives the same payoff x in each period, the repeated game payoff is $\Pi_i = x + \delta x + \delta^2 x + \ldots = x/(1 - \delta)$. This example helps to understand the definition of average payoff: in this case $(1 - \delta)\Pi_i$ is indeed equal to x, the payoff per period.

A *history* up to time t is the sequence of realized action profiles before t s: $h^t = (a(0), a(1), \ldots, a(\mathrm{t} - 1))$. A *repeated game strategy* for player *i,* denoted by s_i, is a complete contingent action plan, which specifies a current action after any history: $a_i(t) = s_i(h^t)$ (a minor note: to determine $a_i(0)$, we introduce a dummy history h^0 such that

$a_i(0) = s_i(h^0)$). A repeated game strategy profile $s = (s_1,\ldots, s_N)$ is a *subgame perfect equilibrium* if it specifies mutual best replies after any history.

The Folk th

Despite the fact that a repeated game has an infinite number of strategies, which can be arbitrarily complicated, we do have a *complete* characterization of equilibrium payoffs. The folk theorem shows exactly which payoff points can be achieved in a repeated game.

Before stating the theorem, we need to introduce a couple of concepts. First, let us determine the set of physically achievable average payoffs in a repeated game. Note that, by alternating between two pure strategy outcomes, say u and v, one may achieve any point between u and v as the average payoff profile. Hence, an average payoff profile can be a weighted average (in other words, a convex combination) of pure strategy payoff profiles in the stage game. Let us denote the set of all such points by V Formally, the set of *feasible average payoff profiles V* is the smallest convex set that contains the pure strategy payoff profiles of the stage game.

Second, let us determine the points in V that cannot possibly be an equilibrium outcome. For example, if a player has an option to stay out to enjoy zero profit in each period, it is a priori clear that her equilibrium average payoff cannot be less than zero. In general, there is a payoff level that a player can guarantee herself in any equilibrium, and this is formulated as the *minimax* payoff. Formally, the minimax payoff for player i is defined as $\underline{v}_i = \min_{\alpha-i} \max_{\alpha i} g_i(\alpha)$, where $\alpha = (\alpha_1,\ldots, \alpha_N)$ is a mixed action profile (α_i is a probability distribution over player i's pure actions) and $g_i(\alpha)$ is the associated expected payoff. To understand why min and max are taken in that particular order, consider the situation where player i always *correctly anticipates what others do*. If player i knows that others choose $\alpha_{-i} = (\alpha_1,\ldots, \alpha_{i-1}, \alpha_{i+1},\ldots, \alpha_N)$, he can play a best reply against $\alpha_{-\mathrm{i}}$ to obtain $\max_{\alpha_i} g_i(\alpha)$. Note well that $\max_{\alpha_i} g_i(\alpha)$ is a function of α_{-i}. In the worst case, where others take the most damaging actions α_{-i} player i obtains the minimax payoff (this is exactly what the definition says). From this definition it is clear that, in any equilibrium of the repeated game, *the average payoff to each player is at least her minimax payoff.* In any equilibrium, each player correctly anticipates what others do, and simply by playing the stage game best reply in each period, any player can make sure that her average payoff is more than her minimax payoff. (A comment: we consider mixed strategies in the definition of the minimax payoff because in many games the minimax payoff is smaller when we consider mixed strategies.)

From what we saw, now it is clear that the set of equilibrium average payoff profiles of a repeated game is *at most* $V^* = \{v \in V \mid \forall i\ v_i > \underline{v}_i\}$. (The points with $v_i = \underline{v}_i$ are excluded to avoid minor technical complications.) The set V^* is called the *feasible and individually rational payoff set*. This is the set of physically achievable average payoff profiles in the repeated game where each player receives more than her minimax payoff. The folk theorem shows that any point in this 'maximum possible region' can indeed be an equilibrium outcome of the repeated game. (Throughout this article, I maintain a minor technical assumption that each player has a finite number of actions in the stage game.)

Folk th In an N-player infinitely repeated game, any feasible and individually rational payoff profile $v \in V^*$ can be achieved as the average payoff profile of a subgame perfect equilibrium when the discount factor 8 is close enough to 1, provided that either N = 2, or N $\geq$ 3 and no two players have identical interests. Formally, no two players have identical interests if there are *no* players i and j ($i \neq j$) whose payoffs satisfy $g_i(a) = bg_j(a) + c$, $b > 0$ (that is, no two players have the same preferences over the stage game outcomes). This is a 'generic' condition that is almost always satisfied: the case where players have identical interests is very special in the sense that the equality $g_i(a) = bg_j(a) + c$ fails by even a slight change of the payoff functions. Hence, the folk theorem provides a general theory of efficiency: it shows that, for virtually any game, any mutually beneficial outcome can be achieved in a long term relationship, if the discount factor is close to 1. *Although game-theoretic predictions quite often depend on the fine details of the model, this result is a notable exception for its generality.*

The crucial condition in the folk theorem is a high discount factor. The discount factor δ may measure the (subjective) patience of a player, or, it may be equal to $1/(1 + r)$, where r is the interest rate per period. Although the discount factor may not be directly observable (in particular, in the former case), it should be high when one period is short. Hence, an empirically testable implication is that players who have daily interaction (such as the gas stations in our ex) have a better scope for cooperation than those who interact only once a year. An important message of the folk theorem is that a high *frequency of interaction* is essential for the success of a long term relationship.

The name 'folk th' comes from the fact that game theorists had anticipated that something like it should be true long before it was precisely formulated and proved. In this sense, the assertion had been folklore in the game theorist community. The proof is, however, by no means obvious, and there is a body of literature to prove the theorem in various degrees of generality. Early contributions include Aumann (1959), Friedman (1971) and Rubinstein (1979). The statement above is based on Fudenberg and Maskin (1986) and its generalization by Abreu et al. (1994). The proof is constructive: a clever strategy, which has a rather simple structure, is constructed to support any point in V^*.

Repeated Games Versus Formal Contracts

To discuss the scope of applications, I now compare a long-term relationship (repeated game) and a formal contract as a means to enforce efficient outcomes. As our gas station example shows, quite often an agent has an incentive to deviate from an efficient outcome, because it increases her private returns at the expense of the social benefit. Such a deviation can be deterred if we impose a sufficiently high penalty so that the *incentive constraint*

$$\text{gain from deviation} \leq \text{penalty}$$

is satisfied. This is the basic and common feature of repeated games and contracts. A formal contract explicitly specifies the penalty and it is enforced by the court. In repeated games, the penalty is indirectly imposed through future interaction. In this sense the theory of repeated games can be regarded as the theory of *informal or relational contracts.*

When is a long-term relationship a better way to achieve cooperation than a formal contract? First, a long-term relationship is useful when a formal contract is too costly or impractical. For example, it is often quite costly for a third party (the court) to verify whether there was any deviation from an agreement, while defections may be directly observed by the players themselves. In practice, what constitutes 'cooperation' is often so fuzzy or complicated that it is hard to write it down explicitly, although the players have a common and good understanding about what it is. 'Pulling enough weight' in a joint research project may be a good example. In those situations, a long-term relationship is a more practical way to achieve cooperation than a formal contract. In fact, a classic study by Macaulay (1963) indicates that the vast majority of business transactions are executed without writing formal contracts. Second, there are some cases where a court powerful enough to enforce formal contracts simply does not exist. For example, in many problems in development economics and economic history, the legal system is highly imperfect. Even for developed countries in the modern age, there are no legal institutions which have enough binding power to enforce international agreements. Hence, repeated games provide a useful framework to address such problems as the organization of medieval trade, informal mutual insurance in developing countries, international policy coordination, and measures against global warming. Lastly, there is no legal system to enforce cartels or collusion, because the existing legal system refuses to enforce any contract that violates antitrust laws. Hence a long-term relationship is the only way to enforce a cartel or collusive agreement.

Is the Folk Theorem a Negative Result?

The theory of repeated games based on the folk theorem is often criticized because it does not, as the criticism goes, have any predictive power. The folk theorem basically says that anything can be an equilibrium in a repeated game. One could argue, however, that this criticism is misplaced if we regard the theory of repeated games as a theory of informal contracts. Just as anything can be enforced when the party agrees to sign a binding contract, in repeated games any (feasible and individually rational) outcome is sustained if the players agree on an equilibrium. Enforceability of a wide range of outcomes is the essential property of effective contracts, formal or informal. The folk theorem correctly captures this essential feature.

This criticism is valid, however, in the sense that the theory of repeated games does not provide a widely accepted criterion for equilibrium selection. When we regard a repeated game as an informal contract, where the players explicitly try to agree on which equilibrium to play, the problem of equilibrium selection boils down to the problem of bargaining. In such a context, it is natural to assume that an efficient point (in the set of equilibria) is played. In the vast majority of applied works of repeated games with symmetric stage games (such as the gas stations ex), it is common to look at the best symmetric equilibrium. In contrast, when players try to find an equilibrium through trial and error, the theory of repeated games is rather silent about which equilibrium is likely to be selected. A large body of computer simulation literature on the evolution of cooperation, pioneered by Axelrod (1984), may be regarded as an attempt to address this issue.

Imperfect Monitoring

So far we assumed that players can perfectly observe each other's actions. In reality, however, long term relationships are often plagued by *imperfect monitoring.* For example, a country may not verify exactly how much CO_2 is emitted by neighbouring countries. Workers in a joint project may not directly observe each others' effort. Electronic appliance shops often offer secret discounts for their customers, and each shop may not know exactly how much is charged by its rivals. In such situations, however, there are usually some pieces of information, or *signals,* which imperfectly reveal what actions have been taken. Published meteorological data indicates the amount of CO_2 emission, the success of the project is more likely with higher effort, and a shop's sales level is related (although not perfectly) to its rivals' prices.

According to the nature of the signals, repeated games with imperfect monitoring are classified into two categories: the case of *public monitoring*, where players commonly observe a public signal, and the case of *private monitoring*, where each player observes a signal that is not observable to others. Hence, the CO_2 emission game and the joint-project game are examples with imperfect public monitoring (published meteorological data and the success of the project are publicly observed), while the secret price-cutting game by electronic shops is a good example with imperfect private monitoring (one's sales level is private information).

This difference may appear to be a minor one, but, somewhat surprisingly, it is not. The imperfect *public* monitoring case shares many features with the *perfect* monitoring case, and we now have a good understanding of how it works. In contrast, the imperfect private monitoring case is not fully understood, and we have only some partial characterizations of equilibria. In what follows, I sketch the main results in the imperfect public and private monitoring cases.

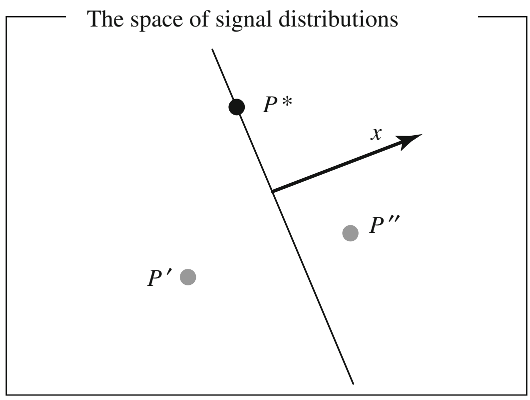
Imperfect Public Monitoring

At first sight, this case might look much more complicated than the perfect monitoring case, but those two cases are similar in the sense that they share a *recursive structure.* Consider the set W^* of all average payoff profiles associated with the subgame perfect equilibria of a perfect monitoring repeated game. Any point $w \in W^*$ is a weighted average of the current payoff g and the continuation payoff w': $(1 - \delta)g + \delta w'$. The continuation payoff typically changes when a player deviates from g, in such a way that the short-term gain from deviation is wiped out. Subgame perfection requires that all continuation payoffs are chosen from the equilibrium set W^*. In this sense, W^* is generated by itself, and this stationary or recursive structure turns out to be quite useful in characterizing the set of equilibria.

The set of equilibria in an imperfect public monitoring game also shares the same structure. Consider the equilibria where the public signal determines which continuation equilibrium to play. When a player deviates from the current equilibrium action, it affects both her current payoff and (through the public signal) her continuation payoff. The equilibrium action should be enforceable in the sense that any gain in the former should be wiped out in the latter, and this is easier when the continuation payoff admits large variations. Formally, given the range of continuation payoffs W, we can determine the set $B(W)$ of enforceable average payoffs. The larger the set W is, the more actions can be enforced in the current period (and therefore the larger the set $B(W)$ is). As in the perfect monitoring case, the equilibrium payoff set $W = W^*$ generates itself: it satisfies the *self-generation condition* of Abreu et al. (1990) $W \subseteq B(W)$. W^* is the largest (bounded) set satisfying this condition, and the condition is in fact satisfied with equality. Conversely, it is easy to show that any (bounded) set satisfying the selfgeneration condition is contained in the equilibrium payoff set W^*.

This provides a simple and powerful characterization of equilibria, which is an essential tool to prove the folk theorem in the imperfect public monitoring case. The folk theorem shows that, despite the imperfection of monitoring, we can achieve any feasible and individually rational payoff profile under a certain set of conditions.

Before presenting a formal statement, let me sketch the basic ideas behind the folk theorem. When monitoring is imperfect, players have to be punished when a ‘bad’ signal outcome ω is observed, and this may happen with a positive probability even if no one defects. For example, in the joint project game, the project may fail even though everyone works hard. A crucial difference between the perfect and imperfect monitoring cases is that, in the latter, punishment occurs *on the equilibrium path.* The resulting welfare loss, however, can be negligible under certain conditions.



Repeated Games, Fig. 1

Consider a two-player game, where the probability distribution of the signal $\omega \in \Omega = \{\omega^1,\ldots, \omega^K\}$, when no one defects, is given by $P^* = (p^*(\omega^1),\ldots, p^*(\omega^K))$ in Fig. 1. Suppose that each player’s defection changes the probability distribution to exactly the same point P'. Then, there is absolutely no way to tell which player deviates, so that the only way to deter a defection is to *punish all players simultaneously,* when a ‘bad’ outcome emerges. This means that surplus is thrown away, and we are bound to have substantial welfare loss. Now consider a case where different players’ actions affect the signal asymmetrically: player 1’s defection leads to point P', while the defection by player 2 leads to P''. In this asymmetric case, one can *transfer* future payoff from player 1 to 2 when player 1’s defection is suspected. Under such a transfer, surplus is never thrown away, and this enables us to achieve efficiency.

More precisely, consider the normal vector x of the hyperplane separating P' and P'' in the figure, and let $w_1 = x$ and $w_2 = -x$ be the continuation payoffs of player 1 and player 2 respectively. Figure 1 indicates that player 1’s expected

R

continuation payoff $P \cdot w_1 = P \cdot x$ is reduced by her own defection ($P' \cdot x < P^* \cdot x$). Similarly, player 2's defection reduces her expected continuation payoffs ($P^* \cdot (-x) > P''(-x)$). Note that this asymmetric punishment scheme does not reduce the joint payoff, because by construction $w_1 + w_2$ is identically equal to 0. This is an essential idea behind the folk theorem under imperfect public monitoring: *When different players' deviations are statistically distinguished, asymmetric punishment deters defections without welfare loss.*

When can we say that different players' deviations are statistically *distinguished?* Note well that the above construction is impossible when P'' is exactly in between P^* and P' (that is, when P'' is a convex combination of P^* and P'). Such a case can be avoided if P^*, P' and P'' are linearly independent. The linear independence of the equilibrium signal distribution (P*) and the distributions associated with the players' unilateral deviations (P' and P''), is a precise formulation of what it means when the signal 'statistically distinguishes different players' deviations'.

Let us now generalize this observation. Given an action profile (for simplicity of exposition, assume it is pure) to be sustained, there is an associated signal distribution P^*. Consider any pair of players i and j, and let $|A_k|$ be the number of player k's actions ($k = i, j$) in the stage game. Since each player $k = i, j$ has $|A_k| - 1$ ways to deviate, we have $|A_i| + |Aj| - 2$ signal distributions associated with their unilateral deviations. If those distributions and the equilibrium distribution P^*, altogether $|A_i| + |Aj| - 1$ vectors, are linearly independent, we say that the signal can discriminate between deviations by i and deviations by j. This is called the *pairwise full rank condition.* This holds only when the dimension of the signal space ($|\Omega|$, the number of signal outcomes) is larger than the number of those vectors (that is, $|A_i| + |Aj| - 1$). Conversely, if this inequality is satisfied, the pairwise full rank condition holds 'generically' (that is, it holds unless the signal distributions have a very special structure, such as exact symmetry). This leads us to the folk theorem under imperfect public monitoring (this is a restatement of Fudenberg et al. 1994, in terms of genericity):

Folk Theorem Under Imperfect Public Monitoring Suppose that the signal space is large enough in the sense that $|\Omega| \geq |A_i| + |Aj|$ 1 holds for each pair of players i and j. Then, for a generic choice of the signal distributions and the stage game, any feasible and individually rational payoff profile $v \in V^*$ can be asymptotically achieved by a sequential equilibrium as the discount factor δ tends to 1.

In contrast to the perfect monitoring case, the proof is non-constructive. Rather than explicitly constructing equilibrium strategies, the theorem is proved by showing that any smooth subset of V^* is self-generating. In fact, the exact structure of the equilibrium strategy profile to sustain, for example, an efficient point is not so well understood. Sannikov (2005) shows that detailed structure of equilibrium strategies can be obtained if the model is formulated in continuous time.

Imperfect Private Monitoring

Now consider the case where all players receive a private signal about their opponents' actions. Although this has a number of important applications (a leading example is the secret price cutting model), this part of research is still in its infancy. Hence, rather than just summarizing definitive results as in the previous subsections, I explain in somewhat more technical detail the source of difficulties and the nature of existing approaches.

The difficulties come from a subtle but crucial difference from the perfect or public monitoring case. I explain below the difference from three viewpoints, in the increasing order of technicality.

1. In the perfect or public monitoring case, players share a mutual understanding about when and whom to punish. They can cooperate to implement a specific punishment, and, more importantly, they can mutually provide the incentives to carry out the punishment. This convenient feature is lost when players have diverse private information about each others' actions.
2. In the perfect or public monitoring case, public information directly tells the opponents' future

action plans. In the private monitoring case, however, each player has to draw statistical inferences about the history of the opponents' private signals to estimate what they are going to do. The inferences quickly become complicated over time, even if players adopt relatively simple strategies.

3. In the perfect or public monitoring case, the set of equilibria has a recursive structure, in the sense that a Nash equilibrium of the repeated game is always played after any history. Now consider a Nash equilibrium of, for example, the repeated Prisoner's Dilemma with imperfect private monitoring. After the equilibrium actions in the first period, say (C, C), players condition their action plans on their private signals ω_1 and ω_2. Hence the continuation play is a *correlated equilibrium,* where it is common knowledge that the probability distribution of the correlation device (ω_1, ω_2) is given by $p(\omega_1, \omega_2|C, C)$. When player 1 deviates to D in the first period, however, the distribution of correlation device is *not* common knowledge: player 1 knows that it is $p(\omega_1, \omega_2|D, C)$, while player 2 keeps the equilibrium expectation $p(\omega_1, \omega_2|C, C)$. Hence, after a deviation, the continuation play is no longer a correlated equilibrium in the usual sense. In addition, the space of the correlation device (the history of private signals) becomes increasingly rich over time. Therefore, the equilibria in the private monitoring case do not have a compact recursive structure; a continuation play is chosen from a different set, depending on the history.

One way to get around these problems is to allow communication (Compte 1998; Kandori and Matsushima 1998). In their equilibrium, players truthfully communicate their private signal outcomes in each period. The equilibrium is constructed in such a way that each player's report of her signal is utilized to discipline *other* players and does *not affect one's own continuation payoff.* This implies that each player is indifferent about what to report, and therefore truth telling is *a* best reply. Such an equilibrium, which depends on the history of publicly observable messages, works in much the same way as the equilibria in the public monitoring case. Hence, with communication, the folk theorem is obtained in the private monitoring case.

The remaining issue is to characterize the equilibria in the private monitoring case without communication. From the viewpoint of potential applications, this is important, because collusion or cartel enforcement is a major applied area of repeated games, where communication is explicitly prohibited by the antitrust law.

One may expect that, when players' private information admits sufficient positive correlation, an equilibrium can be constructed in a similar way to the public monitoring case. Sekiguchi (1997) is the first to construct a non-trivial (and nearly efficient) equilibrium in the private monitoring game without communication, and his construction is basically built on such an idea. Strong correlation of private information is, however, not assumed in his model but is derived endogenously. He assumes that private signals provide nearly perfect observability and considered *mixed* strategies. In such a situation, the privately observed random variables, the action-signal pairs, are strongly correlated (because a player's random action is strongly correlated with another player's signal under nearly perfect observability). Mailath and Morris (2002) show that, in general, there is 'continuity' between the public and private but sufficiently correlated monitoring cases, in the sense that any strategy with a *finite memory* works in either case.

Those papers are examples of the *belief-based approach,* which directly addresses the statistical inference problem (see point 2. above). Some other papers follow this approach, and they provide judiciously constructed strategies in rather specific examples, where the inference problem becomes tractable. Aside from the case with near perfect correlation, however, we are yet to have generally applicable results or techniques from this approach.

More successful has been the *belief-free approach,* where an equilibrium is constructed in such a way that the inference problem becomes *irrelevant.* As a leading example, here I explain Ely and Valimaki's work (2002) on the repeated

Prisoner's Dilemma with imperfect private monitoring. Each player's strategy is a Markov chain with two states, R (reward) and P (punishment). A specific action is played in each state (C in R, and D in P), and the transition probabilities between the states depend on the realization of the player's private signal. Choose those transition probabilities in such a way that the *opponent* is always indifferent between C and D *no matter which state the player is in*. This requirement can be expressed as a simple system of dynamic programming equations, which has a solution when the discount factor is close to 1 and the private signal is not too uninformative. By construction, any action choice is optimal against this strategy after any history, and in particular this strategy is a best reply to itself (so that it constitutes an equilibrium). Note that one's incentives do not depend on the opponent's state, and therefore one does not have to draw the statistical inferences about the history of the opponent's private signals.

There are certain difficulties, however, in obtaining the folk theorem with such a class of equilibria. First, players may be punished simultaneously in this construction, and our discussion about the public monitoring case shows that some welfare loss is inevitable (unless monitoring is nearly perfect). Second, even if we restrict our attention to the nearly perfect monitoring case, there is a certain set of restrictions imposed on the action profiles that can be sustained by such a belief-free equilibrium.

Those difficulties can be resolved when we consider *block strategies*. Block strategies treat the stage games in T consecutive periods as if they were a single stage game, or a block stage game, and applies the belief-free approach with respect to those block stage games. It is now known that, by using the block strategies, the folk theorem under private monitoring holds in the nearly perfect monitoring case (Hörner and Olszewski 2006) and for some two-player games where monitoring is far from perfect (Matsushima 2004). In the former, the block structure of the stage game helps to satisfy the restrictions imposed on the sustainable actions in belief-free equilibria. In the latter, an equilibrium is constructed where players choose constant actions in each block. This means that players have T samples of private signals for the constant actions, so that the observability practically becomes nearly perfect when T is large. With this increased observability and some restrictions on payoff functions, the folk theorem is obtained. For this construction to be feasible, the signals have to satisfy certain strong conditions, such as independence (across players).

The general folk theorem, or a general characterization of equilibria, for the private monitoring case is yet to be obtained, and it remains an important open question in economic theory. A comprehensive technical exposition of the perfect monitoring, imperfect public monitoring, and private monitoring cases can be found in Mailath and Samuelson (2006).

See Also

▶ Cartels
▶ Cooperation
▶ Reputation
▶ Social Norms

Bibliography

Abreu, D., D. Pearce, and E. Stacchetti. 1990. Towards a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58: 1041–1064.

Abreu, D., P. Dutta, and L. Smith. 1994. The folk theorem for repeated games: A NEU condition. *Econometrica* 62: 939–948.

Aumann, R. 1959. Acceptable points in general cooperative N-person games. In *Contributions to the theory of games*, ed. R.D. Luce and A.W. Tucker, vol. 4. Princeton: Princeton University Press.

Axelrod, R. 1984. *Evolution of cooperation*. New York: Basic Books.

Benoit, J.P., and V. Krishna. 1985. Finitely repeated games. *Econometrica* 53: 905–922.

Coase, R. 1937. The nature of the firm. *Economica* n.s. 4: 386–405.

Compte, O. 1998. Communication in repeated games with imperfect private monitoring. *Econometrica* 66: 597–626.

Ely, J., and J. Valimaki. 2002. A robust folk theorem for the Prisoner's Dilemma. *Journal of Economic Theory* 102: 84–105.

Friedman, J. 1971. A non-cooperative equilibrium for supergames. *Review of Economic Studies* 38: 1–12.

Fudenberg, D., and E. Maskin. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54: 533–554.
Fudenberg, D., D. Levine, and E. Maskin. 1994. The folk theorem with imperfect public information. *Econometrica* 62: 997–1040.
Hörner, J., and W. Olszewski. 2006. The folk theorem for games with private almost- perfect monitoring. *Econometrica* 74: 1499–1544.
Kandori, M., and H. Matsushima. 1998. Private observation, communication and collusion. *Econometrica* 66: 627–652.
Macaulay, S. 1963. Non-contractual relations in business: A preliminary study. *American Sociological Review* 28: 55–67.
Mailath, G., and S. Morris. 2002. Repeated games with imperfect private monitoring: Notes on a coordination perspective. *Journal of Economic Theory* 102: 189–228.
Mailath, G., and L. Samuelson. 2006. *Repeated games and reputations: Long-run relationships*. Oxford: Oxford University Press.
Matsushima, H. 2004. Repeated games with private monitoring: Two players. *Econometrica* 72: 823–852.
Pearce, D. 1990. Repeated games: Cooperation and rationality. In *Advances in economic theory*, ed. J. Laffont. Cambridge: Cambridge University Press.
Rubinstein, A. 1979. Equilibrium in supergames with overtaking criterion. *Journal of Economic Theory* 21: 1–9.
Sannikov, Y. 2005. *Games with imperfectly observable actions in continuous time*. Berkeley: Mimeo, University of California.
Sekiguchi, T. 1997. Efficiency in repeated Prisoner's Dilemma with private monitoring. *Journal of Economic Theory* 76: 345–361.

Replacement Policy

John J. McCall

Technological change, variable tastes and the consequent replacement of obsolete physical and human capital comprised the driving forces of the industrial revolution. They continue to be the most important dynamic elements of the modern fluctuating economy. The birth and death of firms and industries, the birth of job vacancies in new and expanding firms and industries, and the elimination of jobs in dying industries and bankrupt firms – these replacements are the essence of a modern capitalist economy. Critics of this system emphasize the mobility costs implied by this dynamism, whereas champions of the capitalist system marvel at the speed and efficiency of these adjustments. Our task is not to evaluate these positions or suggest optimal solutions to this grand replacement problem. We only mention it to remind the reader that 'replacement policy' when broadly construed is at the heart of any economic system and *that it includes both physical and human capital.* Our essay concentrates on the optimal replacement of stochastically failing equipment.

There are several fine books and surveys on capital theory beginning with I. Fisher (1930a) and including Hirshleifer (1970), Jorgenson (1977) and Nickell (1978). Becker (1964) is the standard reference for human capital theory. A remarkable coincidence is the simultaneous appearance in 1930 of I. Fisher's *Theory of Interest* and R.A. Fisher's *The Genetical Theory of Natural Selection.* In the former we learn when a growing asset (a tree) should be cut; in the latter we are shown how to calculate the discounted value of future offspring of individuals of age *x*.

Replacement Theory

Replacement theory resides at the centre of reliability theory, a vital area of applied probability that is a dynamic mechanism of microeconomics. The branch of applied probability that contains reliability is usually called operations research or management science. Inventory theory, queuing, and simulation are some of the other members of this discipline. Economists frequently cloak this vital discipline with an assortment of static 'production functions' thereby ignoring the very quick of the production process.

There are vast theoretical and applied literatures on both reliability theory and replacement policy. Fortunately, there are also several excellent surveys. The theory and statistical methods of reliability are surveyed in Thompson (1981), Bergman (1985), and Pierskalla and Voelker (1976) reviews the maintenance literature. The books, Barlow and Proschan (1965), Gnedenko et al. (1969), and Arrow et al. (1958, 1962) are classics.

From a probabilistic perspective replacement is a special topic in renewal theory. After a brief survey of renewal theory we turn to the economic problem – when should a piece of stochastically failing equipment be replaced. That is, at what point should the stochastic process be renewed. We use optimal stopping theory to answer this economic question. There are many versions of this problem. We consider three. A simple preventive maintenance model, a shock replacement policy, and an adaptive replacement policy. The concluding section observes that whereas economists sometimes overlook the dynamic stochastic aspects of production, engineers frequently tend to ignore incentive problems and miscalculate opportunity costs.

Renewal Theory and Counting Processes

Let $(T_i; i = 1,2,\ldots)$ be a sequence of independent identically distributed random variables with distribution function F. The T's are non-negative and T_j denotes the time between the jth and $(j+1)$st event (failure or replacement). The mean time m between successive events i and $i+1$ is

$$m = E[T_i] = \int_0^\infty s\,\mathrm{d}F(s)$$

with $0 < m \leqslant \infty$.

The time of the nth occurrence is denoted by Z_n, where

$$Z_0 = 0 \text{ and } Z_n = \sum_{i=1}^{n} T_i, \quad n = 1, 2, \ldots$$

Finally, the number of occurrences $N(t)$ by time t is the largest n such that the time of the nth occurrence is less than or equal to t, that is,

$$N(t) = \sup\{n : Z_n \leqslant t\}.$$

The stochastic counting process $N(t)$ is called a *renewal process.*

The expected value of $N(t)$ is called the renewal function $\rho(t)$,

$$\rho(t) = \sum_{n=1}^{\infty} F^n(t),$$

where F^n is the n-fold convolution of F, that is, $F^n(t) = p\{T_1 + T_2 + \cdots + T_n \leq t\}$.

A Simple Preventive Maintenance Model

In a world of certainty one would never replace a piece of equipment until it was just about to fail (unless there were economies of scale in the very act of replacing a number of items at the same time). If the cost of replacement before failure is less than replacement after failure, replacements would be scheduled at the instant before failure. However, if failures occur stochastically, it may pay to replace well in advance of a failure even when the decision-maker is risk neutral.

Consider the optimal preventive maintenance policy for a single piece of equipment with a time-to-failure distribution characterized by an increasing failure rate. If the distribution of time-to-failure exhibits an increasing failure rate (IFR), then, by definition, the conditional probability $h(t)$ of failure at any specified instant t, given that the equipment has not failed prior to that specified instant t, is an increasing function of t. In this sense, the equipment can be said to be wearing out. The lifetime of the equipment is a random variable X with distribution function F. The failure rate function h, given by $h(t) = f(t)/[1 - F(t)]$ for $t > 0$, is assumed to be increasing. The equipment is continuously monitored and is replaced by a new item whenever it fails or reaches age N. A new item costs K (dollars). The marginal cost of replacement before (after) failure is $\alpha_1(\alpha_2)$, $\alpha_2 > \alpha_1$. Replacement is instantaneous, and the planning horizon is infinite. Because there (presumably) will be infinitely many replacements over the infinite future, the risk neutral agent may minimize either the expected cost per unit time or the expected discounted costs. To simplify the analysis, the average expected cost criterion is adopted.

The optimal replacement policy is periodic. The expected average cost $A(N)$ of 'policy N' is the ratio of the expected cost $C(N)$ of a cycle to the expected cycle length $L(N)$, where a cycle is the time between replacements (renewals). Thus,

$$A(N) = C(N)/L(N), \quad N > 0, \tag{1}$$

where

$$C(N) = K + \alpha_2 F(N) + \alpha_1 [1 - F(N)] \tag{2}$$

and

$$L(N) = \int_0^N \mathrm{d}F(t) + N[1 - F(N)]. \tag{3}$$

To find the optimal value of N, differentiate A:

$$A'(N) = \{(\alpha_2 - \alpha_1) f(N) L(N) - [1 - F(N)] \times [K + \alpha_1 + (\alpha_2 - \alpha_1) F(N)]\}/L(N)^2. \tag{4}$$

The solution, N^* is the optimal replacement interval. If $N^* = \infty$ and $\alpha_2 > \alpha_1$, the item must have a constant or a decreasing failure rate. Thus, the increasing failure rate *and* the larger in-service replacement cost justify preventive ($N^* < \infty$) maintenance.

A Shock Model of Preventive Maintenance

An equipment is bombarded by a random sequence of shocks. The amount of damage caused by each shock is also a random variable. Any shock can cause the equipment to fail, but the probability of failure by a shock at t is monotone increasing in the accumulated damage at t. Replacement is restricted to shock times. The accumulated damage is determined by a semi-Markov process, that is, the probability of moving from damage state i to damage state j is given by P_{ij}, $0 < i < j$, $i, j \in (1,2,3, \ldots)$ with F_{ij} the waiting time from i given that j is the next state.

The replacement decision can be formulated as an optimal stopping problem. The solution has the following *'control limit'* structure: there is a critical number ξ such that the optimal replacement policy is the optimal stopping time $N_\xi = \inf(n \in N\colon X_n \geqslant \xi)$, where X_n is the cumulative damage at n. This structure is identical to that of those search models possessing the *'reservation wage' property.* See Lippman and McCall, volume 1 (1986).

Adaptive Replacement Policies

The critical numbers characterizing an optimal replacement policy can be estimated when F is unknown. For moderate discount rates the policy designed by Fox (1965) is optimal, whereas for discount rules close to one stochastic approximation can be applied (Frees and Ruppert 1985).

Conclusion

In this brief essay it is impossible to mention let alone exposit all the significant aspects of replacement policy. There are three of such paramount importance that their absence would eviscerate the essay. First, the incentive problem is manifest. The organization must be designed so that the 'true' opportunity costs of a replacement relative to a failure are correctly transmitted across all levels. For example, information about the state of an aircraft's components flows back and forth between the operations personnel and the maintenance crew. The repair activity may be hierarchical so that the information must cross several levels before it reaches the actual repair crew. Furthermore, maintenance is embedded in the overall production and repair activity that includes inventory control, production scheduling, queuing and transportation. The incentive structure must be such that management's assessment of the relative maintenance costs is reflected in the behaviour of the operations personnel, the information flow, and the work effort of the maintenance crew and its affiliated network.

The existence of this incentive problem reminds us that replacement is essentially a topic in the theory of insurance. Thus all of the difficulties attending the insurance activity and their practical resolution are pertinent. Finally, preventive medicine is one of the most important applications of replacement theory. Thus, while optimal replacement may appear to be a narrow and routine technical problem, it is in fact quite broad, ranging from extinction to the body's replacement of red blood cells, and its practical applications are riddled with complex incentive problems. The technical problems also are profound, entailing the basic physics of the deterioration process.

See Also

▶ Birth-and-Death Processes
▶ Inventory Policy Under Certainty

Bibliography

Arrow, K.J., S. Karlin, and H. Scarf. 1958. *Studies in the mathematical theory of inventory and production*. Stanford: Stanford University Press.

Arrow, K.J., S. Karlin, and H. Scarf (eds.). 1962. *Studies in applied probability and management science*. Stanford: Stanford University Press.

Barlow, R.E., and F. Proschan. 1965. *Mathematical theory of reliability*. New York: Wiley.

Barlow, R.E., and F. Proschan. 1975. *Statistical theory of reliability and life testing*. New York: Holt, Rinehart & Winston.

Becker, G.S. 1964. *Human capital*. New York: Columbia University Press for the National Bureau of Economic Research.

Bergman, B. 1985. On reliability and its applications. *Scandinavian Journal of Statistics* 12: 1–42.

Derman, C. 1963. Ch. 9: On optimal replacement rules when changes of state are Markovian. In *Mathematical optimization techniques*, ed. R. Bellman. Berkeley/Los Angeles: University of California Press.

Derman, C., and J. Sacks. 1960. Replacement of periodically inspected equipment. *Naval Research Logistics Quarterly* 7: 597–607.

Fisher, I. 1930a. *The theory of interest*. New York: Macmillan.

Fisher, R.A. 1930b. *The genetical theory of natural selection*. Oxford: Oxford University Press.

Fox, B. 1965. *An adaptive age replacement policy* (Report No. 65–17(RR)). Berkeley: Operations Research Center.

Gnedenko, B.V., Yu.K. Belyayev, and A.D. Solouyev. 1969. *Mathematical methods of reliability theory.* New York: Academic.

Hirshleifer, J. 1970. *Investment, interest, and capital*. Englewood Cliffs: Prentice-Hall.

Jorgenson, D.W. 1971. Econometric studies of investment behavior: A survey. *Journal of Economic Literature* 9: 1111–1147.

Jorgenson, D.W., J.J. McCall, and R. Radner. 1967. *Optimal replacement policy.* Amsterdam: North-Holland.

Lippman, S.A., and J.J. McCall. 1986. *The economics of search*, vol. 1. Oxford: Basil Blackwell.

Lotka, A.J. 1939. A contribution to the theory of self-renewing aggregates with special reference to industrial replacement. *Annals of Mathematical Statistics* 10: 1–25.

Nickell, S.J. 1978. *The investment decisions of firms*. Cambridge: Cambridge University Press.

Pierskalla, W.P., and J.A. Voelker. 1976. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly* 23: 353–388.

Smith, W.L. 1958. Renewal theory and its ramifications. *Journal of the Royal Statistical Society, Series B* 20: 243–302.

Thompson Jr., W.A. 1981. On the foundations of reliability. *Technometrics* 23: 1–13.

Representation of Preferences

Peter C. Fishburn

Three facets of subjective preferences have played central roles in economics. They are the qualitative structure of an agent's preferences, numerical representations of preferences, and the use of numerical representations or utility functions in economic analysis. We consider various representations and their ties to qualitative preference structures.

Preferences themselves are described by a binary relation $\succ$ *is preferred to,* on a nonempty set X. Axioms or assumptions about the behaviour of $\succ$ on X identify a qualitative preference structure. A representation provides a correspondence between this structure and properties of real valued functions based on X.

Elements in X are often viewed as decision alternatives or outcomes of choice. They may be arbitrary or have a prescribed structure, as when each x in X is a commodity bundle in some Euclidean space or a probability distribution (lottery) on wealth or on increments to current wealth.

Representations of preferences between lotteries date to Bernoulli (1738), who sought to explain why agents often prefer a sure level of wealth to a lottery with larger expected value than the sure level. Representations of preferences between commodity bundles were used by Jevons, Menger, Walras and Edgeworth in the late 19th century to examine the economic consequences of consumer behaviour within the theory of marginal analysis (Samuelson 1947; Stigler 1950). The commodity space and lottery contexts remain the preeminent structures for research in the representation of preferences.

Despite the early beginnings, detailed attention to qualitative preference structures for various representations is comparatively recent. Three examples are Frisch's axiomatization of comparable preference differences and Ramsey's theory of utility and subjective probability for decisions under uncertainty from the 1920s, and the axioms for expected utility of von Neumann and Morgenstern (1944). For general discussion of representations of various types, *see* MEASUREMENT, THEORY OF.

Classification

The structure of X and the degree of transitivity are two useful factors for classifying representations. The five most prominent structures are

S1. X is arbitrary except perhaps for cardinality or topological properties,
S2. $X = X_1 \times X_2 \times \cdots \times X_n$ or $X_1 \times X_2 \times \ldots,$
S3. X is a set of probability distributions,
S4. same as S3 except the outcomes are multidimensional as in S2,
S5. X is a set of mappings from a set S of states into an outcome set.

S2 includes commodity spaces and time streams, S3 is the setting for expected utility, and S5 is Savage's (1954) formulation for decisions under uncertainty. When S is countable, S2 can be used instead of S5 with X_i the possible outcomes for state i.

Our discussion is organized around S1–S5 with transitivity as a subsidiary factor. It is assumed in all cases that $\succ$ is asymmetric, so $x \succ y$ precludes $y \succ x$. The agent's indifference relation $\sim$ and preference-or-indifference relation are defined by

$$\begin{aligned} x \sim y, &\quad \text{if neither } x \succ y \text{ nor } y \succ x, \\ x \succ y, &\quad \text{if } x \succ y \text{ or } x - y. \end{aligned}$$

Comparable preferences differences are discussed in the final section.

The preference relation $\succ$ is *transitive* if, for all x, y and z in X, $x \succ z$ whenever $x \succ y$ and $y \succ z$. Similar definitions pertain to $\sim$ and $\succsim$. Three levels of transitivity are

T1. both $\succ$ and $\sim$ are transitive,
T2. only $\succ$ is assumed to be transitive,
T3. neither $\succ$ nor $\sim$ is assumed to be transitive.

T1 is the usual assumption employed in economic analysis. T2 has little relevance before 1960 and T3 has little relevance before 1970.

T1 implies that $\succsim$ also is transitive. Under T1, $\succ$ is a 'weak order', and $\succsim$ is a 'weak order' or 'complete preorder'. T2 says that preferences are partially ordered; indifference need not be transitive. I include acyclic preferences–it is never true that $x_1 \succ x_2 \succ \cdots \succ x_k \succ x_1$ – under T2. Unordered or non-transitive preferences fall under T3, which allows preference cycles such as $x \succ y, y \succ z$, and $z \succ x$, or $x \succ y \succ z \succ x$ for short.

Arbitrary Sets: S1 The basic representation for weak orders is

$$x \succ y \Leftrightarrow u(x) > u(y), \tag{1}$$

where u is a real function on X. This and later expressions apply to all x, y, ... in X. The function

u in (1) is unique up to transformations that preserve order and is called an *ordinal* utility function.

T1 is necessary and sufficient for (1) when X is countable, but not otherwise. The general case also requires X to have a countable *order – dense* subset Y (Cantor 1895) such that, whenever then $x \succ z$ then $x \succsim y \succsim z$ for some y in Y. When (1) fails under T1 because no countable subset is order-dense in X, $\succ$ can be represented by vectors of utilities ordered lexicographically (Chipman 1960; Fishburn 1974). The finite-dimensional lexicographic representation is

$$x \succ y \Leftrightarrow [u_1(x), \ldots, u_n(x)] >_{\mathrm{L}} [u_1(y), \ldots, u_n(y)], \tag{2}$$

where each u_i is real valued and $(a_1, \ldots, a_n) >_{\mathrm{L}} (b_1, \ldots, b_n)$ if $a_i \neq b_i$ for some i and $a_i > b_i$ for the smallest such i.

Other conditions than order denseness can be used for (1) when X is a topological space. If X is connected and separable and T1 holds, there is a continuous u that satisfies (1) if, for each y in X, $\{x : x \succ y\}$; and $\{x : y \succ x\}$ are open sets in X's topology. This and related contributions on continuity appear in Debreu (1964) and Fishburn (1970a).

Under T2, (1) is replaced by the one-way representation

$$x \succ y \Rightarrow u(x) > u(y). \tag{3}$$

T2 is sufficient for (3) when X is countable, but not otherwise (Fishburn 1970a, b). For comments on continuity, see Sondermann (1980). Specialized partial orders use two functions for two-way representations. For example, if X is countable and $\{x \succ a, y \succ b\}$ implies $x \succ b$ or $y \succ a$ then there are real functions f and $\rho > 0$ on X such that

$$x \succ y \Leftrightarrow f(x) > f(y) + \rho(y).$$

Such an $(X, \succ)$ is called an *interval order*. The more specialized case in which ρ is a positive constant is known as a *semiorder;* see Fishburn (1985) for details.

Under T3, $(X, \succ)$ can be represented by a skew-symmetric $[\phi(y, x) = -\phi(x, y)]$ real function ϕ on $X \times X$ as

$$x \succ y \Leftrightarrow \phi(x, y) > 0. \tag{4}$$

This requires only asymmetry, and $\phi(x, y)$ can be set equal to 1, 0 or -1 when $x \succ y, x \sim y$ *or* $y \succ x$ respectively. We can view (1) as the specialization of (4) in which

$$\phi(x, y) = u(x) - u(y).$$

Product Sets: S2 When $X = X_1 \times X_2 \times \cdots \times X_n$ with $x = (x_1, x_2, \ldots, x_n)$, (1) is

$$x \succ y \Leftrightarrow u(x_1, x_2, \ldots, x_n) > u(y_1, y_2, \ldots, y_n) \tag{1 *}$$

It is often assumed that each X_i is a real interval or a convex subset of a connected and separable topological space, and that u increases and is continuous in each component (Debreu 1964; Fishburn 1970a). When X is the positive orthant of n-dimensional Euclidean space, the indifference classes form a layered array of isoutility contours away from the origin. Isoutility contours that are convex to the origin are often presumed in the marginal analysis of consumption theory.

Houthakker (1961) provides a survey of consumption theory, including the fundamentals of demand as a function of prices and income, revealed preference, direct utility as a function of commodity bundles, and indirect utility. An example of the indirect approach, which expresses utility as a function of prices $p_1, \ldots, p_n$ and total expenditure $\mathrm{m} > 0$, m is the indirect addilog function

$$v(p_1/m, \ldots, p_n/m) = \sum_{i=1}^{n} a_i (p_i/m)^{bi}.$$

A related direct addilog function for quantities $q_1, \ldots, q_n$ is

$$u(q_1, \ldots, q_n) = \sum_{i=1}^{n} \alpha_i {q_i}^{\beta i}.$$

These functions are special cases of the additive-utility specialization of (1), i.e.

$$x \succ y \Leftrightarrow u_1(x_1) + \cdots + u_n(x_n) > u_1(y_1) + \cdots + u_n(y_n), \quad (5)$$

where u_i is a real function on X_i. This presumes the *independence condition* which says that, whenever the n factors are partitioned into two parts, the preference order over one part conditioned on fixed values of the X_i in the other part is independent of the particular fixed values used. Other axioms are also needed for (5).

Necessary and sufficient conditions for (5) and its one-way counterpart under T2 when X is finite appear in Fishburn (1970a) and Krantz et al. (1971). Conditions sufficient for (5) with infinite X_i appear in Fishburn (1970a) and Krantz et al. (1971). The latter conditions imply that the u_i in (5) are unique up to similar positive linear transformations, so that $v_1, \ldots, v_n$ satisfy (5) in place of $u_1, \ldots, u_n$ if and only if there are numbers $\beta_1, \ldots, \beta_n$ and $\alpha > 0$ such that

$$v_i(x_i) = \alpha u_i(x_i) + \beta_i, \quad \text{for all } i \text{ and all } x_i \text{ in } X_i.$$

Debreu and Koopmans (1982) study the conjunction of additive utilities and quasiconcavity of u when $u(x) = u_1(x_1) + \cdots + u_n(x_n)$.

The basic lexicographic representation for $X = X_1 \times X_2 \times \cdots$ (Chipman 1960) with hierarchical importance ordering $1, 2, \ldots$ is

$$x \succ y \Leftrightarrow [u_1(x_1), u_2(x_2), \ldots] >_L [u_1(y_1), u_2(y_2), \ldots].$$

Luce (1978) combines the lexicographic and additive ideas in a two-factor model whose lexicographic part applies under significant differences in the dominant factor, and whose additive part applies otherwise.

One T3 representation is the additive-difference model in Tversky (1969) where

$$x \succ y \Leftrightarrow \sum_{i=1}^{n} f_i[u_i(x_i) - u_i(y_i)] > 0.$$

This is a special case of (4). Here f_i is an odd $[f_i(-a) = -f_i(a)]$, continuous and increasing real function. Fishburn (1980) combines the additive-difference and lexicographic notions.

Other T3 representations are implicit in Mas-Colell (1974) and elsewhere in a topological setting. A key axiom in this work is that $\{y : y \succ x\}$ is a convex subset of X for each x in X.

The homogeneous case $X = A^n$ *or* $X = A \times A \times \cdots$ provides a setting for time preference. Notions of persistence, impatience and stationarity for denumerable-period contexts are analysed by Koopmans (1960), Koopmans et al. (1964), and Fishburn and Rubinstein (1982). Fishburn (1970a) considers finite periods. One representation here is

$$x \succ y \Leftrightarrow \pi_1 u(x_1) + \pi_2 u(x_2) + \cdots > \pi_1 u(y_1) + \pi_2 u(y_2) + \ldots, \quad (6)$$

where $\pi_i \geq 0$ is an importantance of weight for period i. A particular cases is $\pi_i = \sigma^{i-1}$ which obtains for the additive model (5) if preferences are ‘stationary’.

Probability Distributions: S3. This section and the next assume that X is a convex set of probability measures (distributions) on an outcome algebra $\mathscr{A}$. $x(A)$ is the probability that x yields an outcome in set A. When $0 \leq \lambda \leq 1$ and x and y are in X, $\lambda x = (1-\lambda)y$ denotes of linear convex combination of x and y that has $[\lambda x + (1-\lambda)y](A) = \lambda x(A) + (1-\lambda)y(A)$ each A in $\mathscr{A}$. We say that $(X, \succ)$ is linear if has $\lambda x + (1-\lambda)z \succ \lambda y + (1-\lambda)z$ whenever $x \succ y$, z is in X, and $0 < \lambda < 1$.

The von Neumann and Morgenstern theory (Fishburn 1970a, 1982a) uses T1, linearity and a continuity condition to obtain a u on X that satisfies (1) and

$$u[\lambda x + (1-\lambda)y] = \lambda u(x) + (1-\lambda)u(y). \quad (7)$$

Such a u is unique up to an arbitrary positive linear transformation $[v(x) = \alpha u(x) - \beta, \alpha > 0]$ and is sometimes called a *cardinal utility function.* If the outcome set C is finite and $\mathscr{A}$ includes each singleton $\{c\}$ for c in C, then (7) yields the expected-utility form

$$u(x) = \sum_C x(c)u(c).$$

The extension of this to $u(x) = \int u(c)\mathrm{d}x(c)$ for infinite outcome sets is discussed in Fishburn (1970a, 1982a).

Generalizations of the von Neumann and Morgenstern theory that retain T1 but weaken the linearity axiom are discussed by Allais (1953), Kahneman and Tversky (1979), Machina (1982), Chew (1983), and Fishburn (1983). For example, Machina assumes a smooth preference field over X that is approximately linear locally, and Chew and Fishburn axiomatize the representation

$$x \succ y \Leftrightarrow u(x)w(y) > u(y)w(x), \qquad (8)$$

in which each of u and w is linear, as in (7), and w is non-negative. When w is constant, (8) reduces to the von Neumann and Morgenstern case.

Generalizations that retain linearity but weaken T1–T2 appear in Fishburn (1970a, 1982a). Other generalizations retain T1 and linearity but drop continuity to obtain lexicographic expected utility representations (Chipman 1960; Fishburn 1982a).

Axioms for unordered and nonlinear preferences over X are presented in Fishburn (1982b). Assumptions of continuity, convexity–such as $x \succ y$ and $z \succ y$ imply $\lambda x + (1-\lambda)z \succ y$, and symmetry are shown to be necessary and sufficient for the unordered representation (4) in which ϕ on $X \times X$ is skew-symmetric and *bilinear,* that is linear separately in each argument. Such a ϕ is unique up to an arbitrary similarity transformation of the form $\phi'(x,y) = x\phi(x,y)$ with $\alpha > 0$. The von Neumann and Morgenstern model results when ϕ can be decomposed as $\phi(x,y) = u(x) - u(y)$, and (8) corresponds $\phi(x,y) = u(x)w(y) - u(y)w(x)$.

Multiple Attributes Under Risk: S4 Continuing with X as a set of probability distributions, we now assume that the outcome set C is a product set with $C = C_1 \times C_2 \times \cdots \times C_n$. We assume also that the basic expected utility axioms hold for $(x, \succ)$, so that, for all distributions x and y in X with finite supports,

$$\begin{aligned} x \succ y \Leftrightarrow & \sum_C x(c_1,\ldots,c_n)u(c_1,\ldots,c_n) \\ & > \sum_C y(c_1,\ldots,c_n)u(c_1,\ldots,c_n). \end{aligned} \qquad (9)$$

A generalization of (9) is noted at the end of the section.

Representation (9) has several specializations involving decompositions of $u(c_1, \ldots, c_n)$. Many of these are reviewed in Keeney and Raiffa (1976), Fishburn (1977), and Farquhar (1978).

The additive decomposition is

$$u(c_1,\ldots,c_n) = u_1(c_1) + \ldots + u_n(c_n), \qquad (10)$$

in which u_i is a real function on C_i and the u_i are unique up to similar positive linear transformations. When (10) holds, the sum in (9) simplifies to

$$\sum_C x(c_1,\ldots,c_n)u(c_1,\ldots,c_n) = \sum_{i=1}^{n}\sum_{C_i} x_i(c_i)u_i(c_i),$$

where x_i denotes the marginal distribution of x on C_i. Given (9), a necessary and sufficient condition for (10) is $x \sim y$ whenever $x_i = y_i$ *for* $i = 1, \ldots, n$. The same result holds (Fishburn (1982a)) when C is only assumed to be a subset of $C_1 \times \cdots \times C_n$, but in this case the preceding uniqueness property may fail.

Multiplicative decompositions of $u(c_1, \ldots, c_n)$ arise from independence conditions that are similar to the condition following (5). For any non-trivial two-part partition $\{I, J\}$ of $\{1, 2, \ldots, n\}$ we say that I is *utility independent* of J if the preference order over distributions on the product of the C_i for i in I, conditioned on fixed values of the C_i for all j in J, is independent of those fixed values. Moreover, I is *generalized utility independent* of J if any two such conditional preference orders are identical, duals, or one is empty. The importance of these notions is that if I is generalized utility independent of J then $u(c_i, c_j)$ where c_j

is in the product of the C_i for $i \in I$ and similarly for c_j, decomposes as

$$u(c_I, c_J) = f(c_J) + g(c_J)h(c_I).$$

If I is utility independent of J, then g is a strictly positive function.

We mention one consequence of this two-part decomposition. If $\{1, \ldots i-1, i+1, \ldots, n\}$ generalized utility independent of $\{i\}$ for each i in $\{1, \ldots, n\}$, then there is a real function u_i on C_i for each i such that either (10) holds or there is a non-zero constant k such that, under a suitable rescaling of u,

$$ku(c_1, \ldots, c_n) + 1 = [ku_1(c_1) + 1] \ldots [ku_n(c_n) + 1].$$

Other types of independence among factors in the context of (9) are analysed by Farquhar (1975), Keeney and Raiffa (1976), and Fishburn and Farquhar (1982).

States of the World: S5 In our final setting, X is a set of functions from a set S of states of the world into a set C of outcomes or consequences. The set C may be unstructured or have one of the forms considered previously. For example, C could be a set of probability distributions defined on another set.

Following Savage (1954), we refer to each x in X as an *act* and to each subset of S as an *event.* It is presumed that exactly one state in S is the true state and that the agent is uncertain as to which state this is. Moreover, the true state, or state that 'obtains', is determined by circumstances beyond the agent's control. If the agent chooses act x and state s obtains, then $x(s)$ in C is the consequence that occurs as the result of the choice.

The best-known representation for $(X, \succ)$ is Savage's (1954) subjective expected utility model, which was inspired by Ramsey's earlier outline of a theory of preferences and beliefs under uncertainty, de Finetti's (1937) work in subjective probability, and the von Neumann and Morgenstern (1944) theory of expected utility. Savage's representation is

$$x \succ y \Leftrightarrow \int_s u\Big[x(s)dP(s) > \int_s u[y(s)dP(s), \quad (11)$$

where u is a real function on C and P is a finitely-additive probability measure on the set of all events that is 'continuously divisible' in the sense that, for every event A and every $0 < \lambda < 1$ there is another event $B \subseteq A$ for which $P(B) = \lambda P(A)$. In addition, u is bounded (Fishburn 1970a) and unique up to a positive linear transformation, and P is unique.

Savage's assumptions include strong structural conditions on X, T1, a few independence axioms, and a continuity condition that generates the form of P noted above. Criticisms of his conditions and alternative ways of conceptualizing decisions under uncertainty have stimulated a number of people to develop alternatives to Savage's theory with representations that are more or less similar to (11). The alternatives are reviewed in detail in Fishburn (1981).

Several authors (see Fishburn 1981) derive Savage's representation for finite as well as infinite S by taking C as a set of probability distributions or lotteries. The same device is used extensively in Fishburn (1982a), which includes a one-way representation under T2. Schmeidler (1984) keeps T1 in the lottery approach but weakens independence to obtain a representation with monotonic but non-additive 'probabilities' that accommodates preference patterns that are inconsistent with Savage's theory (Ellsberg 1961).

Loomes and Sugden (1982) propose a finite-state model for decision under uncertainty that allows non-transitive preferences and therefore falls in transitivity class T3. Their representation for n states is

$$x \succ y \Leftrightarrow \sum_{i=1}^{n} P(S_i)\phi[x(S_i), y(S_i)] > 0,$$

where $P(s_i)$ is the agent's subjective probability for states s_i and ϕ is skew-symmetric. Lottery-based axioms for this and other T3 models appear in Fishburn (1984).

Comparable Preference Differences In contrast to preceding representations, we now consider representations based on a binary relation $\succ *$ on $X \times X$. A common intuitive interpretation of $\succ *$ is that $(x, y) \succ *(z, w)$ signifies that the difference in preference between x and y exceeds the difference in preference between z and w, or that the intensity of preference for x over y exceeds the intensity of preference for z over w. When $\succ *$ is used, $\succ$ is usually defined by $x \succ y$ if only need if $(x, y) \succ *(y, y)$.

A basic representation in this setting is

$$(x, y) \succ *(z, w) \Leftrightarrow u(x) - u(y) > u(z) - u(w). \quad (12)$$

This requires $\succ *$ to be weak order on $X \times X$ and entries other conditions like

$$(x, y) \succ *(z, w) \Leftrightarrow (x, z) \succ *(y, w) \text{ and } (x, y) \succ *(z, y) \Leftrightarrow (x, w) \succ *(z, w).$$

Early axiomatizations of (12) from the 1920s and 1930s are due to Frisch, Lange and Alt. Comments on these and more recent axiomatizations appear in Fishburn (1970a, Chap. 6). The axioms for infinite X essentially use a bisection procedure to determine utility midpoints, as when $u(x) - u(z) = u(z) - u(y)$, or $u(z) = [u(x) + u(y)]/2$, and the resultant u is unique up to a positive linear transformation.

Specialized representations in the context of (12) arise when $X = X_1 \times \cdots \times X_n$. Dyer and Sarin (1979) and Kirkwood and Sarin (1980) consider decompositions of $u(x_1, \ldots x_n)$ that are similar to ones mentioned under S4, and Fishburn (1970a) discusses the weighted additive form used in (6) when $X = A^n$.

See Also

- Preferences
- Social Choice
- Utility Theory and Decision Theory

Bibliography

Allais, M. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Americaine. *Econometrica* 21: 503–546.

Bernoulli, D. 1738. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientarium Imperialis Petropolitanae* 5: 175–192. Trans. Sommer, L. 1954. *Econometrica* 22: 23–36.

Cantor, G. 1895. Beiträge zur Begründung der Transfinite Mengenlehre. *Mathematische Annalen* 46: 481–512. and 49 (1897), 207–246. English translation: *Contributions to the founding of the theory of transfinite numbers*. New York: Dover, n.d.

Chew, S.H. 1983. A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* 51: 1065–1092.

Chipman, J.S. 1960. The foundations of utility. *Econometrica* 28: 193–224.

Debreu, G. 1964. Continuity properties of Paretian utility. *International Economic Review* 5: 285–293.

Debreu, G., and T.C. Koopmans. 1982. Additively decomposed quasiconvex functions. *Mathematical Programming* 24: 1–38.

de Finetti, B. 1937. La prévision: ses logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7: 1–68. Trans. Kyburg, H.E. 1964. *Studies in subjective probability*, ed. H.E. Kyburg and H.E. Smokler, 93–158. New York: Wiley.

Dyer, J.S., and R.K. Sarin. 1979. Measurable multiattribute value functions. *Operations Research* 27: 810–822.

Ellsberg, D. 1961. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* 75: 643–669.

Farquhar, P.H. 1975. A fractional hypercube decomposition theorem for multiattribute utility functions. *Operations Research* 23: 941–967.

Farquhar, P.H. 1978. Interdependent criteria in utility analysis. In *Multiple criteria problem solving*, ed. S. Zionts, 131–180. Berlin: Springer.

Fishburn, P.C. 1970a. *Utility theory for decision making*. New York: Wiley.

Fishburn, P.C. 1970b. Intransitive indifference in preference theory: A survey. *Operations Research* 18: 207–228.

Fishburn, P.C. 1974. Lexicographic orders, utilities, and decision rules: A survey. *Management Science* 20: 1442–1471.

Fishburn, P.C. 1977. Multiattribute utilities in expected utility theory. In *Conflicting objectives in decisions*, ed. D.E. Bell, R.L. Keeney, and H. Raiffa, 172–194. New York: Wiley.

Fishburn, P.C. 1980. Lexicographic additive differences. *Journal of Mathematical Psychology* 21: 191–218.

Fishburn, P.C. 1981. Subjective expected utility: A review of normative theories. *Theory and Decision* 13: 139–199.

Fishburn, P.C. 1982a. *The foundations of expected utility.* Dordrecht: Reidel.

Fishburn, P.C. 1982b. Nontransitive measurable utility. *Journal of Mathematical Psychology* 26: 31–67.

Fishburn, P.C. 1983. Transitive measurable utility. *Journal of Economic Theory* 31: 293–317.

Fishburn, P.C. 1984. SSB utility theory and decision-making under uncertainty. *Mathematical Social Sciences* 8: 253–285.
Fishburn, P.C. 1985. *Interval orders and interval graphs*. New York: Wiley.
Fishburn, P.C., and P.H. Farquhar. 1982. Finite-degree utility independence. *Mathematics of Operations Research* 7: 348–353.
Fishburn, P.C., and A. Rubinstein. 1982. Time preference. *International Economic Review* 23: 677–694.
Houthakker, H.S. 1961. The present state of consumption theory. *Econometrica* 29: 704–740.
Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.
Keeney, R.L., and H. Raiffa. 1976. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.
Kirkwood, C.W., and R.K. Sarin. 1980. Preference conditions for multiattribute value functions. *Operations Research* 28: 225–232.
Koopmans, T.C. 1960. Stationary ordinal utility and impatience. *Econometrica* 28: 287–309.
Koopmans, T.C., P.A. Diamond, and R.E. Williamson. 1964. Stationary utility and time perspective. *Econometrica* 32: 82–100.
Krantz, D.H., R.D. Luce, P. Suppes, and A. Tversky. 1971. *Foundations of measurement. Volume I: Additive and polynomial representations*. New York: Academic.
Loomes, G., and R. Sugden. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal* 92: 805–824.
Luce, R.D. 1978. Lexicographic tradeoff structures. *Theory and Decision* 9: 187–193.
Machina, M.J. 1982. 'Expected utility' analysis without the independence axiom. *Econometrica* 50: 277–323.
Mas-Colell, A. 1974. An equilibrium existence theorem without complete or transitive preferences. *Journal of Mathematical Economics* 1: 237–246.
Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.
Savage, L.J. 1954. *The foundations of statistics*, 2nd rev. ed. New York: Wiley. Dover Publications, 1972.
Schmeidler, D. 1984. Subjective probability and expected utility without additivity. Preprint no. 84, Institute for Mathematics and its Applications, University of Minnesota.
Sondermann, D. 1980. Utility representations for partial orders. *Journal of Economic Theory* 23: 183–188.
Stigler, G.J. 1950. The development of utility theory I, II. *Journal of Political Economy* 58: 307–327; 373–396.
Tversky, A. 1969. Intransitivity of preferences. *Psychological Review* 76: 31–48.
von Neumann, J., and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press; 2nd ed, 1947; 3rd ed, 1953.

Reputation

Martin W. Cripps

Abstract

We explain what reputation effects are, how they arise and the factors that limit or strengthen them.

Keywords

Collusion; Complete information games; Conflicting interest games; Extensive form games; Imperfect monitoring; Incomplete information games; Industrial organization; Jensen's inequality; Markov equilibria; Martingales; Prisoner's dilemma; Repeated games; Reputation; Signalling; Tit-for-tat; Uncertainty

JEL Classifications

C7

In a dynamic setting signals sent now may affect the current and future behaviour of other players; thus, signals can have effects unrelated to their current costs and benefits. It is the interplay between signals and their long-run consequences that is studied in the literature on reputation.

The literature on reputation has two main themes. The first is that introducing a small amount of incomplete information in a dynamic game can dramatically change the set of equilibrium payoffs: introducing something to signal can have big implications in a dynamic model. These kinds of result can also be interpreted as providing a robustness check. Dynamic and repeated games typically have many equilibria, and reputation results allow us to determine which equilibria continue to be played when a game is 'close' to complete information. The second theme of the literature on reputations is that introducing incomplete information in a dynamic game may introduce new and important signalling dynamics in the players' strategies. Thus reputation effects tell us something about behaviour. This theme is particularly important in applications to macroeconomics and to industrial organization, for example. For either of these themes to be relevant it is necessary to have a dynamic game with incomplete information, so work on reputation has been influenced by, and influences, the larger literature on repeated and dynamic games of incomplete information. An excellent detailed treatment of reputation can be found in Mailath and Samuelson (2006).

An Example

Most of the results below will be described in the context of a simple infinitely repeated trading game. The row player is a seller who can produce high or low quality. The column player is a buyer. Producing high quality is always expensive for the seller, so she would rather produce low quality; the buyer, however, wants to buy only a high-quality product. The only non-standard element is that the buyer regrets not buying a high-quality product. The trading game (Fig. 1) has a unique equilibrium (L, N).

Let us record some facts about this game. The set

$$V \equiv \{(x,y) : x > 0,\ \ y > -1/3,\ \ y \leq x \text{ and } y \leq 3 - 2x\} \times \subset \mathbb{R}^2,$$

illustrated in Fig. 2, is the set of feasible and strictly individually rational payoffs for the trading game. The axes are drawn through the minmax payoffs to make V clear. If the seller could commit to a pure strategy, she would prefer to choose H as the buyer's best response to this is B. However, she could do even better by committing to a mixed strategy; playing (3/4,1/4) for

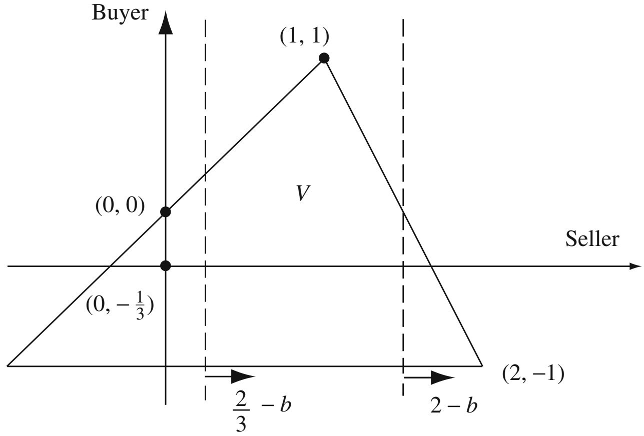
		Buy (B)	Not Buy (N)
High Quality	(H)	(1, 1)	(−1, −1)
Low Quality	(L)	(2, −1)	(0, 0)

Reputation, Fig. 1 A trading game

example would also ensure the buyer played B and give the seller a bigger payoff. Reputation arguments can provide ways for these commitment payoffs to be achieved by sellers who are not actually committed to anything.

The trading game is played in each of the periods $t = 1,2,...$ with perfect monitoring; at the end of the period the players get to observe all payoffs and the pure action taken by their opponent. If both players' discount factors, $\delta < 1$, were sufficiently large, any point in V could be sustained as an equilibrium payoff. If the seller is long lived but faces an infinite sequence of buyers who each live one period, then any point on the line segment joining (0,0) to (1,1) is an equilibrium payoff. (No seller payoff above 1 is achievable if mixed actions are not observable; see Fudenberg et al. 1990.)

The stage is now set. To understand how reputation works we will need to introduce something for the seller to signal. Its commitment to high quality? Its low cost of high quality? Its commitment to always ripping off customers. . . ? At this stage it is unnecessary to be specific, and we will concentrate on the general issues of learning. There are two types of sellers, 'strong' and 'normal', that the buyer may face in a game. The seller is told their type by nature at time $t = 0$. The buyer, however, is unaware of nature's selection and spends the rest of the game looking at the seller's behaviour and trying to figure out what type she is. The normal seller plays action $a \in \{H,\ L\}$ with probability $\tilde{\sigma}^t(a)$ at time t, and the strong seller plays a with probability $\hat{\sigma}^t(a)$ at time t. Everything we say in the section below applies to the case where normal and strong sellers follow history-dependent strategies. (These behaviour strategies do depend on the – public – history of play before time t, but let us keep this out of our notation.) An initially uninformed buyer attaches probability p^t to the strong type and $1 - p^t$ to the normal type at time t; again this depends on the observed history. Our buyer expects the seller to play a with probability $\bar{\sigma}^t(a) = p^t\tilde{\sigma}^t(a) + (1 - p^t)\hat{\sigma}^t(a)$, and as time passes the buyers observe the outcomes of this strategy and revise their prior accordingly.



Reputation, Fig. 2 Sets of equilibrium payoffs and reputation bounds

Tricks with Bayes's Rule and Martingales

Now we generate three properties of learning that are extensively used in the reputations literature. We will call them the 'merging' property, the 'right ballpark' property and the 'finite surprises' property. These properties are based on some simple facts about how Bayesian agents revise their beliefs, that is, how uncertainty about the seller's type is processed by the buyers or any other observer of its behaviour. A more advanced treatment of these results can be found in Sorin (1999). We defer any derivation of reputation results to the next section, so a reader could skip this section.

How does the buyer revise his or her beliefs in the light of an observed action a^t? A plain application of Bayes' rule tells us

$$p^{t+1} = \frac{Pr(a^t \cap \text{Strong})}{Pr(a^t)} = \frac{p^t \widetilde{\sigma}^t(a^t)}{\overline{\sigma}^t(a^t)}$$

Or, in terms of the change in the beliefs

$$\begin{aligned} p^{t+1} - p^t &= \frac{p^t\left[\widehat{\sigma}^t(a^t) - \overline{\sigma}^t(a^t)\right]}{\overline{\sigma}^t(a^t)} \\ &= \frac{p^t(1-p^t)\left[\widehat{\sigma}^t(a^t) - \widetilde{\sigma}^t(a^t)\right]}{\overline{\sigma}^t(a^t)} \end{aligned}$$

These equalities are powerful tools when combined with the properties of the priors.

Merging property. This tells us exactly how the long-run behaviour of the sellers is related to the buyer's long run beliefs. Either $p^t(1-p^t) \to 0$ and the buyer eventually learns the type of the seller and can perfectly predict their actions, or all types of the seller end up behaving in the same way $\widehat{\sigma}^t(a^t) - \widetilde{\sigma}^t(a^t) \to 0$ and again the buyer can perfectly predict their actions. Nothing else can happen!

The stochastic process $\{p^t\}$ is a martingale on [0,1] with respect to public histories. To see this there is a simple calculation we can do.

$$\begin{aligned} E(p^{t+1} | h_t) &= \sum_{a^t} Pr(a^t) p^{t+1} = \sum_{a^t} \overline{\sigma}^t(a^t) \frac{p^t \widetilde{\sigma}^t(a^t)}{\overline{\sigma}^t(a^t)} \\ &= p^t. \end{aligned}$$

(The expectation $E(\cdot)$ is taken with respect to the buyer's beliefs about future play.) Bounded martingales converge almost surely (see Williams 1991, for example), which implies $|p^{t+1} - p^t| \to 0$ almost surely. Applying this to the second equality above (noting that $|\overline{\sigma}^t(a^t)| \leq 1$), we get

$$p^t(1-p^t)\left|\widehat{\sigma}^t(a^t) - \widetilde{\sigma}^t(a^t)\right| \to 0, \quad \text{(Merging)}$$

almost surely. This kind of result is extensively used in Hart (1985) and the literature that stems from his work.

Right ballpark property. The strong seller knows that the future will evolve according to the strategy $\widehat{\sigma}$ (we use $\widehat{Pr}(\cdot)$ and $\widehat{E}(\cdot)$ to denote her probability measure and its expectation). This seller might ask, as she plays out an equilibrium, how little probability the buyers can attach to the strong seller, or how low p^t could get when she plays $\widehat{\sigma}$. Of course, when the seller is in fact the strong type it is very unlikely that $p^{t.}$ becomes low – beliefs must stay in the right ballpark. (For example, if $\widehat{\sigma}$ was actually a pure strategy the strong seller cannot ever believe p^t will decrease. As she plays $\widehat{\sigma}$ there will be periods in which the normal type of seller could have done something different, so observing the actions of $\widehat{\sigma}$ will cause buyers to revise p^t upwards.)

From the perspective of the strong seller, the likelihood ratio is a martingale:

$$\widehat{E}\left(\frac{1-p^{t+1}}{p^{t+1}}\,|\,h_t\right)=\frac{1-p^t}{p^t}.$$

(The calculation is just like the earlier one for p^t, where we use $\widehat{Pr}(a^t)=\widehat{\sigma}^t(a^t)$.) Let τ be the first time, s say, that $p_s\le v$ and let C^t be the event that $\tau\le t$. That is, sometime in the first t periods $p_s<\nu$. Then the martingale property combined with the optional stopping theorem (for example, Williams 1991) implies

$$\frac{1-p^0}{p^0}=\widehat{E}\left(\frac{1-p^{t+1}}{p^{t+1}}\right)\ge\widehat{Pr}(C^t)E\left(\frac{1-p^\tau}{p^\tau}\,|\,C^t\right)\ge\widehat{Pr}(C^t)\frac{1-v}{v}.$$

The above gives an upper bound on $\widehat{Pr}(C^t)$ that is independent of t. Thus it also bounds the probability that p^t is ever below ν:

$$\widehat{Pr}(\exists t\ \text{s.t.}\ pt<v)\le\frac{v}{p^0}.\quad\text{(Right Ballpark)}$$

Hence, the strong seller knows that it is very unlikely that the buyer's posterior will ever be close to certain she is actually the normal seller.

Finite surprises property. The strong seller might also ask how many times (as she plays $\widehat{\sigma}$) the uninformed buyers will make a big mistake in predicting her strategy, that is, how many periods does $\left\|\widehat{\sigma}^t-\overline{\sigma}^t\right\|>v$ occur when the seller actually plays $\widehat{\sigma}$. Here we are helped by the fact that our seller has only two actions, so the variation distance between the mixed actions is just twice the difference in probability of the realized action $\left\|\widehat{\sigma}^t-\overline{\sigma}^t\right\|=2\left|\widehat{\sigma}^t(a^t)-\overline{\sigma}^t(a^t)\right|$. Let M_N be the event that there are more than N *mistakes*, $\left\|\widehat{\sigma}^t-\overline{\sigma}^t\right\|>v$, before time T. The finite surprises property is that independently of the equilibrium $\widehat{Pr}(M_N)\to 0$ as $T,\ N\to\infty$. Thus, it is very unlikely that there are many periods in which the buyers do not think the seller will play as the strong type if the seller is indeed this type.

Jensen's inequality applied to the likelihood ratio above implies that the prior is a submartingale, that is, $\widehat{E}(p^{t+1}|\,h_t)\ge p^t$. There is a second property of martingales we can now use: they cannot move around very much: $\sum_{t=1}^T\widehat{E}\left(\left(p^{t+1}-p^t\right)^2\right)\le 1$. (A proof of this fact follows from $\widehat{E}(p^{t+1}-p^t)^2\le\widehat{E}\left((p^{t+1})^2-(p^t)^2\right)$. A substitution from the first Bayes' rule equality above then tells us

$$1\ge\sum_{t=1}^T\widehat{E}\left(\left(p^t\left[\widehat{\sigma}^t(a^t)-\overline{\sigma}^t(a^t)\right]\right)^2\right).$$

It is obvious that only a few of the (non-negative) terms in the sum above can be much above zero, otherwise the upper bound will be violated. The right ballpark property tells us it is very unlikely that $p^t<v$. On the event $\{p^t\ge v\ \forall\ t\}\cap M_N$, the $p^{t.}$ in the above expectation is greater than v and there are at least N differences that are bigger than $v/2$, so the sum is at least $Nv(v/2)^2$, hence

$$1\ge\sum_{t=1}^T\widehat{E}\left(p^t\left[\widehat{\sigma}^t(a^t)-\overline{\sigma}^t(a^t)\right]\right)^2\ge\widehat{Pr}(\{p^t\ge v\forall t\}\cap M_N)\frac{Nv^3}{4}.$$

Using the fact that $Pr(A\cap B)\ge Pr\,(B)-\ Pr\,(A^c)$ we now have an upper bound on $\widehat{Pr}(M_N)$.

$$\frac{4}{Nv^3}+\widehat{Pr}(\exists t\ s.t.\ pt<v)\ge\widehat{Pr}(M_N).$$

The right ballpark property gives us

$$\widehat{Pr}(M_N) \leq \frac{v}{p^0} + \frac{4}{Nv^3}.\text{(Finite Surprises)}$$

As the size of the surprises becomes small $v \to 0$ and the number of surprises becomes large $Nv^3 \to \infty$, the strong seller must attach smaller and smaller probability to M_N. Fudenberg and Levine (1989, 1992), for example, invoke this property.

Basic Reputation Results: Behaviour

The three tools above are sufficient to establish most well-known reputation results. The arguments below are entirely general, and are widely applied, but we use them only in the trading game. To make things simple, suppose that for some reason the strong seller is committed to playing $(b, 1 - b)$, that is, in every period t the strong buyer provides high quality with probability b. We reserve the discussion of more complicated types of reputations for a later section.

From the perspective of the buyer, any equilibrium will consist of two phases: an initial phase when there is learning and signalling about the seller's type (this is sometimes called reputation building, although often reputation destruction is what occurs), and a terminal phase when the learning has virtually settled down. It is the merging property that tells us there must be this latter phase. The play in the game moves into this second phase either because the buyer is almost sure he knows the type of the seller (reputation considerations have vanished) or because the sellers are playing in the same way. Thus the equilibria of dynamic signalling games are inherently non-stationary, which is in contrast to much of the work on repeated games. Of course, Markovian equilibria can be calculated but these too will exhibit the two phases of play. The initial learning, when reputation builds or is destroyed, depends on the particular equilibrium and the game being studied. This phase may last only one period (if a once and for all revealing action is taken by a seller) but frequently it is long and has a random duration (if both types of seller randomize, for example).

Let us first examine reputation destruction in the case where $b \approx 1$, so the strong seller is committed to high quality and only very occasionally slips up. There is an equilibrium of this game where the normal type of seller will offer low quality more often than the strong type, and thereby gradually reveal her type (destroy her reputation for being good). Nevertheless, as this occurs she will enjoy heightened payoffs. The trade-offs our normal seller experiences in this game are what drive the reputation destruction. A seller offering low quality today enjoys the benefit of a higher payoff now, but the observation of low quality typically leads the buyers to revise downwards their probability of the strong seller and buy less in the future, whereas a seller offering high quality will lead the buyer's posterior on the strong seller to be revised upwards and an increased likelihood of buying in the future.

Exactly how the normal seller chooses to trade off long-run benefits and short-run costs is unclear. It is possible that pooling dominates and that future buying is so strong that the normal seller prefers to offer high quality today even if it costs something in the short run. However, in this equilibrium the normal seller perceives the long-run benefits to be relatively small and prefers to offer low quality today. The normal seller can be thought of as exploiting, or cashing in, the value of her accumulated reputation. We also know, from the finite surprises property, that there will be finite opportunities for the normal seller to do this. Relatively soon there will come a time where the buyers know the seller is normal and purchase accordingly.

Reputation building (as opposed to destruction) is more likely in a world where there is the possibility that one is thought to be bad, for example, if the strong type is committed to ripping customers off and only occasionally produces a good product ($b \approx 0$). In such a world the normal seller wants to tell buyers she is not this type, because by playing as the strong type she is doomed to never trade. She is building a reputation for *not* being the strong type. To do this the normal type will have to incur the cost of repeatedly offering high quality, even if the buyer is not buying. This is expensive and will drag down the normal seller's equilibrium payoff. But, as above, it will increase the likelihood of future buying by decreasing the likelihood of a strong seller. In contrast to the reputation destruction case, there are short-run costs borne by the normal type to achieve long-run gains. Again, the nature of these costs and benefits rely on the buyers' uncertainty about the seller's type.

Basic Reputation Results: Payoffs

Reputation issues can have an extreme effect on payoffs, and this is what first came to the attention of economists. The general question of how the presence of something to signal in the repeated game affects the equilibrium payoffs could be answered in a number of ways. One way would be to calculate equilibria explicitly. This is usually difficult and would not establish results that hold for *all* equilibria.

Instead, a different approach is taken that is described in the following recipe:

1. If the seller is strong, then in finite time the buyers will believe they face a seller who plays arbitrarily close to $(b, 1 - b)$ for ever.
2. Figure out what the buyers will do when the seller is strong.
3. Use step 2 to evaluate the normal seller's payoff if she pretends to be strong for ever.
4. At a Nash equilibrium the answer to step 3 is a lower bound on the normal seller's equilibrium payoff.

Step 1 is independent of the model and is a result of our earlier calculations. The right ballpark property tells us that p^t does not tend to zero when the seller is strong. The merging property then implies either $p^t \to 1$, or eventually all remaining normal types of buyer are also playing arbitrarily close to $(b, 1 - b)$. In either of these cases, at a large but finite time the buyers believe that they face a seller who will always play $(b, 1 - b)$.

Before proceeding to apply this recipe, we illustrate its power with the remarkable results we expect to get. Let us first consider a world where buyers are short run. We will show that introducing an arbitrarily small probability that there is a strong seller places a lower bound on the normal seller's equilibrium payoffs of $2 - b$ (when $b > 1/3$). Thus for b close to 1/3 the equilibrium payoffs in the complete information game (the segment joining (0,0) and (1,1)) and the incomplete information game are disjoint! Moreover, the normal seller can get almost his maximum feasible payoff at every equilibrium. In the second case, where buyers are also long run, we will get less strong conclusions; nevertheless, we will show that the normal type of seller must get at least $2/3 - b$ when $b > 1/3$. These payoffs are illustrated in Fig. 2.

The really difficult part of our recipe is step 2, because we have to understand how the buyers will behave in equilibrium. We therefore need to consider as separate cases what happens if buyers are short run or long run. Also, the amount of discounting that the sellers do affects the answer to step 3, so we need to consider different arguments for different amounts of discounting. The following catalog moves from simple to more elaborate arguments and from stronger to weaker reputation effects.

Reputation Without Discounting: Short-Term Buyers

When a buyer lives only one period he plays a best response to the seller's current action. By step 1 in the very long run this will be B if $b > 1/3$ and N if $b > 1/3$. Step 3 is simple; by playing $\widehat{\sigma}$ for ever the normal seller knows that in a large but finite time she can ensure the buyer will behave as above and so she will receive a stage game payoff approximately $R^*(b)$, where

$$R^*(b) := \begin{cases} 2 - b & b > 1/3, \\ -b & b < 1/3. \end{cases}$$

If there is no discounting, and limits are correctly taken, $R^*(b)$ will equal the normal type's payoff from playing $\widehat{\sigma}$ for ever. Thus, at step 4, at any Nash equilibrium the normal type must get at least $R^*(b)$.

In a general game R^* is equal to the seller's payoff from playing the strong type's stage game strategy when the buyer plays his or her unique best response. (If the best response is not unique this is not correct.)

Reputation with Discounting: Short-Term Buyers

Step 2 is as above – we still have short-term buyers. When the normal seller discounts payoffs, however, playing $\widehat{\sigma}$ and eventually getting $R^*(b)$ every period does not tell us what her payoff discounted to time zero will be. There is an order of limits issue; as the discounting of the seller becomes weaker ($\delta \to 1$), it could be that the equilibria change and there are more and more periods where the seller is not getting $R^*(b)$. It is now that the finite surprises property plays an important role. First notice that, when ν is chosen appropriately and $\|\overline{\sigma} - \widehat{\sigma}\| < v$, then playing a best response to $\overline{\sigma}$ is the same as playing a best response to $\widehat{\sigma}$. Hence, it is only when a surprise occurs that the normal seller is not getting $R^*(b)$ from playing $\widehat{\sigma}$. But the probability of more than N surprises can be made very small *independently of the discounting*. So, as the discounting becomes weak and N periods have a small effect on total discounted payoff, there is a small probability of the normal seller of getting anything less than $R^*(b)$ when she plays $\widehat{\sigma}$. Any Nash equilibrium, therefore, gives the normal seller at least $R^*(b)$. This is the kind of argument first made in specific cases by Kreps and Wilson (1982) and Milgrom and Roberts (1982), and generalized in Fudenberg and Levine (1989, 1992).

Reputation Without Discounting: One Long-Run Buyer

If the buyer lives for many periods, he will not necessarily play a short-run best response to $(b, 1 - b)$ even if he expects it to be played for ever. We can, however, use some weaker information. At an equilibrium the buyer must on average get at least $-1/3$ (his minmax payoff) against $(b,1 - b)$. This implies that the buyer has to buy with at least probability 1/3 when $b > 1{=}3$ and buy with at most probability 1/3 when $b < 1{=}3$. There are, consequently, some bounds on the normal seller's payoff when she has played $\widehat{\sigma}$ for a sufficiently long time. While playing $(b,1 - b)$ she gets $2 - b$ when the buyer buys and $-b$ if not; thus, if the buyer buys with probability greater than 1/3, she expects to receive a payoff of at least $2/3 - b$. If the buyer buys with at most probability 1/3, she expects to get at least $-b$. The seller is not discounting, so what she gets in the long run from playing $\widehat{\sigma}$ is also what she expects to get at time zero. Our answer to step 3, therefore, is

$$R^{\dagger}(b) := \begin{cases} (2/3) - b & b > 1/3, \\ -b & b < 1/3; \end{cases}$$

and we have a weaker lower bound on the normal type's payoff.

In an arbitrary game $R^{\dagger}$ is equal to the seller's worst payoff from playing as the strong type when the buyer plays a response that gives him more than his minmax payoff. In certain cases this can be a very strong restriction – for example, if the seller has a pure strategy that minmaxed the buyer and there is a unique response for the buyer that ensured he received his minmax payoff. Certain games, known as games of conflicting interests, have the property that the best action for the seller to commit to is pure and minmaxes the buyer. $R^{\dagger}$ is a very tight bound for such games.

Reputation with Discounting: One Long-Run Buyer

This final case combines most of the above issues. If the seller discounts the future much less than the buyer, then in the long run the seller must get $R^{\dagger}(b)$ from playing $\widehat{\sigma}$. If a normal seller pretends to be strong, the buyers think there are at most N periods when the strong strategy is not played. Imagine now we have a buyer who cares only about what happens in the next t' periods. Such a buyer can think there are at most $t'N$ periods in which $\widehat{\sigma}$ is not played for the next t' periods. (This kind of argument is due to Schmidt 1993.) As the seller becomes very patient Nt' periods become of vanishing importance and the normal seller's payoff is bounded below by $R^{\dagger}(b)$. If the seller and buyer discount equally, however, reputation effects cannot be found except in some very special cases.

Imperfect Monitoring: Temporary and Bad Reputation

The analysis of reputation given above presupposes perfect monitoring by the buyers and sellers of each others' actions. In many dynamic and repeated games this is not likely. To what extent do the above results continue to hold when the players are not able to see exactly what their opponent did in any one period? Perhaps reputations are harder to establish if the observed behaviour is noisy? On the other hand, perhaps deviations from the strong type's action are harder to detect and so reputations last longer and are more valuable.

The merging, right ballpark and finite surprises properties all hold true under imperfect monitoring, with a suitable redefinition, provided there is enough statistical information for the buyer to eventually identify the seller's behaviour. (This is a full-rank condition on the players' signals.) As a result, the bounds on payoffs given in the previous section continue to hold.

Under imperfect monitoring with adequate statistical information there is one new behavioural feature of these games – reputation is almost always temporary, that is, the buyer will eventually get to know the seller's type. To see why this is so, let us amend the game in Fig. 1 by restricting the buyer to imperfectly observe the seller's action. With probability $1 - \varepsilon$ the buyer observes the seller's true action in the current period, but with probability ε he observes the reverse action. (We must also assume the buyer does not see his own payoffs, otherwise he can deduce the seller's action from his payoff.) Consider a game where the seller always provides high quality ($b = 0$) and suppose that reputation is permanent in such a game. Then p would, at least some of the time, converge to a number that is not zero or one. (Remember beliefs have to converge.) The merging property tells us that, in this case where the limit of beliefs is between zero and one, the buyer will be certain the normal seller is always providing high quality. Such buyers will ignore the occasional low-quality product as just unlucky outcomes, and there will be no loss of seller reputation if the buyer ever receives low quality. The normal type of seller can, therefore, deviate from always providing high quality, gain one unit of profit, and not face any costs in terms of loss of reputation. This cannot be an equilibrium. The initial claim that reputation is permanent has to be false as a result of this contradiction. The details of this argument can be found in Cripps et al. (2004).

When the monitoring is not statistically informative, 'bad reputation', due to Ely and Valimaki (2003), is a possibility. Uninformative monitoring is a particular problem in repeated extensive form games, because players do not get to see the actions their opponent would have taken on other branches of the game tree. Bad reputation may arise in our example if the buyer could take an action (such as not to buy) that stopped the seller being able to signal her type. Then, the normal seller might find herself permanently stuck in a situation where she cannot sell. This is not particularly surprising if the buyers were strongly convinced they faced a strong seller that almost always provided low quality. However, in certain circumstances this problem is much more severe: even if the buyers were almost certain the seller were normal, every equilibrium has trade ending in a bounded and finite time. Thus, it is possible that introducing something for the seller to signal has huge negative costs for her equilibrium payoffs. To illustrate this, suppose the seller were a restaurant with imperfect control over quality, although it does have a strategy (for example, doubling the butter and salt content!) that makes it more likely the buyer will think the meal he received is good – but is actually damaging to the buyer.

When play has reached the position where just one more bad meal will lead the buyer to permanently avoid the restaurant, then the restaurant will choose to use this unhealthy strategy. Knowing this, the buyer will choose to go elsewhere for his last but one meal too, and there is an unravelling of the putative equilibrium. Buyers eat at the restaurant only if they get very few bad meals, because they know they are in for clogged arteries and high blood pressure after that. Bad reputation arises because the seller cannot resist the temptation of taking actions that are actually unfavourable to the buyer in an effort to regain his good opinion. They actually have the reverse effect of ultimately driving the buyers away.

Reputation for What?

In our discussion we consider a strong type of seller who is committed to playing a particular fixed (random) action in each period. Is this form of uncertainty the only relevant one, or are there other potential types of strong seller that may do even better for our normal seller? There are two alternatives to consider: the strong seller is committed to playing a history-dependent strategy, or the strong player is equipped with a payoff function and her strategy is determined by an equilibrium.

If the seller faces a sequence of short-term buyers, then committing to a fixed stage game action is the best she could ever do, because each buyer's optimization focuses on what the seller does in the current period – the future is irrelevant. Even when the buyers are long lived, there are circumstances where committing to play a fixed action imparts a strategic advantage in repeated play, for example in most coordination or common interest games. However, there are other repeated games, such as the Prisoner's Dilemma, and dynamic games where committing to a fixed stage action is worthless. What the seller would like to do is to commit to a strategy, such as tit-for-tat, which would persuade a sufficiently patient buyer to cooperate with the strong type. Provided some rather strong conditions are satisfied, this is possible.

Our recipe for reputation results will break down when we consider strong sellers with payoffs rather than actions; nevertheless, reputation results are possible. For example, if the strong seller had payoffs of 2 for high quality and zero for low quality he would be strategically identical to a seller who always provided high quality.

Many Players: Social Reputation and Other Considerations

Thus far we have resolutely stuck to a model of two players, but it is clear that reputation is a pervasive social and competitive phenomenon. Here we sketch some of the issues in many-player reputation. The literature on this area is in its infancy; very little can be said with much certainty now.

The easiest case to deal with is what happens as the number of uninformed players (the buyers in our example) increases. Here the benefit to the seller of building a reputation for high quality increases, as providing a good product today means the seller is more likely to trade with many buyers tomorrow. In a way, increasing the number of buyers is like making the seller more patient, and so we would expect the seller to be more inclined to build a reputation in this case.

A second case would be where there are very large numbers of informed buyers trying to acquire reputations for individual or group characteristics. Models of career concerns are similar to reputation models and have many workers trying to acquire reputations for individual characteristics. Also, there are models of group reputation, such as Tirole (1996), where a particular class of individuals behaves in a particular way to perpetuate the 'group's' reputation. In both these types of model the large numbers assumption allows one individual's reputation decision to be treated as virtually independent of others. Thus they can be analysed using quite simple tools.

A final case is where a few informed agents are in competition or collusion with each other. Collusion in team reputation obviously introduces a public goods issue. If one player contributes to the good name of the group, he or she does not get to enjoy the full benefits of the contribution.

Typically, therefore, reputations for such teams are harder to establish. One might conjecture that competition appears to drive a player towards excessive investment in reputation, but there are many effects at work that we do not completely understand. For example, competitors may also act to undermine their rival's reputation and to interfere with its development. This is a fertile region for applied and theoretical investigations.

See Also

► Repeated Games
► Signalling and Screening

Bibliography

Cripps, M.W., G.J. Mailath, and L. Samuelson. 2004. Imperfect monitoring and impermanent reputations. *Econometrica* 72: 407–432.

Ely, J., and J. Valimaki. 2003. Bad reputation. *Quarterly Journal of Economics* 118: 785–814.

Fudenberg, D., D. Kreps, and E. Maskin. 1990. Repeated games with long-run and short-run players. *Review of Economic Studies* 57: 555–574.

Fudenberg, D., and D.K. Levine. 1989. Reputation and equilibrium selection in games with a patient player. *Econometrica* 57: 759–778.

Fudenberg, D., and D.K. Levine. 1992. Maintaining a reputation when strategies are imperfectly observed. *Review of Economic Studies* 59: 561–579.

Hart, S. 1985. Nonzero-sum two-person repeated games with incomplete information. *Mathematics of Operations Research* 10: 117–153.

Kreps, D., and R. Wilson. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27: 253–279.

Mailath, G.J., and L. Samuelson. 2006. *Repeated games and reputations: Long-run relationships*. Oxford: Oxford University Press.

Milgrom, P., and J. Roberts. 1982. Predation, reputation and entry deterrence. *Journal of Economic Theory* 27: 280–312.

Schmidt, K. 1993. Reputation and equilibrium characterization in repeated games of conflicting interests. *Econometrica* 61: 325–351.

Sorin, S. 1999. Merging, reputation, and repeated games with incomplete information. *Games and Economic Behavior* 29: 274–308.

Tirole, J. 1996. A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies* 63: 1–22.

Williams, D. 1991. *Probability with martingales*. Cambridge: Cambridge University Press.

Resale Markets

Philippe Jehiel

Abstract

Resale markets are necessary to correct misallocations of assets, but do they always ensure that goods end up in the hands of those who value them most? This article reviews theoretical arguments as to why this need not necessarily be so and when inefficiencies might be expected despite the presence of resale markets. Policy implications are also suggested.

Keywords

Akerlof, G; Allocative externalities; Asymmetric information; Auctions; Bargaining; Chicago School; Coase Theorem; Commitment; Efficient allocation; Incentive constraints; Interdependent values; Limited liability; Market failure; Market imperfections; Mechanism design; Participation constraints; Private information; Property rights allocation; Resale markets; Reservation utility; Revelation principle; Risk aversion; Transferability

JEL Classifications

L1; D8

Resale markets seem necessary to correct misallocations of assets, where misallocations may be the result of mistakes in initial purchasing decisions, or more generally of changes in the state of the economy. For the sake of illustration, a car owner may after a while find it desirable to buy a new car, and he may be willing to resell his old car on the second-hand market. A manager of a firm holding a Universal Mobile Telephone System (UMTS) licence may be willing to resell her licence to another firm if she realizes that the firm is unable to cover its cost (generated by the licence acquisition). A homeowner may need to resell his house if he has to move to another country or jurisdiction.

A question of primary interest is whether such resale markets are good for the economy. Or, to put it differently, whether, when and how should such resale markets be regulated? This article starts with the laissez-faire viewpoint on this issue; it then proceeds to show how asymmetric information and commitment issues mitigate that viewpoint.

The Laissez-Faire Viewpoint

The classical neoliberal viewpoint as represented by the Chicago School would favour laissez-faire. Within the present context, this would imply that resale markets should not be regulated. The premise of this line of thought is that resale markets give the right flexibility so that assets can be allocated to the right agents at any point in time. This view has important consequences for the theory of mechanism and market design. Indeed, it implies that the initial allocation of property rights is irrelevant, as resale markets should be able to correct any misallocations (this is one version of the so-called Coase Theorem – Coase 1960). Thus, according to this view, a government interested in maximizing economic efficiency should worry neither about the method of privatization nor about how to allocate licences for operating public services. It should simply allow for well-functioning resale markets.

Of course, very few economists truly believe that real resale markets can achieve such a fantastic job of always allocating assets to the right agents at the right time. On the academic side, Akerlof (1970) provides an early theoretical example of market failure in the context of the market for used cars (more on this below). Coase himself argues that transaction costs which are numerous are likely to invalidate the above angelic view about resale markets. On the 'real world' side, it seems implausible that the method of privatization or the allocation of licences for the use of public services is irrelevant for economic efficiency. In fact, recent years have seen a rapid growth of auction methods to allocate licences or privatize publicly owned firms, suggesting an interest on the part of practitioners in market design. It is worth pointing out that, in the case of licence auctions, most governments have chosen not to allow for resale markets, suggesting some distrust towards their functioning.

In the tradition of Coase, the words 'transaction costs' will be interpreted to mean any reason why inefficiencies may arise in transactions. Of course, some of the reasons need not be related to the intuitive notion of transaction costs, and one could alternatively use the more neutral terminology of 'market imperfections'. The rest of this article will review how theoretical insights from the mechanism design literature and the bargaining literature help identify significant sources of transaction costs. The review will abstract from transferability issues, which is a legitimate idealization for transactions that are not too big for the financial capabilities of the parties. The theoretical insights will then be used to shed some light on whether and how to regulate resale markets.

The Role of Private Information

It is relatively intuitive to see why private information may be a source of inefficiency in transactions. A seller who privately knows her valuation for the object for sale has an incentive to pretend that she values the object more than she really does, in the hope that this will lead the buyer to increase his purchasing price. Similarly, a buyer has an incentive to pretend that he values the good less than he really does, in the hope that he will obtain a lower selling price. But such distortions inevitably induce inefficiencies whenever the gains of trade are not large enough. This intuition has been formalized in the work of Myerson and Satterthwaite (1983), who show that, if the distributions of valuations are independently distributed between a seller and a buyer, and if it is not known who values the good more, inefficiencies must arise in any bargaining game in which no outside money is given to the bargaining parties. One of the strengths of Myerson and Satterthwaite's work is that it applies to any bargaining game, including protocols in which a broker could help improve the bargaining outcome and protocols allowing for several stages of bargaining. The result is obtained by relying on the so-called revelation principle, which allows for the derivation of constraints that should be satisfied in any Nash–Bayes equilibrium of any game (whether static or dynamic): these constraints are the so-called incentive constraints – an agent with valuation v should find his own strategy no worse than the strategy of the same agent with valuation v' – and the participation constraints – an agent should get at least what he could get by staying outside the game. Myerson and Satterthwaite then proceed to show that these constraints together with the constraint that the bargaining parties receive no outside money cannot be simultaneously satisfied unless there are inefficiencies (see Milgrom 2004, for an exposition of this and other impossibility results).

The above buyer–seller set-up assumes that agents know how valuable the good is to them. This is referred to as a 'private values set-up'. Akerlof (1970) identifies another source of bargaining inefficiency in set-ups in which the value to the buyer is a function of the information held by the seller – this is sometimes called an informational externality and referred to as an 'interdependent values set-up'. For example, a seller of a used car may know the quality of his or her car, and the quality of the car obviously affects the valuation of both the seller and the buyer. In an elegant example, in which the buyer is known to value the good α times as much as the seller with $2 > \alpha > 1$ and the quality (identified here with the valuation of the seller) is distributed uniformly on [0,1], Akerlof shows that there can be no trade. The no-trade result arises because a selling price of p would be acceptable to the seller only if the quality is below p, resulting in an average quality of $p/2$. But such an average quality does not justify buying the good at price p for the buyer, as $\alpha p/2 - p < 0$. One of the beauties in Akerlof's example is that it illustrates that, even in situations in which it is common knowledge that the buyer values the good more than the seller, there is no trade in equilibrium. Even though Akerlof restricts his analysis to special trading mechanisms, the inefficiency he identifies can be shown to arise in any equilibrium of any bargaining game, with the use of the same mechanism design techniques as those of Myerson and Satterthwaite. It also extends (even though not in the extreme form of no trade) to other classes of problems with interdependent values (see Samuelson 1984).

In the above bargaining set-ups, a specific form of property rights was assumed. Within the same examples, other efficiency conclusions would arise with alternative property right structures, thereby illustrating how the initial allocation of property rights may affect efficiency in the presence of informational asymmetries. Obviously, in Akerlof's interdependent values example, if the person valuing the good more is initially the owner of the good there is no inefficiency, which thereby offers a simple illustration of this idea. (See Jehiel and Pauzner 2006, for further elaboration.) In the private values situation considered by Myerson and Satterthwaite, if the two parties are *ex ante* symmetric and initially own 50 per cent shares of the object, a double auction (in which the party quoting the highest price would buy the 50 per cent shares of the other party at a selling price in between the two quoted prices) would result in an efficient allocation of property rights. Cramton et al. (1987) generalize the latter insight by showing that mixed ownership is economically superior in partnership dissolution problems with private values.

The above bargaining inefficiencies implicitly assume that no outside money can be introduced on to the bargaining table. Otherwise, with large enough subsidies, efficiency could be obtained in the above bargaining set-ups, thereby suggesting that an appropriate public intervention may eliminate the inefficiency due to asymmetric information. However, in interdependent values situations in which agents hold multidimensional signals that are independently distributed across agents, Jehiel and Moldovanu (2001) show that the sole incentive constraints make it generically impossible to achieve the first-best allocation no matter how much money is introduced on to the bargaining table. This result is especially relevant in transactions involving several items because then private information is naturally multidimensional. The result then implies that no public intervention can eliminate the bargaining inefficiencies. (A similar conclusion arises even with one-dimensional private information if the single crossing condition is violated; see Maskin 1992.)

The above results assume that there is no correlation in the private information held by the various agents. Whenever there are correlations, incentive constraints are less severe because the report made by agent *i* can be used to deter misreports by agent *j*. The works of Crémer and McLean (1985, 1988) and Johnson et al. (1990) (see also Myerson 1981) suggest that inefficiencies can be totally eliminated even under moderate correlations if agents are risk neutral and transfers can be arbitrarily large. However, limited liability and risk aversion (which seem plausible, especially if very large transfers are involved) ensure that the qualitative insights obtained for the case without correlation continue to hold with moderate correlation (see Robert 1991). Hence, inefficiencies due to asymmetric information continue to hold even in the correlated case, as long as correlation is not too large. (See also Compte and Jehiel 2006, who argue within Myerson and Satterthwaite's private values set-up that inefficiencies may arise even with large correlation whenever agents have the option to leave the bargaining table at any time, thereby obtaining their reservation utility.)

As already mentioned, the above inefficiencies hold even if multiple stages of bargaining are allowed, as long as the only inferences of the players come from the equilibrium play of the other parties and not from the release of new hard information (either in an exogenous manner or through endogenous information acquisition). If new information becomes available, the situation is different. Obviously, if the private information held by the various agents become public at some stage, then at this stage bargaining parties with full commitment abilities should be able to implement an efficient agreement. This is because, if inefficiencies were to arise at that stage, a party could propose a Pareto improvement with no further move, keeping the generated surplus for herself: this can be viewed as an application of the Coase Theorem. But, even if one adopts the view that eventually private information becomes publicly available, a critical issue is about how long this takes. If it takes very long, inefficiencies are still likely to be significant because the transitory phase is long. If it does not take long and full commitments are possible, efficiency can be expected.

The Role of Commitment

The above reported results assume full commitment abilities on the part of the bargaining parties. Another major source of inefficiencies is the limited commitment abilities of the agents. From the viewpoint of mechanism design, the relaxation of commitment abilities of the proposing party (sometimes called Principal) is generally thought of as a bad thing. But one should be cautious here about the criterion used to assess what 'good' or 'bad' means. Clearly, from the viewpoint of the Principal limited commitment ability is a bad thing because it puts additional constraints on the Principal's maximization exercise. However, from the viewpoint of society (as measured by social welfare), the conclusion is far from clear. For example, Coase's conjecture suggests that a monopolist with no commitment ability may end up pricing his good efficiently if consumers are forward-looking (they anticipate the distribution of future prices correctly) and patient enough. In a similar vein, the commitment ability of an auctioneer may allow him to use inefficient reserve prices, which he might be unable to exploit under weaker commitment scenarios. (See McAfee and Vincent 1997, for a formal approach, and Zheng 2002, for an optimal auction model in which, even though the seller can commit not to lower his reservation price if there is no interested buyer, buyers can resell the object if they wish.) Clearly, more work is required to understand the pros and the cons of commitment from a mechanism design perspective with non-benevolent principals.

In a number of transactions, the transacting parties impose a cost or benefit on third parties: think of the sale of pollution rights or the sale of technologies through patents in imperfectly competitive markets. From the viewpoint of the transaction, this corresponds to an externality in the sense that the trade between a subset of agents affects the payoffs of other agents (see Jehiel et al. 1996). Abstracting from informational asymmetries, Jehiel and Moldovanu (1999) in a one-object environment and Gomes and Jehiel (2005) in a general multi-object environment study resale markets in such set-ups with allocative externalities. They establish that the lack of commitment ability may induce long-run inefficiencies in resale markets whenever there are allocative externalities and agents are patient and forward-looking. Furthermore, if we take as given the legal constraints governing how goods can be exchanged, the initial allocation of property rights is shown to have no effect on the long-run properties of the equilibrium pattern of sales in such markets, as long as parties are forward-looking and patient enough. Thus, in such a complete information world, the lack of commitment ability induces inefficiencies in the presence of allocative externalities and at the same time makes it irrelevant how the initial property rights are allocated.

Practical Implications

What are the lessons to be drawn from these theoretical observations? What do these results imply for the desirability of resale markets?

A first category of problems concerns those situations in which private information is persistent. Then the above inefficiency results show that in most scenarios, no matter how exchanges are organized, no matter whether or not resales are permitted, and no matter how well resale markets work, inefficiencies are inevitable. In interdependent value situations with multi-dimensional signals, even subsidies may not be enough to eliminate the inefficiencies.

Full commitments including controls over resales would seem desirable from a mechanism design viewpoint, as long as the proposing parties seek to maximize total welfare. However, with non-benevolent agents there is no reason in general to expect the full commitment scenario to be preferable to weaker commitment scenarios whenever private information is persistent.

A second category of problems concerns those situations with vanishing private information that will be identified with complete information. Then resale markets permit an efficient allocation of goods whenever agents care solely about their own allocation (that is, when there are no externalities). However, when there are allocative externalities in the sense that the allocation of agent i directly influences the well-being of agent j, resale markets do not allow parties with limited commitment abilities to reach an efficient state of the economy. Yet, even when there are allocative externalities, the efficiency of the economy is unaffected by the initial allocation of property rights, suggesting that in such situations the only role for government interventions is through the legal framework, not the allocation of property rights. For example, it may be desirable from this perspective to require by law that the transacting parties compensate those agents suffering from the transaction.

In complete information situations, it would seem that full commitments including controls over resales should improve efficiency. However, that view ignores the reality of a changing environment, which is one of the basic rationales for the existence of resale markets. Because the economy is changing, resale markets are necessary. The complete contracting scenario implicitly assumed by the full commitment idea is impractical in that it might involve agents that are not even present in the economy (think of a future homeowner who may not yet be born and whose future possession already exists). From a practical viewpoint, the main issue is about understanding the effect of the legal framework that governs resale markets on the overall efficiency of the economy. Some insights about how the legal framework might improve the economic performance of resale markets have been suggested above (see the idea of compensating those agents who suffer from the transaction). Admittedly, more work on both the theoretical and empirical sides is required to understand this as well as the additional effect of persistent private information on resale markets.

See Also

- ▶ Bargaining
- ▶ Coase Theorem
- ▶ Efficient Allocation
- ▶ Incentive Compatibility
- ▶ Market Failure
- ▶ Mechanism Design

Bibliography

Akerlof, G. 1970. The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84: 488–500.

Coase, R. 1960. The problem of social cost. *Journal of Law and Economics* 3: 1–44.

Compte, O., and P. Jehiel. 2006. *Veto constraint in mechanism design: Inefficiency with correlated types*. Mimeo: Paris-Jourdan Sciences Economiques and University College London.

Cramton, P., R. Gibbons, and P. Klemperer. 1987. Dissolving a partnership efficiently. *Econometrica* 55: 615–632.

Crémer, J., and R. McLean. 1985. Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica* 53: 345–362.

Crémer, J., and R. McLean. 1988. Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica* 56: 1247–1257.

Gomes, A., and P. Jehiel. 2005. Dynamic processes of social and economic interactions: On the persistence of inefficiencies. *Journal of Political Economy* 113: 626–667.

Jehiel, P., and B. Moldovanu. 1999. Resale markets and the assignment of property rights. *Review of Economic Studies* 66: 971–991.

Jehiel, P., and B. Moldovanu. 2001. Efficient design with interdependent valuations. *Econometrica* 69: 1237–1259.

Jehiel, P., and A. Pauzner. 2006. Partnership dissolution with interdependent values. *RAND Journal of Economics* 37: 1–22.

Jehiel, P., B. Moldovanu, and E. Stacchetti. 1996. How (not) to sell nuclear weapons. *American Economic Review* 86: 814–829.

Johnson, S., J. Pratt, and R. Zeckhauser. 1990. Efficiency despite mutually payoff-relevant private information: The finite case. *Econometrica* 58: 873–900.

Maskin, E. 1992. Auctions and privatization. In *Privatization*, ed. H. Siebert. Kiel: Institut für Weltwirtschaften der Universität Kiel.

McAfee, P., and D. Vincent. 1997. Sequentially optimal auctions. *Games and Economic Behavior* 18: 246–276.

Milgrom, P. 2004. *Putting auction theory to work*. Cambridge: Cambridge University Press.

Myerson, R. 1981. Optimal auction design. *Mathematics of Operations Research* 6: 58–73.

Myerson, R., and M. Satterthwaite. 1983. Efficient mechanisms for bilateral trading. *Journal of Economic Theory* 28: 265–281.

Robert, J. 1991. Continuity in auction design. *Journal of Economic Theory* 55: 169–179.

Samuelson, W. 1984. Bargaining under asymmetric information. *Econometrica* 52: 995–1005.

Zheng, C. 2002. Optimal auction with resale. *Econometrica* 70: 2197–2224.

Research and Experimental Development (R&D) and Technological Innovation Policy

Henri Delanghe and Ugur Muldur

Abstract

Research and experimental development (R&D), when appropriately valorised, lead to technological innovation in the form of new products and processes, which contribute to growth, competitiveness and job creation, and which produce other societal benefits. Because of market failures, the private sector, left to its own devices, invests in R&D in sectors not always fully aligned with, and at levels below, the socially desirable, and is unable to fully valorise its research output, which justifies public intervention. The latter needs to be thought through carefully based on *ex ante* impact assessment informed by credible *ex post* evaluation.

Keywords

Economic growth; Economic policy; Private sector; R&D; Technological innovation

JEL Classifications

O3; O31; O320; O380

R

Defining 'Research and Experimental Development' (R&D)

The focus of this article is on public policy in support of research and experimental development (R&D) and technological innovation. According to the Frascati Manual of the Organisation for Economic Cooperation and Development (OECD), a basic reference in the R&D field with respect to definition and measurement, 'R&D comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications', and cover three activities: basic research, applied research, and experimental development (a distinction currently losing ground in the scientific community, if not in research administration, because of the dividing lines between basic and applied research, for instance, becoming blurred and concepts like frontier research gaining ground).

> 'Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view. Applied research is also original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective. Experimental development is systematic work, drawing on existing knowledge gained from research and/or practical experience, which is directed to producing new materials, products or devices, to installing new processes, systems and services, or to improving substantially those already produced or installed. R&D covers both formal R&D in R&D units and informal or occasional R&D in other units'. (OECD 2002) (See also Technology in this dictionary.)

Defining 'Innovation'

According to the OECD's Oslo Manual, the Frascati Manual's equivalent in the field of innovation, 'an innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations'. Technological innovation concerns mainly product and process innovation. A product innovation is defined as 'the introduction of a good or service that is new or significantly improved with respect to its characteristics or intended uses. This includes significant improvements in technical specifications, components and materials, incorporated software, user friendliness or other functional characteristics'. A process innovation is defined as 'the implementation of a new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software'. Many other forms of innovation (e.g. marketing innovation, organisational innovation) exist besides technological innovation. These fall outside the scope of this article (OECD 2005).

R&D and Innovation Can Produce Large-Scale Societal Impacts

R&D and technological innovation can produce large-scale private and public benefits and impacts. For instance, modern mainstream economic theory (e.g. Solow 1957; Abramowitz 1986; Romer 1990; Baumol 2002; Aghion and Howitt 2006) has recognised for quite some time that technological progress and innovation are the main engines of economic growth: markets' competitive pressures give rise to R&D and technological innovation involving new products and new processes that increase productivity, the source of sustainable economic growth.

Basic research generates direct economic benefits in the form of useful new information; new instrumentation and methodologies; skills that yield economic benefits when individuals move from basic research carrying codified and tacit knowledge; access to networks of experts and information; expertise in the solving of complex technological problems; and spin-off companies (Martin et al. 1996).

The rate of return for public R&D lies between 20% and 50% (e.g. Griliches 1995). The private rate of return to company R&D commonly falls in the 20–30% range and the social rates of return are in the 30–40% range (e.g. the overviews of available evidence in Margolis and Kammen 1999; Griffith 2000; Jones and Williams 1998). The ability to innovate is positively related to firms' export performance.

Empirical work has established robust relations at the macroeconomic level between investment in innovation and productivity. Positive and significant relations have been established, for instance, between increases in public R&D and productivity growth, increases in business R&D and productivity growth, increases in business R&D investment and multi-factor productivity growth, increases in business investment in intangible capital (which includes scientific R&D) and overall economic labour productivity growth, increases in total R&D and output per capita growth, and increases in investment in intangible assets (which includes scientific R&D) and multi-factor productivity growth (e.g. Guellec and van Pottelsberghe de la Potterie 2001). In the process, technological change boosts employment, though it is increasingly recognised that the latter holds under certain conditions only (e.g. Van Reenen 1997).

Self-evidently, R&D and technological innovation can also contribute to the resolution of large-scale societal challenges, such as disease, ageing, climate change and energy insecurity. It used to be assumed that not all of these benefits could be achieved at the same time and that they were to a certain extent mutually exclusive. Greening the economy, for instance, would mean sacrificing growth. It is now increasingly recognised, however, that these benefits of R&D and technological innovation can be achieved together (for instance, OECD 2011). (See also Economic growth in this dictionary.)

The Justification for Public Intervention

Given these potential large-scale growth, competitiveness, labour and societal benefits of R&D and technological innovation, the focus, level and output of investment in research and innovation by the private sector, left to itself, are generally considered to be sub-optimal from a societal point of view. This is due to so-called market failures that prevent business from investing in research and innovation in areas and at levels that are socially desirable, or from fully valorising its research output. These market failures concern, for instance, risk generated by uncertainty and high costs, in combination with spillovers; imperfect or asymmetric information; and systemic or coordination failures. They are generally speaking more acute the further the research concerned is removed from the market. As for risk, at the start of a research project, it is not at all sure that the research efforts undertaken will actually result in new knowledge and innovation (uncertainty). At the same time, the cost of R&D is high and rising: it becomes more expensive to carry out research because the price of research inputs is rising and the life cycle of products is shortening dramatically. Even when R&D and technological innovation projects are successful and generate new knowledge and innovation, they may not be profitable, because markets do not exist (yet), because competing products are not priced correctly (negative externalities, e.g. pollution effects, are not taken account of), or because knowledge spills over and not all benefits can be appropriated privately so that social rates of return exceed private ones (public goods).

Companies may be reluctant to invest in research out of fear that the new products they may come up with may make obsolete the products they are currently deriving substantial profits from. In the case of imperfect/asymmetric information, companies may not get access to the required capital because financiers, unlike the companies involved, fail to understand the potential of the R&D and innovation projects concerned. Where there are systemic/coordination failures, the innovation systems literature argues that what matters for an economy's innovation performance are the linkages and flows of information between the different actors in the innovation system and these linkages and flows are often suboptimal preventing the full valorisation of research results. (See also Market failure in this dictionary.)

There are Limits to the Public Support that can be Provided: State Aid Considerations

To the extent that R&D and technological innovation produce societal benefits, and that market failures affecting the level, focus and research output valorisation of private sector investment in R&D and technological innovation exist, a justification exists for a measure of public intervention in the form of public support for R&D and technological innovation. For this reason, state aid for R&D and innovation is treated as a special case in competition policy and allowed subject to certain limitations.

Public Policy Should Aim for the Achievement of 'Additionality'

The literature shows that public support for R&D and technological innovation can make a real difference, i.e. achieve tangible 'additionality', defined as the difference in outcome between scenarios including and excluding public support, testifying to the existence of market failures. As already mentioned above, public research generates direct economic benefits. Public R&D also increases the pay-off to private R&D and supports innovation. For instance, increases in university R&D lead to increases in corporate patenting. A sizeable share of new products and new processes would have been developed with a substantial delay in the absence of academic research. A significant share of private sector innovations are partially based on public sector research, and university and government R&D laboratory research is critical to industrial R&D in a small number of industries. Public research also affects industrial R&D across much of the manufacturing sector (e.g. Mansfield 1998; Tijssen 2002; Cohen et al. 2002). Furthermore, high-quality public research attracts private R&D. For instance, the number of relevant scientific publications by scientists based in the host country has a substantial positive impact on the propensity to conduct foreign R&D; multinational corporations' R&D location decisions and investment levels are influenced by scientific output and institutional quality; industry-financed R&D is positively associated with both the per capita number of highly cited researchers and expenditure on higher education R&D; private sector R&D laboratories cluster disproportionately around highly rated university research departments; and small firms benefit from localised university–industry knowledge transfer (e.g. Belderbos et al. 2009; Dosi *et al*. 2009; Guimón 2008; Abramovsky et al. 2007). Fourth, public subsidies for private research increase the total amount of research expenditure (input additionality, crowding-in effect, leverage effect) (e.g. Czarnitzki and Licht 2006; Guellec and van Pottelsberghe de la Potterie 2000; European Commission 2004). Finally, public subsidies for private research increase the total amount of innovation (output additionality) and also achieve other types of additionality including project additionality (pure or partial), scope or challenge additionality, network or cooperation additionality, acceleration additionality, cognitive capacity additionality, management additionality and follow-up additionality (e.g. Czarnitzki and Licht 2006).

The R&D Policy Toolbox

The achievement of additionality effects and their maximisation requires the careful design of policies to avoid government failures. This requires first of all the correct choice of policy instrument from the R&D and innovation policy toolbox, which is large. Various categorisations of policy tools exist. The OECD, for instance, makes a distinction between population-targeted (e.g. focused on small and medium-sized enterprises (SMEs) and young firms, as well as particular sectors) versus generic (non-population-targeted) instruments; technology-targeted versus generic (non-technology-targeted) instruments; financial versus non-financial instruments; direct (e.g. credit loans and guarantees, repayable advances, competitive grants, technology consulting services and extension programmes, innovation vouchers, equity financing and venture capital investments etc.)

versus indirect (e.g. tax incentives on R&D and innovation, which may be both expenditure-based (R&D tax credits, R&D tax allowances and payroll withholding tax credit for R&D wages) or income-based (preferential rates on royalty income and other income from knowledge capital)) financing instruments; competitive versus non-competitive instruments; and supply-side versus demand-side (aiming to stimulate and articulate public demand for innovative solutions and products from firms) instruments (OECD 2010; OECD 2012).

The Need for Horizontal and Vertical Policy Coordination

Whatever the specific policy tool chosen from the R&D and innovation policy toolbox, it never operates in isolation. It is always part of a broader portfolio of different R&D and innovation, and also of other policy measures. All these measures need to be coordinated, both 'horizontally' and 'vertically'. As for horizontal coordination, coordination is needed between different R&D and innovation policy instruments and with other knowledge triangle policies. Further horizontal coordination is needed between knowledge triangle and other, e.g. structural policies. In other words, the broader framework conditions for innovation – which touch upon education, the macroeconomic environment, the functioning of product, labour and financial markets, infrastructure, the regulatory framework, the protection of intellectual property rights etc. – need to be supportive. Vertical coordination is also needed between regional, national, supranational, inter-governmental and international policies (OECD 2010).

Impact Assessment, a Tool for Carefully Considered Policy Development

A key tool for ensuring that policy instruments respond to market failures, are suitable and well-coordinated, and maximize their impact, consists of *ex ante* impact assessment (IA). IA is a key tool to ensure that new initiatives and legislation are prepared on the basis of transparent, comprehensive and balanced evidence. IA is focused on a number of key questions. What is the nature and scale of the problem, how is it evolving, and who is most affected by it? What are the views of the stakeholders concerned? What objectives should be set to address the problem? What are the main policy options for reaching these objectives? What are the likely economic, social and environmental impacts of those options? How do the main options compare in terms of effectiveness, efficiency and coherence in solving the problems? How could future monitoring and evaluation be organised?

The Challenge of *ex post* Evaluation in the Field of R&D and Innovation

IA needs to be fed by a wide range of evidence. A key part of this evidence needs to be derived from monitoring and evaluation exercises. In this sense, IA constitutes the bridge tween *ex post* evaluation and new policy development, thereby closing the policy cycle, which comprises:

- Agenda-setting and problem identification. At this stage, policy problems are defined and policy issues are raised, introduced to the political stage by different governmental institutions, individuals, interest groups or specific events.
- Policy formulation. At this stage, analysis and politics determine how the agenda item is translated into legislation. This stage encompasses several stages, including the development and consideration of alternative policy options, the selection of a preferred option and its adoption.
- Implementation. At this stage, the adopted policy is implemented – administered and enforced – by the bureaucracy, by an agency of government. The bureaucracy or agency interprets the policy into a concrete set of actions and makes judgments as to intent, goals, timetables, program design, reporting methods.

- Evaluation. The implementation of policy is evaluated to assess what is working and what is not. The impacts of the policy are assessed. If goals exist, the effectiveness of the policy and its components can be determined. Side-effects must also be discovered and reckoned.

Ex post evaluation of R&D and technological innovation policy encounters its own problems, however. The core problem is that it is difficult to define precisely the concrete contribution made to a particular innovation or societal impact by a particular research project. Research projects achieve intended and unintended results, and only the intended and 'additional' ones can be linked back to policy. Research results take time to generate impact and also spill over to and impact upon other players, sectors and countries, so that impacts are rarely direct, which complicates the identification of useful inputs. Particular innovation or societal impacts are usually the result not of a single research project but of a series or portfolio of research projects, once more complicating the identification of useful inputs (OECD 2008).

Disclaimer

All views expressed herein are entirely of the authors, do not reflect the position of the European Institutions or bodies and do not, in any way, engage any of them.

See Also

- ► European Union (EU) Research and Experimental Development (R&D) Policy
- ► Information Sharing Among Firms
- ► Research Joint Ventures

Bibliography

Abramowitz, M. 1986. Catching up, forging ahead, and falling behind. *Journal of Economic History* XLV(2): 385–406.

Abramovsky, L., R. Harrison, and H. Simpson. 2007. University research and the location of business R&D. *Economic Journal* 117(519): 114–141.

Aghion, P., and P. Howitt. 2006. Appropriate growth policy: A unifying framework. *Journal of the European Economic Association* 4: 269–314.

Baumol, W.J. 2002. *The free-market innovation machine. Analyzing the growth miracle of capitalism*. Princeton/Oxford: Princeton University Press.

Belderbos, R., B. Leten, and S. Suzuki. 2009. *Does excellence in academic research attract foreign R&D?* Strategy and Innovation, Working Paper Series. Leuven: Department of Managerial Economics, Katholieke Universiteit.

Cohen, W.M., R.R. Nelson, and J.P. Walsh. 2002. Links and impacts: the influence of public research on industrial R&D. *Management Science* 48(1): 1–23.

Czarnitzki, D., and G. Licht. 2006. Additionality of public R&D grants in a transition economy. The case of Eastern Germany. *The Economics of Transition* 14(1): 101–131.

Dosi, G., P. Llerena, and M. Sylos Labini. 2009. Does the 'European Paradox' still hold? – did it ever? In *European science and technology policy: Towards integration or fragmentation?* ed. H. Delanghe, U. Muldur, and L. Soete, 214–236. Cheltenham: Edward Elgar.

European Commission. 2004. *European Competitiveness Report 2004*. Enterprise and Industry Publications, European Communities.

Griliches, Z. 1995. R&D and productivity. In *Handbook of industrial innovation*, ed. P. Stoneman, 52–89. London: Blackwell Press.

Griffith, R. 2000. *How important is business R&D for economic growth and should the government subsidise it?* Briefing Note No. 12. London: The Institute for Fiscal Studies.

Guellec, D., and B. van Pottelsberghe de la Potterie. 2000. *The impact of public R&D expenditure on business R&D*, STI Working Papers 2000/4. Paris: OECD.

Guellec, D., and B. van Pottelsberghe de la Potterie. 2001. *R&D and productivity growth: Panel data analysis of 16 OECD countries*, STI Working Papers 2001/3. Paris: OECD.

Guimón, J. 2008. *Government Strategies to Attract R&D-Intensive FDI*. OECD Global Forum on International Investment, 27 and 28 March.

Jones, C.I., and J.C. Williams. 1998. Measuring the social return to R&D. *The Quarterly Journal of Economics* 113(4): 1119–1135.

Mansfield, E. 1998. Academic research and industrial innovation: An update of empirical findings. *Research Policy* 26(7–8): 773–776.

Margolis, R.M., and D.M. Kammen. 1999. Evidence of under-investment in energy R&D in the United States and the impact of Federal policy. *Energy Policy* 27: 575–584.

Martin, B.R., A. Salter, D. Hicks, K. Pavitt, J. Senker, M. Sharp, and N. von Tunzelmann. 1996. *The relationship between publicly funded basic research and economic performance*. London: SPRU Review. HM Treasury.

OECD. 2002. *Frascati manual*. Paris: OECD.
OECD. 2005. *Oslo manual*. Paris: OECD.
OECD. 2008. *Science, technology and industry outlook 2008*. Paris: OECD.
OECD. 2010. *Science, technology and industry outlook 2010*. Paris: OECD.
OECD. 2011. *Towards green growth*. Paris: OECD.
OECD. 2012. *Science, technology and industry outlook 2012*. Paris: OECD.
Romer, P.M. 1990. Endogenous technological change. *Journal of Political Economy* 98(5): S71–S102.
Solow, R.M. 1957. Technical change and the aggregate production function. *The Review of Economics and Statistics* 39: 312–320.
Tijssen, R.J.W. 2002. Science dependence of technologies: Evidence from inventions and their inventors. *Research Policy* 31: 509–526.
Van Reenen, J. 1997. Employment and technological innovation: Evidence from UK manufacturing firms. *Journal of Labour Economics* 15(2): 255–284.

Research Joint Ventures

Bruno Cassiman

Abstract

A research joint venture (RJV) is an agreement between two or more partners to perform research and development (R&D). RJVs provide a mechanism to bridge the divide between the optimal public R&D policy – free dissemination of knowledge – and private incentives to invest in R&D – appropriation of returns to investments. Three important issues related to appropriation of returns to R&D condition the private incentives to form a RJV: coordination of R&D investments between RJV partners, free-riding inside and outside the RJV, and information sharing between RJV partners.

Keywords

Adverse selection; Antitrust; Cartels; Collusion; Competition policy; Externalities; Free riding; Incentive compatibility; Industrial organization; Information sharing among firms; Innovation; Innovation policy; Mergers; Public goods; R&D contracting; R&D cooperation; Repeated games; Research and development (R&D); Research joint ventures; Risk sharing; Spillovers; Technology licensing; Transaction costs; Transfer of technology

JEL Classifications

L24

A *research joint venture* (RJV) is an agreement between two or more partners to perform research and development (R&D), where each partner has an active role in the generation of new knowledge and technology. As such, a RJV is distinct from the *ex ante* or *ex post* agreement to acquire knowledge or technology as in *R&D contracting* or the *licensing* of technology respectively. Many times RJV and *R&D cooperation* are used as synonyms in the literature.

Two features distinguish R&D from ordinary capital investments. First, R&D is a public good (Arrow 1962). The use by one firm of the information produced by its R&D investments does not diminish the amount of information available to other firms. Second, and related to its public good nature, R&D investment is plagued by an *externality* problem. Firms investing in R&D typically cannot fully appropriate the returns to their own R&D investments. This tends to reduce the incentive to invest in R&D when firms act non-cooperatively (Spence 1984; d'Aspremont and Jacquemin 1988).

Both of these characteristics of R&D investments have a profound impact on the optimal way of organizing R&D, as they affect the incentives to invest in R&D.

From a welfare perspective the optimal economy-wide organization would involve the free distribution of the knowledge produced by these R&D investments. However, such a policy would provide little incentive for private investment in R&D in the first place. RJVs provide a mechanism to bridge this divide between public policy and private incentive.

Incentives to Form RJVs

Given the public-good nature of R&D, firms do have an incentive to jointly develop technology and share the costs and risk of these projects. Mariti and Smiley (1983) provide evidence for the importance of cost and risk-sharing for the success of R&D cooperation. Developing new technology from scratch implies incurring a high (fixed) cost. Transferring and sharing knowledge that is already developed has a low (marginal) cost. Therefore, firms with complementary products (Röller et al. 1997) or complementary knowledge (Sakakibara 1997) have an incentive to form RJVs to share knowledge for the development of new products. Furthermore, from a *transaction costs* perspective R&D collaboration allows access to specialized and complementary know-how, while at the same time allowing for a transfer of technology at lower transaction costs than with arm's length arrangements. As a result the total cost of developing new knowledge through a RJV is reduced (Pisano 1990; Oxley 1997).

While knowledge transfer and cost sharing provide the most common and trivial incentive for the formation of RJVs, the industrial organization literature emphasizes competitive motives for engaging in R&D cooperation and RJVs. R&D is imperfectly appropriable and R&D results, therefore, leak out involuntarily to rival firms. These models concentrate on horizontal R&D cooperation among rival companies as a mechanism to internalize these *spillovers*. The R&D process is represented as a two-stage, non-tournament model where in a first stage firms make R&D investments that (strategically) affect second-stage output market decisions through either a cost-reducing or a demand-enhancing effect. Firms can cooperate – form an RJV – in the R&D stage, but may continue to compete in the product market (for example, Katz 1986; d'Aspremont and Jacquemin 1988; Kamien et al. 1992; Suzumura 1992; Leahy and Neary 1997). From this literature we discern three important issues conditioning the interrelation between the profitability of RJVs and spillovers: *coordination, free-riding* and *information sharing*.

Coordination

Cooperation in these models is typically industry-wide and takes the form of firms coordinating R&D choices in order to maximize joint profits. As a result investment in R&D in an RJV is increasing in the level of the spillover as the firms internalize the positive effect these spillovers have on their partners. In addition, when spillovers are high enough – that is, above a critical level – coordination in R&D will result in higher R&D investment than in non-coordinating firms. At the critical spillover level, the profitability of cooperative and non-cooperative R&D strategies coincides. (When goods are substitutes, the level of product differentiation and the number of rivals are important parameters that determine the critical spillover level; de Bondt et al. 1992.)

Coordination through joint profit maximization without incurring any explicit costs to R&D cooperation increases the firms' profitability in these models. But, more importantly, spillovers increase the profitability of cooperation in R&D. Furthermore, for spillovers above the critical level, firms have an increasing incentive to engage in R&D coordination (De Bondt and Veugelers 1991). This means that, when spillovers are high enough, firms have an increasing incentive to engage in R&D coordination. Such cooperation would furthermore enhance welfare as R&D investment and market output increase.

Free Riding

Most models focus on the welfare and profitability of R&D cooperation, ignoring the stability of such cooperation. The stability of RJVs can be threatened by free riding of non-participating companies on the output of the venture, or by free riding by partners who may conceal their technological expertise while trying to absorb as much as possible of the partner's knowledge (Shapiro and Willig 1990). Kesteloot and Veugelers (1994) find that cooperative agreements that are profitable, and at the same time also stable, require involuntary – outgoing – spillover levels that are not too high. (Using a repeated game, cheating can be prevented by grim-trigger strategies specifying an eternal dissolution of an industry-wide venture. An alternative approach to solve the internal stability problem is through the organizational design of the venture. Perez-Castrillo and Sandonis 1996, characterize incentive compatible and individually rational contracts that lead to disclosure of knowledge and, hence, the formation of profitable research joint ventures.) Hence, although higher spillover levels increase the profits from cooperation through coordination, they also increase the profits from cheating by a partner and from free riding by an outsider to the cooperative agreement. Therefore, cooperative ventures become more profitable the more able firms are to restrict outgoing spillovers by protecting their information while selectively sharing information with partners.

Information Sharing

Some models take into account the fact that firms can indeed manage spillovers by voluntarily increasing the spillovers among cooperating partners. Such information sharing is found to further increase the profitability of cooperation in R&D. In addition, information sharing not only increases the profitability of R&D cooperation; it also makes such agreements more stable. Eaton and Eswaran (1997) show that, when technology trading cartels are not necessarily industry-wide, information sharing is an even stronger stabilizing force. In this case a much stronger punishment can be specified, namely, the ejection of the cheating firm from a technology-trading coalition, followed by the continuation of information sharing by the non-cheating members. Similarly, De Bondt and Wu (1997) find that information sharing produces larger coalition sizes that are both internally and externally stable.

Katsoulacos and Ulph (1998) explicitly model the choice of spillovers by cooperating and non-cooperating firms, and find that RJVs will always share at least as much information as non-cooperating firms because the former maximize joint profits. When firms act non-cooperatively, however, one would expect that the aim is to minimize the creation of spillovers – the *outgoing* spillovers – through the use of effective legal and strategic protection measures while at the same time to maximize the *incoming* spillovers. Kamien and Zang (2000) show that firms that coordinate their R&D expenditures maximize information flows – their *incoming* spillovers – through the choice of very broad research directions for the RJV. If the firms cannot coordinate their R&D expenditures, they are more concerned about managing their *outgoing* spillovers by choosing a more narrow research approach. This result emphasizes a potential dual role of spillovers: outgoing spillovers which might jeopardize the cooperative agreement, and incoming spillovers which increase the attractiveness of the cooperative agreement. In an empirical paper Cassiman and Veugelers (2002) indeed show that *incoming* spillovers and appropriability have important and separately identifiable effects: firms with higher incoming spillovers and better appropriation have a higher probability of cooperating in R&D.

RJVs and Social Welfare

When firms are allowed to form RJVs, R&D investments increase with the level of spillovers, exceeding the non-cooperative investment level when the spillovers are substantial (d'Aspremont and Jacquemin 1988). Competing firms that cooperate in R&D might thus increase not only profits but also welfare when the spillovers are substantial. Policywise, a case can then be made for allowing RJVs to form when spillovers are high. However, when spillovers are low firms acting non-cooperatively with respect to R&D bring about higher welfare than when allowed to form an RJV (Suzumura 1992). The only effect of a RJV in this case is to reduce R&D competition, which in turn decreases welfare (Katz 1986). (It has often been suggested that RJVs might also facilitate collusion in the output market. A necessary condition for a RJV to be welfare improving in this case is that total R&D investments increase. Martin 1997, analyses the increased potential for tacit collusion in RJVs, while Yi 1995, looks at the welfare effects of product market collusion by an industry-wide RJV. Greenlee and Cassiman 1999, discuss the effects of collusion in the output market on RJV formation.) This theoretical finding has fuelled the debate on the issue of relaxing antitrust regulation with respect to RJVs. In evaluating cooperative R&D, regulators often use the same “rule of reason” as in the case of mergers. Given the dynamic nature of R&D, insensitive application of static merger guidelines may lead to undesirable outcomes (Ordover and Willig 1985). Appropriate standards for evaluating RJVs should be developed. Jorde and Teece (1990) propose the creation of an administrative procedure for evaluating and possibly certifying cooperative R&D agreements in order to establish a safe harbour from antitrust litigation. But Shapiro and Willig (1990) argue that this would provide too much protection to RJVs, especially because the regulator needs a great deal of information to evaluate a RJV, and much of this information might be proprietary.

Policymakers have attempted to address these issues. In the USA firms can register their RJVs under the National Cooperative Research Act (NCRA). By registering under the NCRA, firms become exempt from treble damages under antitrust regulation. However, cooperative R&D ventures need not register under the NCRA. In that case, they are liable under the usual antitrust regulation. Scott (1988) actually notes that cooperative research registered under the NCRA predominantly falls into industries without severe appropriability problems, while supposedly welfare-enhancing RJVs do not seem to register, leading to a suspicion of adverse selection of RJVs under the NCRA (Cassiman 2000).

In Europe the 1986 Single European Act amendments to the Treaty of Rome gave the Community specific responsibility for strengthening “the scientific and technological basis of European industry”. In addition to the EEC block exemption of Article 85(1) of the EC treaty for cooperative ventures in R&D, a variety of programmes were initiated, many of which explicitly fostered inter-firm cooperation tied to Community funding for part of the R&D costs of the proposed projects (Martin 1996). Nevertheless, the debate on the exact implementation of these policies is still ongoing and has initiated a broader debate on the interaction between innovation policy and competition policy.

Conclusion

While the industrial organization models of RJVs have focused on imperfect appropriation among competitors, several areas for research on RJVs remain thoroughly unexplored. First, empirical work has indicated that most RJVs are formed with customers, suppliers or research organizations rather than with competitors. Recent empirical work has started to tackle the issue of different types of partners for the RJVs, but little theoretical work has followed (Fritsch and Lukas 2001; Belderbos et al. 2004; Veugelers and Cassiman 2005).

Second, and related, we still know very little about the actual effect of engaging in RJVs on firm (innovation) performance. Brandstetter and Sakakibara (1998) find some evidence of the formation of RJVs on research productivity, and Belderbos, Carree and Lokshin (2004) show that cooperation in R&D leads firms to generate more sales from products that are new to the market. But most empirical studies interpret R&D cooperation as an indirect indication of RJV’s profitability. To really uncover the incentives to engage in RJVs, we need to understand how RJVs improve the innovation performance of firms relative to alternative organizational forms.

Finally, little progress has been made yet in understanding the organization of RJVs from a theory of the firm perspective. Why would firms make joint investments in R&D and share property rights and decision rights over the outcome of future research outcomes? When is this efficient or when does it enhance the competitiveness of firms?

See Also

▶ Externalities
▶ Information Sharing Among Firms

Bibliography

Arrow, K. 1962. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, ed. R. Nelson. Princeton: Princeton University Press.

Belderbos, R., M. Carree, and B. Lokshin. 2004. Cooperative R&D and firm performance. *Research Policy* 33: 1477–1492.

Brandstetter, L., and M. Sakakibara. 1998. Japanese research consortia: A microeconomic analysis of industrial policy. *Journal of Industrial Economics* 46: 207–233.

Cassiman, B. 2000. Research joint ventures and optimal R&D policy with asymmetric information. *International Journal of Industrial Organization* 18: 283–314.

Cassiman, B., and R. Veugelers. 2002. R&D cooperation and spillovers: Some empirical evidence from Belgium. *American Economic Review* 92: 1169–1184.

d'Aspremont, C., and A. Jacquemin. 1988. Cooperative and non-cooperative R&D in duopoly with spillovers. *American Economic Review* 78: 1133–1137.

De Bondt, R., and R. Veugelers. 1991. Strategic investment with spillovers. *European Journal of Political Economy* 7: 345–366.

De Bondt, R., and C. Wu. 1997. Research joint venture cartels and welfare. In *R&D cooperation: Theory and practice*, ed. J. Poyago-Theotoky. London: Macmillan.

De Bondt, R., P. Slaets, and B. Cassiman. 1992. Spillovers and the number of rivals for maximum effective R&D. *International Journal of Industrial Organization* 10: 35–54.

Eaton, B., and M. Eswaran. 1997. Technology trading coalitions in supergames. *RAND Journal of Economics* 28: 135–149.

Fritsch, M., and R. Lukas. 2001. Who cooperates on R&D? *Research Policy* 30: 297–312.

Greenlee, P., and B. Cassiman. 1999. Product market objectives and the formation of research joint ventures. *Managerial and Decision Economics* 20: 115–130.

Jorde, T., and D. Teece. 1990. Innovation and cooperation: Implications for competition and antitrust. *Journal of Economic Perspectives* 4(3): 75–96.

Kamien, M., and I. Zang. 2000. Meet me halfway: Research joint ventures and absorptive capacity. *International Journal of Industrial Organization* 18: 995–1012.

Kamien, Morton I., E. Müller, and I. Zang. 1992. Research joint ventures and R&D cartels. *American Economic Review* 82: 1293–1306.

Katsoulacos, Y., and D. Ulph. 1998. Endogenous spillovers and the performance of research joint ventures. *Journal of Industrial Economics* 46: 333–358.

Katz, M. 1986. An analysis of co-operative research and development. *RAND Journal of Economics* 17: 527–543.

Kesteloot, K., and R. Veugelers. 1994. Stable R&D co-operation with spillovers. *Journal of Economics and Management Strategy* 4: 651–672.

Leahy, D., and P. Neary. 1997. Public policy towards R&D in oligopolistic industries. *American Economic Review* 87: 642–662.

Mariti, P., and R. Smiley. 1983. Co-operative agreements and the organisation of industry. *Journal of Industrial Economics* 38: 183–198.

Martin, S. 1996. Protection, promotion and cooperation in the European semiconductor industry. *Review of Industrial Organization* 11: 721–735.

Martin, S. 1997. Public policy toward cooperation in research and development: The European Union, Japan, the United States. In *Competition policy in the global economy*, ed. L. Waverman, W. Comanor, and A. Goto. London: Routledge.

Ordover, J., and R. Willig. 1985. Antitrust for high-technology industries: Assessing research joint ventures and mergers. *Journal of Law and Economics* 28: 311–333.

Oxley, J.E. 1997. Appropriability hazards and governance in strategic alliances: A transaction costs approach. *Journal of Law, Economics and Organization* 13: 387–409.

Perez-Castillo, D., and J. Sandonis. 1996. Disclosure of know-how in research joint ventures. *International Journal of Industrial Organization* 15: 51–75.

Pisano, G. 1990. The R&D boundaries of the firm: An empirical analysis. *Administrative Science Quarterly* 35: 153–176.

Röller, L., M.Tombak, and R. Siebert. 1997. *Why firms form research joint ventures: Theory and evidence*, Discussion Paper No. 1654, CEPR.

Sakakibara, M. 1997. Heterogeneity of firm capabilities and cooperative research and development: An empirical examination of motives. *Strategic Management Journal* 18: 134–164.

Scott, J.T. 1988. Diversification versus cooperation in R&D investment. *Managerial and Decision Economics* 9: 173–186.

Shapiro, C., and R. Willig. 1990. On the antitrust treatment of production joint ventures. *Journal of Economic Perspectives* 4(3): 113–130.

Spence, M. 1984. Cost reduction, competition and industry performance. *Econometrica* 52: 101–121.

Suzumura, K. 1992. Cooperative and noncooperative R&D in an oligopoly with strategic commitments. *American Economic Review* 82: 1307–1320.

Veugelers, R., and B. Cassiman. 2005. R&D cooperation between firms and universities: Some empirical evidence from Belgian manufacturing. *International Journal of Industrial Organization* 23: 355–379.

Yi, S. 1995. *R&D cooperation, product-market collusion and welfare*, Working paper, Department of Economics, Dartmouth College.

Reservation Price and Reservation Demand

Ian Steedman

Keywords
Factor supply; Free goods; Labour supply; Opportunity cost doctrine; Real cost doctrine; Reservation demand; Reservation price; Wicksteed, P. H

JEL Classifications
D0

The simplest example of a reservation price is that price below which an owner will refuse to sell a particular object in an auction. Since the owner could always, in principle, enforce such a price by outbidding everyone else, this leads immediately to the more general concept of a reservation price as that price at which the owner of a fixed stock will choose to *retain* some given amount from that stock, rather than supply more, and of the amount retained as the owner's 'reservation demand' at the price in question. Considering alternative hypothetical prices, one sees that the owner's supply curve of the commodity can equally well be described as an 'own (reservation) demand' curve, where 'supply' and 'own demand' sum identically to the given stock. The same is naturally true of the market supply curve. Thus consider the standard example of the determination of the price of first-edition copies of a certain old book. A demand curve may be drawn up for those who at present own no copies. Taking account of each present owner's reservation price (or prices for those who possess more than one copy), we may also draw up a supply curve. (Of course 'supply' by present owners may be negative at low prices.) Confrontation of the demand and supply curves will then show the market-clearing price. Equally, however, we could have drawn up the 'reservation demand' curve of present owners, summed it with the demand curve of non-owners and then confronted the 'total' demand curve with the given stock. Since 'supply' and 'reservation demand' sum identically to total stock, at every price, the alternative diagram inevitably shows the same market-clearing price as does the first; it does not show the number of books traded, however.

It will be clear that an agent's reservation price for any type of commodity can be expected to depend on one or more of the following considerations: the scope for direct 'own use' of the commodity; the agent's present need for liquidity; the agent's other resources; the perishability of the commodity and thus the various elements of storage costs (including interest costs); expectations about future prices, there being always a speculative element in the reservation price of any commodity which is not immediately perishable. These considerations all emerge in theories of 'factor supply', for example in the theory of household labour supply. Since 'labour time' is instantly perishable, there is no strictly speculative element to take into account (although someone seeking work may refuse a particular job offer because the wage offered is below a 'reservation wage' based on expectations as to the wage that can be obtained after further job searching). The conventional theory is, however, firmly based on viewing labour supply in terms of the 'reservation demand' for time not spent in market employment, and it is this that leads to the familiar argument that the income effect of a 'wage' change can both be large and contrary to the substitution effect, with the result that labour supply may be either positively or negatively related to the level of the 'wage'. Analogous arguments bear on the supply of land services by landowners who have an 'own use' for their land, on the supply of agricultural products, and so on. The reservation price concept is also useful in the context of privately owned natural resources, a context which introduces two further determinants of reservation price. The lowest price at which a natural resource owner will be prepared to extract the resource will naturally depend on extraction costs, both the present extraction costs and those expected in the future; it will also depend on the expected growth rate, if any, of the resource. It is to be noted that the 'neoclassical rule of free goods' would never have to be applied to primary inputs for which (a) there was a positive price below which supply would be zero, and (b) demand at a zero price would be positive (both conditions holding for all prices of other commodities).

It was noted above, in connection with the market for first-edition copies of a book, that the 'total' demand curve diagram gives the same information with respect to price, and less information with respect to quantity, than does the more conventional supply and demand diagram. How then could P.H. Wicksteed – whose name is so strongly associated with the concept of a supply curve being merely a 'reversed demand curve' – have been so insistent that the former diagram is actually *superior* to the conventional one? (See Wicksteed 1910, Book II, Ch. IV, and 1914.) Because the 'total' demand curve diagram emphasizes the idea that essentially the *same* kind of forces underlie the conventional supply curve as underlie the usual demand curve, thus breaking down the idea that there is an asymmetry in market forces, with subjective factors being dominant on the 'demand side' and objective ones on the 'supply side'. The diagram in which a single demand curve (inclusive of reservation demand) confronts a fixed supply is at once congenial to any author both seeking to stress the subjective elements of the economic process and upholding the opportunity cost doctrine as against the real cost doctrine. While acknowledging that the demand and supply curves diagram illuminates the process through which the market clearing price is *discovered,* therefore, Wicksteed insisted that the other diagram brings out far more clearly the fundamental *determinants* of that price, namely, subjective marginal valuations and given supplies. With reference to continuously produced commodities, as opposed to first-edition copies, maintenance of this viewpoint would presumably require that the 'given stocks' referred to should be those of primary inputs. Here it may be noted that, even in the course of denouncing the conventional supply curve, Wicksteed admitted that 'as we recede from the market and deal with long periods . . . cases may arise in which something like a "supply curve" seems legitimate' and that nature does not have 'reserve prices in which she expresses her own demand!' (1914, p. 16, n.1).

See Also

▶ Wicksteed, Philip Henry (1844–1927)

Bibliography

Wicksteed, P.H. 1910. *The common sense of political economy, including a study of the human basis of economic law*. London: Macmillan.

Wicksteed, P.H. 1914. The scope and method of political economy in the light of the 'marginal' theory of value and distribution. *Economic Journal* 24: 1–23.

Residential Real Estate and Finance

Peter Englund

Abstract

Residential real estate is a major asset for most households. This article focuses on three issues relating to housing as an investment. (a) Are returns to housing investment predictable? (b) What is the optimal fraction of real estate in an investment portfolio? (c) How important are borrowing constraints, and how do they influence housing prices? It concludes that housing risks are difficult to hedge in practice and that developing suitable derivative markets would fulfil an important function.

Keywords

Asset pricing; Case–Shiller indexes; Hedonic indexes; Household portfolios; Housing and mortgage markets; Options; Overlapping generations; Portfolio analysis; Residential real estate and finance; Tenure choice

JEL Classifications

R39

Residential real estate is in any definition a major asset class. The average Swedish household invests three-quarters of its net wealth in its own home. Yet it was not until after 1990 that central questions in finance were asked about real estate. Is the market for real estate informationally efficient? What is the optimal fraction of real estate in a household portfolio? What role do financial constraints play in the pricing of real estate? These are particularly challenging questions in view of the special nature of residential real estate assets: properties are heterogeneous, transactions are infrequent, the trading parties are typically amateurs, and the market is best characterized as a search market where identical properties may trade at quite different prices. For all these reasons the data problems are of a different order of magnitude from those in the core areas of finance. Naturally, progress has been slow and we should not expect answers ever to be as sharp as for assets like stocks and bonds.

This article is organized around the questions posed above. Other areas, in particular the important field of mortgages and mortgage-backed securities, are not discussed.

Market Efficiency

Standard theories of portfolio choice and asset pricing presume that markets are informationally efficient in the sense that it is impossible to make profits from trading strategies based on publicly available information, such as past returns. There is ample evidence indicating that real estate markets are not efficient in this sense. Time series studies of real estate returns typically find a strong pattern of positive autocorrelation on quarterly or yearly data (Case and Shiller 1989; Englund and Ioannides 1996). Such a pattern could in principle reflect time-varying risk premia, but this interpretation appears implausible. A problem with most studies of housing returns is that they measure only the time variation in the capital-gains part of returns and ignore the value of housing services (the implicit rent). An exception is Meese and Wallace (1994), which is based on micro evidence on unregulated rents. They confirm, for the San Francisco Bay Area, that returns on owner-occupied homes are indeed predictable based on past returns, but they also show that the profits involved are not sufficient to cover realistic transaction costs for a round-trip trade. There is no money to be made by shifting between renting and owning, with housing consumption fixed, but it may be profitable to time moves according to predicted returns. A general conclusion is that transaction costs in a broad sense are important in understanding real estate markets.

Portfolio Choice

Research on portfolio choice has been hampered by a lack of reliable high-frequency data. Much recent research has been stimulated by the repeat-sales indexes for US metropolitan areas developed and analysed by Case and Shiller (1989). Goetzmann (1993) uses the Case–Shiller indexes to compute optimal portfolios (efficient frontiers) in mean-variance space, taking into account the idiosyncratic component of housing return, that is, the added risk of an individual home above the general return risk captured by a price index. He finds optimum housing shares to be on the order of 10–50 per cent of household net wealth depending on risk attitudes. It is well known from portfolio analysis of other assets that calculated portfolio shares are quite sensitive to input data, particularly expected return, and hence should be treated with caution. Nevertheless, later studies using data for European countries (using hedonic indexes not available in the United States) have obtained similar results (see, for example, Englund et al. 2002, for Stockholm; le Blanc and Lagarenne 2004, for Paris; and Iacoviello and Ortalo-Magné 2004, for London). The discrepancy between computed optimal portfolio shares and real world numbers, often in the order of several hundred per cent, is striking and has provided a challenge for further research.

The standard mean-variance analysis is obviously oversimplified in several ways. First, it is static. Grossman and Laroque (1990) consider lifetime portfolio choice when utility is derived from a durable good (housing), which can only be traded at a cost (proportional to house value). Housing trades are determined in analogy with Ss-models from inventory theory, and optimal portfolios are shown to be mean-variance efficient like in the static case.

Second, the standard analysis does not account for housing services as a consumption good separate from non-durable goods. Flavin and Yamashita (2002) analyse a two-good version of the Grossman and Laroque model with a stochastic relative price of housing. Based on correlations calculated from Case–Shiller indexes, their model indicates that the optimal fraction of financial assets going into stocks is inversely related to the fraction invested in housing, and hence should increase with age, consistent with empirical observations. More recently some authors have analysed models with finite lifetimes, using numerical solution techniques. A key factor in determining the attractiveness of investing in housing is the correlation between labour income and the returns to housing: the stronger the correlation, the smaller is the optimal housing portfolio share.

Third, we have so far assumed housing to be consumed by owning, disregarding the alternative of renting. The issue of tenure choice is a classic one in the housing literature; see Henderson and Ioannides (1983) for a two-period model that brings out some of the basic features. Only rarely have issues of risk been included in the analysis. Among the exceptions are Rosen et al. (1984) and Turner (2003), who find that volatile house prices deter young households from entering into owner-occupancy. These studies do not explicitly measure the relative risks of owning compared with renting. More recently, Sinai and Souleles (2005) have emphasized that owning one's home is a way of hedging the risk associated with stochastic variations in the cost of renting. This is a particularly important aspect for households with a long expected stay in the same dwelling or the same housing market. Empirically, Sinai and Souleles confirm, for US households, that the probability of homeownership is indeed an increasing function of rent risk.

For most households, net wealth falls far short of the value of the house they demand for consumption purposes. Hence, any portfolio study that includes housing has to take a stand on the availability of borrowing. In fact, most households are constrained in their access to borrowing, at least when young, and financial constraints exert an important influence on savings and housing choices over the life cycle; see King (1980) for an early study emphasizing borrowing constraints. Integrating down-payment constraints into models of dynamic portfolio and tenure choice remains an important topic for future study.

Asset Pricing and Financing Constraints

The standard approach to real estate price determination (as in Poterba 1984) is explicitly couched in asset pricing terms: the price is the discounted value of the housing services generated by the property net of operating and maintenance costs. In principle, housing services could be valued based on market rents for comparable dwellings. In applying this approach, lip service is often paid to risk-adjusting the discount rate. It is fair to say, however, that there is no established theory or pragmatic consensus on the choice of discount rate. In recent years there has been a surge of interest in integrating housing into the standard asset-pricing paradigm. So far, however, interest has focused on the impact on financial asset prices of introducing housing collateral rather than on pricing real estate assets.

More attention has been paid to the direct impact of financial constraints on pricing. If the representative homebuyer is constrained by borrowing opportunities rather than by lifetime resources, then wealth shocks have a direct impact on housing demand. This implies that a shock to the demand and supply of housing services will be reinforced through its impact on financing constraints. An income shock, for example, will increase demand and housing prices, thereby releasing borrowing constraints. This will in turn give an extra boost to demand and prices. There will be a 'financial multiplier': the more important financial constraints are, the more sensitive prices will be to shocks to underlying fundamentals. This view of real estate pricing was formulated by Stein (1995) and has been inserted into an overlapping-generations framework with demographic fluctuations by Ortalo-Magné and Rady (2006). Its empirical validity has been investigated in some studies. As an example, Lamont and Stein (1999) show that variations in the sensitivity of house prices to income shocks across US states can be explained by differences in loan-to- value ratios. Financial constraints may also explain the strong impact of variations in house prices on consumption observed in many studies; see, for example, Case et al. (2005) for the United States and internationally.

Historically, mortgage lending has been further restricted by regulations in virtually all countries. Dismantling these regulations has in many cases caused price booms. But borrowing constraints remain important facts of life even in unregulated market environments, and there are large differences across countries even today, reflecting history and legal institutions. Chiuri and Jappelli (2003) show that average downpayment ratios vary from close to 50 per cent in Italy to a little above 10 per cent in Sweden and United Kingdom. They find that these differences, which they largely ascribe to legal tradition – relating to foreclosure, for example – explain differences in homeownership rates across countries, in particular the age when young households enter into owner occupancy.

The Future

Not only has the area of real estate economics been lagging in its adoption of new analytical frameworks from finance, markets have also been slow in adopting new financial instruments and contracts to handle better the important risks many household confront in relation to their housing investment. While households have access to a wide variety of mortgage instruments, markets remain seriously incomplete and fail to offer flexible and liquid contracts related to housing price risks. As Robert Shiller (2003) has forcefully argued, this is one of the macro risks in society that remain uninsurable despite their fundamental importance for individual welfare. Options or futures on relevant housing price indexes could go a long way towards providing such insurance. It remains to be seen how long it will take to develop liquid markets in such instruments.

See Also

- ▶ Capital Asset Pricing Model
- ▶ Efficient Markets Hypothesis
- ▶ Household Portfolios
- ▶ Housing Supply

Bibliography

Case, K., and R. Shiller. 1989. The efficiency of the market for single-family homes. *American Economic Review* 79: 125–137.

Case, K., J. Quigley, and R. Shiller. 2005. Comparing wealth effects: The stock market versus the housing market. *Advances in Macroeconomics* 5: 1–32.

Chiuri, M., and T. Jappelli. 2003. Financial market imperfections and homeownership: A comparative study. *European Economic Review* 47: 857–875.

Englund, P., and Y. Ioannides. 1996. House price dynamics: An international empirical perspective. *Journal of Housing Economics* 6: 119–136.

Englund, P., M. Hwang, and J. Quigley. 2002. Hedging housing risk. *Journal of Real Estate Finance and Economics* 24: 167–200.

Flavin, M., and T. Yamashita. 2002. Owner-occupied housing and the composition of the household portfolio. *American Economic Review* 92: 345–362.

Goetzmann, W. 1993. The single family home in the investment portfolio. *Journal of Real Estate Finance and Economics* 6: 201–222.

Grossman, S., and G. Laroque. 1990. Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica* 58: 25–51.

Henderson, J., and Y. Ioannides. 1983. A model of housing tenure choice. *American Economic Review* 73: 98–113.

Iacoviello, M., and F. Ortalo-Magné. 2004. Hedging housing risk in London. *Journal of Real Estate Finance and Economics* 27: 191–209.

King, M. 1980. An econometric model of tenure choice and demand for housing as a joint decision. *Journal of Public Economics* 53: 137–159.

Lamont, O., and J. Stein. 1999. Leverage and house-price dynamics in U.S. cities. *RAND Journal of Economics* 30: 498–514.

le Blanc, D., and C. Lagarenne. 2004. Owner-Occupied housing and the composition of the household portfolio: The case of France. *Journal of Real Estate Finance and Economics* 29: 259–275.

Meese, R., and N. Wallace. 1994. Testing the present value relation for housing prices: Should I leave my house in San Francisco. *Journal of Urban Economics* 35: 245–266.

Ortalo-Magné, F., and S. Rady. 2006. Housing market dynamics: On the contribution of income shocks and credit constraints. *Review of Economic Studies* 73: 459–485.
Poterba, J. 1984. Tax subsidies to owner-occupied housing: An asset-market approach. *Quarterly Journal of Economics* 99: 729–752.
Rosen, H., K. Rosen, and D. Holtz-Eakin. 1984. Housing tenure, uncertainty, and taxation. *Review of Economics and Statistics* 66: 405–416.
Shiller, R. 2003. *The new financial order: Risk in the 21st century*. Princeton: Princeton University Press.
Sinai, T., and N. Souleles. 2005. Owner-occupied housing as a hedge against rent risk. *Quarterly Journal of Economics* 120: 763–789.
Stein, J. 1995. Prices and trading volume in the housing market: A model with downpayment constraints. *Quarterly Journal of Economics* 110: 379–406.
Turner, T. 2003. Does investment risk affect the housing decisions of families? *Economic Inquiry* 41: 675–691.

Residential Segregation

Jacob L. Vigdor

Abstract

Housing market equilibria display residential segregation when there are systematic disparities in the physical location of households belonging to different racial, ethnic, socio-economic, or other social groups. Historically, segregation has often been enforced through non-market processes such as legal restrictions. Modern segregation, by contrast, is largely driven by cross-group differences in willingness to pay for housing in group enclaves. Segregation often generates social concern, particularly when the segregated group is of low socio-economic status. Empirical studies, including a few based on randomized mobility experiments, suggest that there are negative consequences of growing up in an enclave neighbourhood.

Keywords

Census data; Dissimilarity index; Ethnic identity; Ghettoes; Housing markets; Immigration; Inequality; Internal migration; Racial segregation; Residential integration; Residential segregation; Socio-economic segregation; Spatial mismatch hypothesis; Spectral segregation index; Tipping; Zoning

JEL Classifications

J15; R21; R29

The term 'residential segregation' describes a housing market equilibrium marked by systematic disparities in the physical location of households belonging to different racial, ethnic, socio-economic, or other social groups.

While history is replete with examples of groups forced to live in complete isolation from the remainder of society, residential segregation is not inherently a dichotomous phenomenon. Rather, housing markets may exhibit varying degrees of segregation; social scientists have endeavoured to quantify this variation for the better part of a century. The term 'ghetto' is often ascribed to social groups experiencing segregation that exceeds a loosely defined threshold.

Residential segregation may be the outcome of a past residential sorting process wherein centralized authorities restricted some agents' location choices. Very simple economic theory, and an increasing amount of empirical evidence, however, point to the conclusion that modern-day residential segregation is driven primarily by the operation of decentralized market forces.

Even if residential segregation is a pure market phenomenon, many observers harbour concerns that segregated housing market equilibria are sub-optimal from a social welfare perspective. Some debate exists as to whether segregated housing markets are inefficient. Arguments hinge on whether households are fully informed at the time they make location decisions, or whether they face borrowing constraints. It is a less controversial observation that residential segregation has important implications for distributional equity. In segregated equilibria, for example, wealthy households have the opportunity to avoid subsidizing the local public good consumption of poorer households. Over the past several

R

decades, there have been many attempts to estimate the relationship between residential segregation and inequality between groups, in both the short term and the long term.

Here, basic evidence is provided on the existence and magnitude of contemporary residential segregation. This evidence draws heavily on the experience of racial and immigrant groups in the United States; the measurement of segregation in other nations is limited in scope and often confined to very recent observations. As discussed below, this is more a reflection of data limitations than any genuine lack of interest. The basic economic theory of why segregation exists is then outlined, and empirical evidence that has been brought to bear on the issue discussed. The concluding discussion considers the potential implications of segregation on socio-economic outcomes and human capital investment.

Measuring Segregation

There are many ways to measure segregation (Massey and Denton 1988). The metrics most commonly used in sociology and economics require the existence of neighbourhood-level data on the distribution of groups in a city or region. Some measures, including the spectral segregation index (Echenique and Fryer 2005), require additional data on the physical location of these neighbourhoods and, in some cases, their land area. A central challenge to the systematic measurement of segregation is the lack of comparable neighbourhood-level data across nations, or even cities within nations, and over time. The United States, for example, has collected data on the race of its inhabitants since 1790, but did not report race at a consistently defined neighbourhood level until 1940. The United Kingdom did not systematically collect information on the ethnic identity of its inhabitants until the 1991 Census.

Given the existence of required data, the most commonly used segregation indices classify the residential separation of any particular group between two extremes: perfect segregation, where group members never share a neighbourhood with individuals not belonging to the group, and perfect integration, where group members form an equal share of the population in all neighbourhoods. The dissimilarity index (Duncan and Duncan 1955), records groups on the scale from 0 (perfectly integrated) to 1 (perfectly segregated) using the following formula:

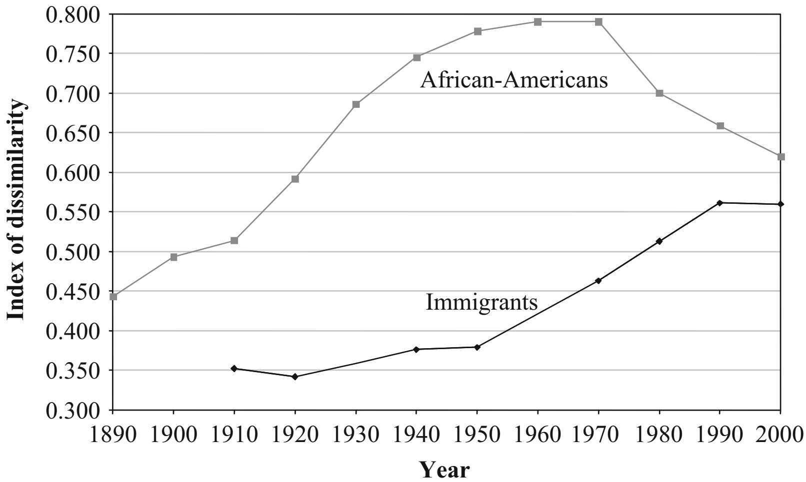
$$D = \frac{1}{2} \sum_i \left| \frac{A_i}{A} - \frac{B_i}{B} \right| \tag{1}$$

where i indexes neighborhoods, A_i and B_i represent the number of group members and others in neighborhood i, respectively, and A and B represent the total population of group members and others in the city or region. The dissimilarity index has a relatively intuitive interpretation: it is the share of group members, or others, who would have to switch neighbourhoods in order to achieve perfect integration. While many demographers state a preference for other indices based on various criteria, the dissimilarity index is most commonly used in existing literature.

Stylized Facts

The absence of neighbourhood-level data makes it difficult to gauge the contemporary level of segregation in many cities, let alone historical levels. The most comprehensive historical data pertain to American cities. Cutler et al. (1999, 2005) use these data to compute long-term trends in dissimilarity for African-Americans and foreign-born individuals, respectively. Figure 1 plots weighted averages of these measures across cities, with weights equal to the population of the group in question in each city. Immigrant segregation is computed separately for each country-of-origin group in each city; the immigrant time series represents the weighted average of these data.

As the relatively low initial levels of black–white dissimilarity indicate, urban ghetto neighbourhoods were relatively uncommon in the United States at the turn of the 20th century. The birth of the African-American ghetto coincided with the so- called Great Migration of blacks from the southern part of the United States



Residential Segregation, Fig. 1 Dissimilarity of African-Americans and immigrants in the United States, 1890–2000

to northern cities between the First World War and 1965. Segregation reached its peak at the end of this period of migration. In 1970 dissimilarity levels in some areas, chiefly large industrial cities of the north-eastern and mid-western United States, were at or near 0.90. Since that time, segregation has fallen pervasively throughout the nation but most acutely in rapidly growing cities in the southern and western parts of the country.

Immigrant segregation, quite strikingly, displays the opposite trend to racial segregation in the United States. Immigrant dissimilarity remained stable at relatively low levels through the first half of the century, then rose steadily. Thus, even as racial ghettos have declined over the past few decades, immigrant segregation has risen. Cutler et al. (2005) present further data indicating that the rise in average segregation can be attributed primarily to the growth of groups that have always experienced high segregation, rather than to the increasing segregation of individual groups. The growing, highly segregated groups generally originate in less developed countries and gravitate toward the largest cities in the United States. The limited amount of data available from other nations supports the general trend found in the United States: individual racial and ethnic group are experiencing stable or declining segregation in most parts of the world.

Socio-economic segregation, or the degree to which households in poverty tend to cluster together in neighbourhoods, increased in the United States in the 1970s and 1980s, but showed some evidence of lessening in the 1990s.

Table 1 presents some representative dissimilarity index values for groups in some of the world's largest cities, using recently available data.

Why Are Groups Segregated?

Theories of racial segregation can be classified into two types. The first type permits some form of discrimination in housing markets. The second models segregation as the equilibrium outcome of a fully competitive market. A potential third class of models explains one form of segregation as the direct consequence of a second form – for example, it explains racial segregation as a consequence of economic segregation. This third class is of less interest to attempts to explain segregation more generally.

In a discrimination-based model, location choices are constrained for members of one group, defined by race, ethnicity, or other observable characteristic. The constraints on location choice might include explicit legal barriers or implicit patterns of 'steering' households towards certain locations. Historical examples of explicit

Residential Segregation, Table 1 Recent dissimilarity indices for various groups in major world cities

City	Group	Year	Dissimilarity
Barcelona	Latin American immigrants	2001	0.290
Cape Town	Blacks[a]	1996	0.928
Chicago	African-Americans	2000	0.778
Cologne	Turkish immigrants[b]	1994	0.337
Lima	High SES households	1993	0.440
London	Blacks (Caribbean, African and Other)[a]	2001	0.468
London	South Asians[a]	2001	0.544
Los Angeles	Mexican immigrants	2000	0.446
Mexico City	High SES households	2000	0.380
New York	African-Americans	2000	0.670
Santiago	High SES households	1992	0.490
Tokyo	Individuals over 65	1995	0.147

Note: Dissimilarity indices measure the separation of each group from the remainder of the population, except as indicated
[a]Dissimilarity from whites
[b]Dissimilarity from Germans
SES socio-economic status
Sources: Barcelona: Martori i Cañas and Hoberg (2004); Cape Town: Rospabe and Selod (2003); Chicago and New York: Glaeser and Vigdor (2002); Cologne: Friedrichs (1998); Lima, Mexico City and Santiago: Arriagada Luco and Vignoli (2003); London: Burgess et al. (2005); Los Angeles: Cutler et al. (2005); Tokyo: Nakagawa (2003)

legal barriers abound. In a few cases, governments have attempted to restrict location choices as a matter of public law, have enforced contracts between private parties restricting racial ownership or occupancy of property, or have adopted policies that had the effect of limiting the residential options of certain groups. In the United States, federal legislation had made most explicit forms of housing market discrimination illegal by the end of the 1960s.

While few observers would argue that explicit racial or ethnic barriers to location choice persist in the developed world, the existence and prevalence of implicit discriminatory patterns is a subject of continuing debate. Government policies such as zoning laws, which local governments use to regulate the density and nature of residential development within their borders, may implicitly perpetuate segregation. Housing audit studies provide evidence of discriminatory behaviour among real estate agents, mortgage brokers, or landlords. In these studies, auditors of different races present carefully matched, fictionalized credentials to housing market agents. The behaviour of these agents is then analysed to uncover any systematic differences in treatment by race. Recent studies, such as Ondrich et al. (2003), find evidence of significant racial disparities in treatment in the United States.

While disparities in treatment of housing market auditors can be interpreted as evidence of continued racism, such behaviour can also be consistent with unbiased, profit-maximizing motives. As in models of statistical discrimination in labour markets or other settings, agent behaviour could be motivated by accurate perceptions of differences in average preferences across racial groups.

Such an interpretation is consistent with the second type of racial segregation theory, which posits that segregated housing market equilibria are fully consistent with decentralized, unconstrained household choices. Preference-based theories of segregation owe some intellectual debt to Tiebout's (1956) vision of residential sorting, but evolve most clearly from Schelling's (1978) simulation of residential sorting in the presence of very slight preferences for neighbours of one's own group. Schelling's simulations show that a small initial concentration of same-group neighbours can rapidly evolve into a vast enclave community. This process of 'tipping' is driven by group members' heightened willingness to pay for locations in close proximity to the initial cluster. As the enclave grows in size, it becomes disproportionately more attractive to group members than to others. So long as the segregated group in question maintains a steady population share in the entire market, the enclave is very unlikely to dissipate.

Much anecdotal evidence supports the Schelling model. The neighbourhood integration that has taken place in the United States since 1970, for example, has left most African-American enclaves untouched. Rather, integration has occurred either in newly developed neighbourhoods on the fringe of urban areas or in locations marked by significant demolition and redevelopment.

While intuitively appealing and supported by anecdotal observation, true empirical tests of preference-based theories are rendered difficult by the unobservability of household preferences. Econometric models associated with the measurement of willingness to pay, such as discrete choice models, often assume away the existence of housing market discrimination (for example, Bayer et al. 2004). Survey-based methods of eliciting preferences are valid only to the extent that respondents can accurately separate their valuation of neighbourhood racial composition from all other attributes, and truthfully reveal this valuation. What survey evidence that exists supports the notion that groups harbour preferences for same-group neighbours (Vigdor 2003).

Why might individuals care about the racial or ethnic composition of their neighbourhoods? Group members may prefer to congregate in enclaves in order to take advantage of scale economies enabling the supply of group-specific community institutions or consumer goods. Individuals may also seek to limit exposure to other groups on the basis of stereotyped perceptions of inferiority, greater criminality, or other characteristics. It is also possible that individuals care, not about the race of their neighbors directly, but about characteristics correlated with race, such as socio-economic status. These varying hypotheses have dramatically different implications for the social value of segregation. Unfortunately, these various explanations are observationally equivalent. Each predicts that segregation occurs in equilibrium because willingness to pay for housing in a group enclave is relatively higher among group members.

While there is currently no consensus on the importance of housing market discrimination in perpetuating segregation, Cutler et al. (1999) present evidence that any such importance has declined. In 1940, at a time when many forms of housing market discrimination were legal – and in some cases practised by government itself – restrictions on African-American location choice had the impact of increasing equilibrium prices in segregated areas. By 1970 that premium had disappeared, suggesting that these artificial barriers to mobility had been removed.

Does Segregation Influence Economic Outcomes?

A number of hypothesized causal mechanisms link segregation to socio-economic outcomes. The ‘spatial mismatch’ hypothesis contends that segregation reduces the average income of certain groups to the extent that their residential enclaves are located at some distance from growing employment centres (Kain 1968). Segregation may also lead to differences in education quality across racial or ethnic groups, to the extent that schooling is tied to residential location. Finally, there may be other localized factors that differ across neighbourhoods and have the net impact of leading to different human capital investment trajectories. For example, children growing up in different neighbourhoods may develop different consumption or investment preferences by being exposed to different types of role models.

Numerous attempts have been made to empirically estimate the impact of segregation on outcomes, whether operating immediately through spatial mismatch- type mechanisms or developmentally. Much of this empirical literature is plagued by a fundamental endogeneity problem: since individuals choose their own neighbourhoods, any correlation between neighbourhood characteristics and individual outcomes might reflect selection rather than any causal effect of the former on the latter. Researchers have implemented three strategies for circumventing these selection problems. The first is to focus on the outcomes of young adults, whose location choices are presumably determined by their parents rather than themselves. Vigdor (2002) points out that the strategy of studying young adults is suspect in the presence of intergenerational transmission of economic outcomes.

A second basic strategy for identifying the impact of segregation on outcomes in the presence of selective migration is to model location choice and socio-economic outcomes simultaneously. Some research in this vein makes use of individual data-sets with detailed geographic identification, recently made available by the US Census Bureau. A simultaneous equation model can uncover the true causal impact of segregation on outcomes if it employs an instrumental variable – a factor than affects location choice but otherwise bears no correlation to individual outcomes. In practice, identifying a valid instrumental variable is very difficult.

Recently, researchers have addressed selective migration concerns by turning their attention to randomized mobility experiments, in which a 'treatment' group is offered a voucher redeemable for housing only in certain neighbourhoods, while a 'control' group is offered no such aid. While these experiments generally do not permit examination of the causal impact of segregation per se, they do allow a more general study of the potential importance of neighbourhood characteristics in determining outcomes. In general, studies find little impact of neighbourhood factors on the socio-economic outcomes of adults. There is more evidence in favour of developmental impacts on youth. Orr et al. (2003) present an overview of research results stemming from one such randomized mobility experiment, the Moving to Opportunity demonstration programme.

See Also

▶ Finance
▶ Ghettoes
▶ Housing Policy in the United States
▶ Immigration and the City
▶ Neighbours and Neighbourhoods
▶ Spatial Mismatch Hypothesis
▶ Symmetry Breaking
▶ Urban Economics
▶ Urban Housing Demand

Bibliography

Arriagada Luco, C., and J. Vignoli. 2003. Segregatión residencial en áreas metropolitanas de América Latina: magnitud, caracteristicas, evolutión e implicaciones de politica. Series Población y Desarrollo No. 47. UN Comisión Económica para América Latina y El Caribe, Santiago de Chile.

Bayer, P., R. McMillan, and K. Rueben. 2004. Residential segregation in general equilibrium. Working Paper no. 11095. NBER, Cambridge. MA.

Burgess, S., D. Wilson, and R. Lupton. 2005. Parallel lives? Ethnic segregation across schools and neighborhoods. *Urban Studies* 42: 1027–1056.

Cutler, D., E. Glaeser, and J. Vigdor. 1999. The rise and decline of the American ghetto. *Journal of Political Economy* 107: 455–506.

Cutler, D., E. Glaeser, and J. Vigdor. 2005. Is the melting pot still hot? Explaining the resurgence of immigrant segregation. Working Paper No. 11295. NBER, Cambridge, MA.

Duncan, O., and B. Duncan. 1955. A methodological analysis of segregation indexes. *American Sociological Review* 20: 210–217.

Echenique, F., and R.G. Fryer, Jr. 2005. On the measurement of segregation. Working Paper No. 11258. NBER, Cambridge, MA.

Friedrichs, J. 1998. Ethnic segregation in Cologne, Germany, 1984–1994. *Urban Studies* 35: 1745–1763.

Glaeser, E., and J. Vigdor. 2002. Residential segregation: Promising news. In *Redefining urban & suburban America: Evidence from census 2000*, ed. B. Katz and R. Lang, Vol. 1. Washington, DC: Brookings Institution Press.

Kain, J. 1968. Housing segregation, negro employment, and metropolitan decentralization. *Quarterly Journal of Economics* 82: 175–197.

Martori i Cañas, J., and K. Hoberg. 2004. Indicadores cuantativos de segregatión residencial. El caso de la población inmigrante en Barcelona. *Scripta Nova* 8(169). Online. Available at http://www.ub.es/geocrit/sn/sn-169.htm. Accessed 1 Aug 2005.

Massey, D., and N. Denton. 1988. The dimensions of racial segregation. *Social Forces* 67: 281–315.

Nakagawa, M. 2003. Analysis of age discrimination in the rental housing market in Japan: An approach using a fair housing audit. Discussion Paper No. 577. Institute of Social and Economic Research, Osaka University, Osaka.

Ondrich, J., S. Ross, and J. Yinger. 2003. Now you see it, now you don't: Why do real estate agents withhold available houses from black customers? *Review of Economics and Statistics* 85: 854–873.

Orr, L., J. Feins, R. Jacob, E. Beecroft, L. Sanbonmatsu, L. Katz, J. Liebman, and J. Kling. 2003. *Moving to opportunity: Interim impacts evaluation*. Washington, DC: Office of Policy Development and Research, US Department of Housing and Urban Development.

Rospabe, S., and H. Selod. 2003. Does city structure cause unemployment? The case study of Cape Town. Unpublished manuscript. Laboratoire d'Economie Appliquée, Paris.

Schelling, T. 1978. *Micromotives and macrobehavior*. New York: W.W. Norton.

Tiebout, C. 1956. A pure theory of local expenditures. *Journal of Political Economy* 64: 416–424.

Vigdor, J. 2002. Locations, outcomes, and selective migration. *Review of Economics and Statistics* 84: 751–755.

Vigdor, J. 2003. Residential segregation and preference misalignment. *Journal of Urban Economics* 54: 587–609.

Residuals

F. J. Anscombe

Most commonly used statistical procedures, for analysis and interpretation of statistical data, rest on assumptions about the behaviour of the data. Quite often these assumptions can be adequately justified, and the procedures accepted as fair and reasonable. But that is not always so, and it behoves the analyst to check consistency of the data with the assumptions. Failure to do this may lead to a grossly misleading analysis and the drawing of wrong conclusions. Just how consistency can be checked depends on the complexity of the data. Often a step is calculation of *residuals,* which are measures of deviation between the observed values of a variable and the fitted (or estimated or predicted) values for that variable, calculated in accordance with the assumptions. The residuals, when found, are sometimes combined into a summary measure of goodness of fit, or sometimes they are displayed graphically, in various possible ways.

A very simple example of this kind of concern is afforded by the common practice of summarizing a single set of readings of a quantitative variable by the average of the readings and their standard deviation. Those two quantities would certainly form a good and convenient summary of the data, useful for a variety of purposes, if we knew that the readings were independent observations of a random variable following a normal (Gauss–Laplace) distribution, or something not very different from that. Usually in practice we do not have such knowledge. We can, however, check to see whether the distribution of the readings, shown perhaps by a histogram, is reasonably consistent with a normal distribution; and if the readings came to us arranged in some meaningful order we could look for evidence of serial dependence. If the readings contained one extreme outlier (a reading very far from all the others), the average and standard deviation calculated from all the readings could be quite different from those calculated from all the readings *except* that one outlier, and for most purposes the average and standard deviation of all the readings would be misleading.

The possibly devastating effect of outliers has suggested to some authorities that 'robust' measures of the centre and spread of a set of readings would be preferable to the traditional average and standard deviation – measures that rest on much weaker assumptions than a nearly normal distribution, measures that would be little affected by inclusion or exclusion of a few outliers if such occurred in the data. Instead of the average one could choose the median of the data, and instead of the standard deviation one could choose the median absolute deviation of readings from their median, or the interquartile range. Such robust measures cannot be said to rest on no assumptions at all – independence is assumed, for example – but they are safer to use if procedures must be used uncritically. A price is paid for the safety. The traditional assumptions permit a considerable body of simple inferential methods, that must be foregone or much modified when only the weaker assumptions for robust procedures are made. Thus in much statistical practice today, analytical procedures based on specific non-robust probabilistic assumptions are still often used, but checking conformity of the data with the assumptions is regarded as essential.

In the above discussion of summarizing a single set of readings of one variable, the word 'residual' has not been mentioned. Residuals could be defined as the differences between each of the readings and their average (or median or whatever central measure is adopted). The central measure is the 'fitted value', the same for all readings. An outlier is a reading whose residual is much larger in magnitude than nearly all the other residuals. In any graphical presentation of the data, the differences between individual readings and the common central value are easily seen, whether or not the residuals have been calculated; and therefore in this context it is hardly necessary to refer to residuals, even though it is just those differences that are of most interest. The most widely used technique in the analysis of statistical data is linear regression, by which the association of a quantitative 'dependent' variable with one or more explanatory variables may be studied. Some of the considerations that arise concerning consistency of the data with the assumptions underlying linear regression can be seen in their simplest form if we consider linear regression of one dependent variable on just one explanatory variable.

For such simple linear regression, the standard least-squares calculation is based on the following theoretical description or 'model': the given number pairs (x_i, y_i) are related by

$$y_i = \beta_0 + \beta_1 x_i + \in_i \quad (i = 1, 2, \ldots, n), \qquad (1)$$

where β_0 and β_0 are constants and the 'errors' $\{\in_i\}$ are drawn independently from a normal probability distribution having zero mean and constant variance. The regression calculation leads to estimates b_0 and b_1 for β_0 and β_1 to the fitted values

$$\widehat{y}_i = b_0 + b_1 x_i = \overline{y} + b_1(x_i - \overline{x}),$$

and to the residuals

$$e_i = y_i - \widehat{y}_i.$$

The sum of squares of the latter, generally called the 'residual sum of squares', leads to an estimate of the variance of the distribution of errors. If the theoretical description were exactly correct (and all calculations were exact, without round-off error), these calculations would be entirely satisfactory, in the sense that b_0, b_1 and the residual sum of squares, together with the number of readings n and the first two moments of the x-values, would constitute sufficient statistics for the original data for all purposes with no loss of information. In practice, we do not know that the theoretical description is correct, we should generally suspect that it is not, and we cannot therefore heave a sigh of relief when the regression calculation has been made, knowing that statistical justice has been done.

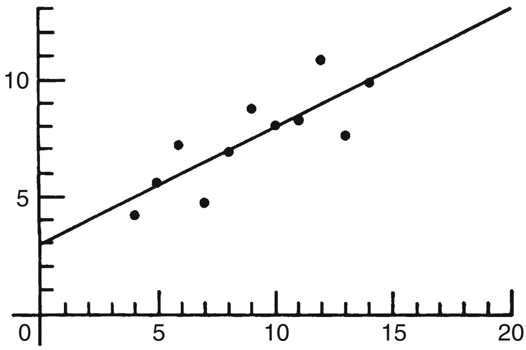
Some of the possibilities for appropriateness or inappropriateness of the standard regression calculation are illustrated by the four artificial data sets given in Table 1. Each data set consists of

Residuals, Table 1 Four artificial data sets, each consisting of eleven (x, y) pairs

Data set		1–3	1	2	3	4	4
Variable		x	y	y	y	x	y
Obs. no.	1:	10.0	8.04	9.14	7.46:	8.0	6.58
	2:	8.0	6.95	8.14	6.77:	8.0	5.76
	3:	13.0	7.58	8.74	12.74:	8.0	7.71
	4:	9.0	8.81	8.77	7.11:	8.0	8.84
	5:	11.0	8.33	9.26	7.81:	8.0	8.47
	6:	14.0	9.96	8.10	8.84:	8.0	7.04
	7:	6.0	7.24	6.13	6.08:	8.0	5.25
	8:	4.0	4.26	3.10	5.39:	19.0	12.50
	9:	12.0	10.84	9.13	8.15:	8.0	5.56
	10:	7.0	4.82	7.26	6.42:	8.0	7.91
	11:	5.0	5.68	4.74	5.73:	8.0	6.89

Residuals, Table 2 The same standard output of a regression analysis of each of the data sets in Table 1

Number of observations (n) = 11
Mean of the x's ($\bar{x}$) = 9.0
Mean of the y's ($\bar{y}$) = 7.5
Regression coefficient (b_1) of y on x = 0.5
Equation of regression line: $y = 3 + 0.5x$
Sum of squares of $x - \bar{x}$ = 110.0
Regression sum of squares = 27.50 (1 d.f.)
Residual sum of squares of y = 13.75 (9 d.f.)
Estimated standard error of b_1 = 0.118
Multiple R^2 = 0.667

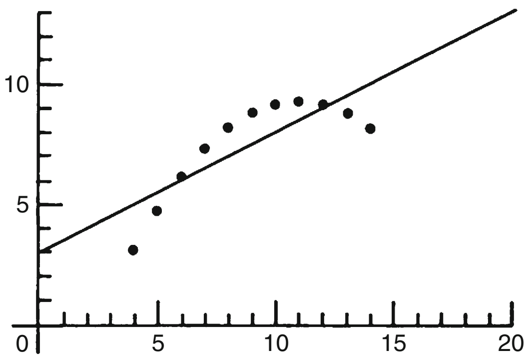


Residuals, Fig. 1

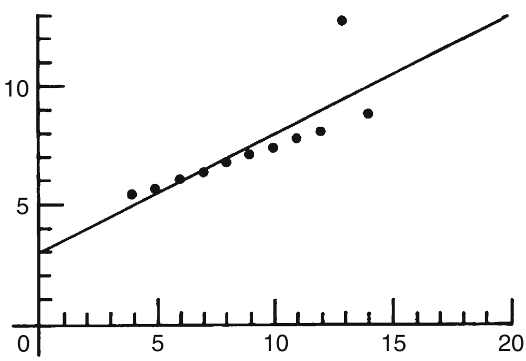
eleven (x, y) pairs. For the first three data sets the x-values are the same, and they are listed only once. The four data sets have been constructed so as to yield the same standard output from a typical regression programme, as shown in Table 2. Thus if equation (1) is a correct theoretical description of the data, all four data sets are equivalent – they mean the same thing.

Regression programmes often list the residuals, in the order in which the data were entered. Since in the present case the data have been entered in a random order, probably little would be seen if the eye were run down such a listing, especially if it were in abominable floating-point notation. Only if the residuals are presented graphically, or perhaps combined into one or more overall measures of goodness of fit, is the viewer likely to realize how very different in character these four data sets are, and therefore how inadequate the information in Table 2 is. The simplest kind of graphical presentation of the data sets is just a scatterplot of the given (x, y) pairs, together with the fitted regression line, as in Figs. 1, 2, 3, and 4.

Figure 1, corresponding to data set 1, is the kind of thing most people would see in their mind's eye, if they were presented with the summary in Table 2. The theoretical description (1) seems to be perfectly appropriate here, and the summary fair and adequate. Figure 2 suggests forcefully that data set 2 does not conform with the theoretical description (1), but rather y has a smooth curved relation with x, possibly quadratic, and there is little residual variability.

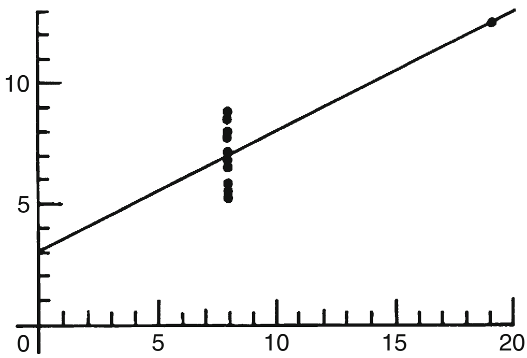


Residuals, Fig. 2



Residuals, Fig. 3

Figure 3 similarly suggests that (1) is not a good description for data set 3: all but one of the observations lie close to a straight line (not the one yielded by the standard regression calculation), namely



Residuals, Fig. 4

$$y = 4 + 0.346x;$$

and one observation is far from this line. Those are the essential facts that need to be understood and reported.

Figure 4, like Fig. 1, shows data apparently conforming well with the theoretical description (1). If all observations are considered genuine and reliable, data set 4 is just as informative about the regression relation as data set 1; there is no reason to prefer either to the other. Yet in most circumstances we should feel that there was something unsatisfactory about data set 4. All the information about the slope of the regression line resides in one observation – if that observation were deleted the slope could not be estimated. Usually we are not quite sure that every observation is reliable. If any one observation were discredited and therefore deleted from data set 1, the remainder would tell much the same story. That is not so for data set 4. Thus the standard regression calculation ought to be accompanied by a warning that one observation has played a critical role. Of course, just one informative observation is much better than none. But we are usually happier about asserting a regression relation if the relation seems to permeate many of the observations and does not inhere mostly in one or two.

Each of the data sets 2, 3, 4 illustrates a peculiar effect in an extreme form. In less extreme forms such effects are often encountered in statistical analysis. There are other kinds of effect that can appear, such as residual variability changing progressively with x. But it is arguable that data sets 2, 3, 4 exemplify the three situations most important to recognize if they should occur, namely that the true regression relationship between the variables is not the linear one fitted, that there are one or more extreme outliers among the residuals, and that there are one or more highly influential x-values.

When regression is done on more than one explanatory variable, similar considerations arise. A simple two-dimensional scatterplot cannot now represent directly the whole of the data, in the style of Figs. 1, 2, 3, and 4 above. It is found that plots of residuals against the fitted values, and also of residuals against the values of each explanatory variable in turn, are often effective in suggesting ways to improve the analysis; and other kinds of plots depending on residuals are sometimes made. The more variables there are, the greater are the possible complexities in the data, and the less sure we can be that all important effects will be perceived. That is the more reason for examining residuals carefully. Various specific test statistics can be formed from residuals and used to detect specific kinds of discrepancy between the data and the assumed theoretical description.

Examination of residuals has been most thoroughly developed for regression. But in many other cases when data are considered in light of a theoretical description, measures of difference between observed values and fitted values can be defined that behave like regression residuals and are similarly useful. Such residuals are, however, often not just simple differences between observed values and fitted values.

[Examination of residuals is discussed by Draper and Smith (1981), Cook and Weisberg (1982), Anscombe (1981), McCullagh and Nelder (1983), Cox and Snell (1968). The tables and figures given above are taken from Anscombe (1973). For robust methods see Tukey (1977) and Huber (1981).]

See Also

- ▶ Outliers
- ▶ Random Variables
- ▶ Regression and Correlation Analysis

Bibliography

Anscombe, F.J. 1973. *American Statistician* 27: 17–21.
Anscombe, F.J. 1981. *Computing in statistical science through APL*. New York: Springer. (Especially Appendix 2).
Cook, R.D., and S. Weisberg. 1982. *Residuals and influence in regression*. London: Chapman & Hall. (Especially Chapter 2).
Cox, D.R., and E.J. Snell. 1968. *Journal of the Royal Statistical Society, Series B* 30: 248–275.
Draper, N.R., and H. Smith. 1981. *Applied regression analysis*, 2nd ed. New York: Wiley.
Huber, P.J. 1981. *Robust statistics*. New York: Wiley.
McCullagh, P., and J.A. Nelder. 1983. *Generalized linear models*. London: Chapman & Hall.
Tukey, J.W. 1977. *Exploratory data analysis*. Reading: Addison-Wesley.

Reswitching of Technique

Roberto Scazzieri

Abstract

Reswitching of technique is the property whereby, when multiple production techniques are available in a wage–profit economy, the same technique may be optimal at different levels of the rate of interest. This means that a virtual movement of the rate of interest in a given direction might make it rational to use techniques that had been previously excluded, so that the rate of interest cannot provide an unambiguous ranking of techniques. This possibility is rooted in the complex interactions (movements of relative prices) that occur in a production economy with different proportions between labour and intermediate inputs.

Keywords

Capital deepening; Capital intensity; Factor-price frontiers; Intermediate products; Recurrence of technique; Reswitching of technique; Technical choice

JEL Classifications

D2

Reswitching of technique refers to the virtual adoption of production techniques, either by the individual producer or by the economic system as a whole. Standard economic theory treats technical adoption on the assumption that there is a multiplicity of techniques for producing any given good, and that the producer, as a rational decision maker, will switch from one technique to another according to a certain hypothetical sequence as the prices of productive factors are changed. This sequence would depend on the ranking of techniques in terms of capital per man or 'capital intensity', so that a lower rate of interest (which is equal to the rate of profit in equilibrium) would be associated with the 'adoption' of a technique characterized by higher capital per man. This process is known as capital deepening.

The development of discrete production models in the 1950s led to the discovery that this view of 'rational' technical adoption is not necessarily well founded. David Champernowne (1953) and Joan Robinson (1956) pointed out that a movement of the rate of interest in a given direction might make it optimal once again to use techniques that had been previously excluded. This phenomenon is known as *reswitching of technique.*

The original discovery was associated with the belief that reswitching was nothing more than a 'curiosum', which could not be left out on grounds of pure logic but was nevertheless unlikely to happen. The discussion of this phenomenon by Piero Sraffa (1960) showed that reswitching is the normal outcome of a situation in which the various production processes are characterized by different proportions between 'direct' labour and the quantity of 'past' labour. (This latter is the quantity of labour that is indirectly required in a production process, being required in producing its intermediate inputs.) Sraffa's analysis also provides a clear insight into the reasons for technical reswitching along the hypothetical sequence associated with changes in the rate of profit. It is worthwhile considering his example in some detail.

R

The 'Pure Products' Case

It is useful to start with the consideration of a special category of commodities, which we might call of the *pure product* type. These are commodities that are never used as productive inputs, so that their price reflects production cost, but cost is never influenced by the variation of their particular prices.

Let a and b be commodities of that type, and let them be produced with different proportions of direct labour to past labour. (This structure of labour requirements is representative of the differences in the proportions between labour and intermediate inputs in the production processes of the two commodities.)

Let a require more labour than b if we consider labour applied eight years before the year in which the product is ready, whereas b requires more labour than a in the cases of labour applied in the current year and 25 years earlier. This situation may be represented as follows (n is the date at which labour is applied):

$$\begin{aligned}
&\text{(i)} \quad n = 8 \\
&\quad l_{a(8)} = v + 20 \\
&\quad l_{b(8)} = v \\
&\text{(ii)} \quad n = 0 \\
&\quad l_{a(0)} = x \\
&\quad l_{b(0)} = x + 19 \\
&\text{(iii)} \quad n = 25 \\
&\quad l_{a(25)} = y \\
&\quad l_{b(25)} = y + 1
\end{aligned}$$

We are now in a position to examine in which way the cost difference between the two products may vary if the rate of profit is raised from 0 to a maximum value of 25 per cent. (An increase of the rate of profit is equivalent to a change in the weight of the different labour terms in each cost equation.)

The cost difference is expressed by the following equation:

$$p_a - p_b = 20w(1+r)^8 - \left[19w + w(1+r)^{25}\right]. \qquad (1)$$

On the assumption that the wage rate (w) is inversely related to the rate of profit according to the following expression:

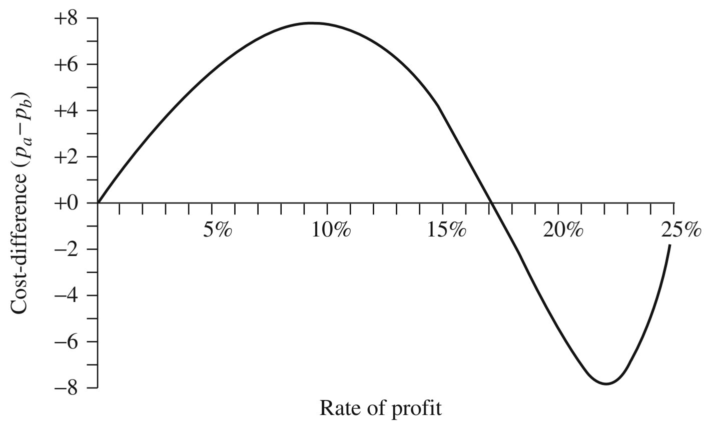
$$w = 1 - \frac{r}{25\%},$$

the cost difference equation will be represented by the curve in Fig. 1.

The cost of a rises relatively to b as r increases between zero and nine per cent. The reason for this is that the change of r leaves the value of current labour unaffected, whereas the 'excess labour' of date 8 is much greater than the excess labour of date 25. The increase in the value of $l_{b(25)}$ is more than offset by the increase in the value of $l_{a(8)}$ and the compound effect of these two variations is an increase in the cost difference. Beyond $r = 9\%$, the increasing weight of remote labour terms brings the cost difference down. This reduction stops at $r = 22\%$, since at this particular level of the rate of profit the decline of the wage rate starts offsetting the increase in the value of remote labour terms due to a higher r.

The above argument has straightforward implications for technical choice in the case of commodities of the pure product type. For in this case we can take for granted that the price of each commodity reflects its cost of production, whereas this price has no influence at all on the cost. Under such conditions, eq. (1) permits us to examine in which way the relative profitability of two techniques is varied as r goes from 0 to r (max). In fact, we may take eq. (1) to illustrate the difference between the unit costs of production of the same commodity produced with two alternative techniques. (For reasons of symmetry with the previous argument we call such alternative techniques a and b respectively.) Figure 1 can be applied to this particular case. An immediate shortcoming would be that a change in the price of direct to 'dated' labour, as reflected in an increasing r, is associated with a positive excess of unit cost p_a over unit cost p_b until the curve intersects the horizontal axis for the first time. This involves that, over this interval, technique b is more profitable than technique a. A further increase of r (until r (max)) is associated with a negative

Reswitching of Technique, Fig. 1



difference ($p_a - p_b$), so that technique a is more profitable than technique b. However, the same figure shows that the *reduction* of the cost difference stops at $r = 22\%$. For any r such that $22\% < r < r(\max)$, the cost difference is increasing once again. This increase stops at $r = 25\%$, when techniques a and b become equally profitable.

The movement of the cost difference when r is increasing shows that the relative profitability of techniques a and b is subject to fluctuations which depend on the particular interval within which r is changed. The relative profitability of technique a with respect to technique b is initially decreasing, then increasing, finally decreasing again. These fluctuations show that the 'unevenness' of the input structure may bring about multiple switches between the two techniques as we consider a steadily increasing r: the same technique might be adopted at low and high rates of profit, with the alternative technique being adopted at intermediate levels of r.

The 'Intermediate Products' Case

It might appear that the above picture gets greatly complicated when we consider the more general case of products that are used as productive inputs either of themselves or of other commodities. For in this new situation the price of a commodity reflects its production cost, but this cost might itself be influenced by that price. (Directly in the case of a product used in its own production, indirectly in the case of a product that is, at some stage, a necessary means of production for at least one of its inputs.)

An immediate consequence of the consideration of interdependence between production processes is that inspection of the cost difference equation is no longer sufficient in order to assess the relative profitability of alternative techniques. The mutual influence between prices and production costs brings about the need of comparing systems of interrelated techniques (production technologies) rather than individual techniques. This requires consideration of the price system that will be associated with each technology at any given distribution of income between wages and profits.

The analysis of the 'intermediate products' case can be carried out by examining a simple model with two alternative two-good technologies A and B, in which all products are used as inputs of themselves and of the other commodity. We shall also assume that the two technologies differ only in the technique used to produce commodity 1.

The two price systems may be written as follows:

$$(a_{11}p_1 + a_{21}p_2)(1 + r) + l_1(a)w = p_1$$

$$(a_{12}p_1 + a_{22}p_2)(1 + r) + l_2(a)w = p_2, \qquad (2)$$

R

$$(b_{11}p_1 + b_{21}p_2)(1+r) + l_1(\mathrm{b})w = p_1$$

$$(b_{12}p_1 + b_{22}p_2)(1+r) + l_2(\mathrm{b})w = p_2, \qquad (3)$$

where a_{ij} $(i, j = 1, 2)$ and b_{ij} $(i, j = 1, 2)$ are the quantities of commodity i required to produce one unit of commodity j with technologies A and B respectively, $l_i(a)$ and $l_i(b)$ are the quantities of labour entering one unit of commodity i with technologies A and B respectively, $p_i (i = 1, 2)$ is the price of product i, w is the unit wage and r the rate of profit. The quantities a_{ij}, b_{ij}, $l_i(a)$ and $l_i(b)$ are known, whereas r, w, p_i are unknown.

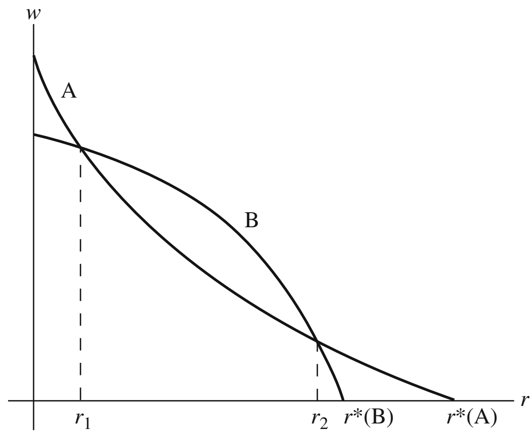
Either product is common to both systems. We may thus choose either commodity 1 or 2 as the common standard of prices (*numéraire*) in both systems. If we put the price of commodity 1 equal to unity, commodity 1 becomes the common *numéraire* of price systems (2) and (3). At this stage, it is found convenient to assess the relative profitability of alternative technologies by considering the functional relationship between r and w for each technology.

The systems of eqs. (2) and (3) would each be associated with a particular relation between the rate of profit and the unit wage. The wage–profit relationships for the two systems would respectively be given by the following expressions:

$$w_A = \frac{1 - (a_{22} + a_{11})(1+r) + (a_{11}a_{22} - a_{21}a_{12})(1+r)^2}{(1+r)[a_{21}l_2(a) - a_{22}l_1(a)] + l_1(a)} \qquad (4)$$

$$w_B = \frac{1 - (b_{22} + b_{11})(1+r) + (b_{11}b_{22} - b_{21}b_{12})(1+r)^2}{(1+r)[b_{21}l_2(b) - b_{22}l_1(b)] + l_1(b)} \qquad (5)$$

It may be immediately noted that w is always a decreasing function of r, independently of the sign of the second order derivative (see also Morishima 1966, p. 521). We may also note that the unit wage is expressed in terms of the same numéraire in (4) and (5). This suggests that the relationships between r and w (also known as factor-price frontiers) can be plotted as negatively sloped curves on the same diagram.



Reswitching of Technique, Fig. 2

The intersections between the two curves occur at those levels of the rate of profit which are associated with the same unit wage in both technologies. The number of intersections can be obtained by equating w in eqs. (4) and (5) and solving for r. The resulting equation will generally have more than one positive solution (Bruno et al. 1966, p. 34). In the case of technologies such that each product is a necessary input for all commodities including itself (all products are *basic commodities*), the maximum number of intersections is given by the number of distinct commodities in the two alternative systems of production (Bharadwaj 1970). This implies that, in the two-good technologies of our example, there will be at most two intersections. Figure 2 represents a case in which there are two intersections in the positive quadrant.

Technologies A and B can now be compared, on grounds of profitability, by considering which technology yields the higher rate of profit for any given wage. (Or, alternatively, which technology yields the higher wage rate for any given rate of profit.)

Figure 2 makes clear that the relative profitability of the two technologies is subject to fluctuation as r increases from 0 to r^*(B) (the maximum rate of profit with technology B). At a low level of the rate of profit ($r < r_1$), technology A is more profitable ('cheaper') than B. At $r = r_1 = r_2$, A and B are equally profitable. At levels of r between r_1 and r_2, B is more profitable than A. But at any rate of profit higher than r_2, A is again more profitable than B.

Reswitching of technique may be shown to be possible between complete production systems as well as between individual techniques. Shortly after the identification of the reswitching possibility by Champernowne (1953) and Robinson (1956), and its subsequent analysis by Sraffa (1960), Morishima (1964) and Hicks (1965), David Levhari (1966) proposed the argument that reswitching between production systems is possible only in the case of a 'reducible' or 'decomposable' technology matrix, so that reswitching would not occur with technologies producing only basic commodities ('irreducible' or 'indecomposable' technologies). Levhari's argument was disproved by Pasinetti and others (Pasinetti 1966; Morishima 1966; Garegnani 1966). It was also acknowledged to be false by Levhari and Samuelson (Levhari and Samuelson 1966; Samuelson 1966). Conditions excluding reswitching were then discovered by Bruno et al.(1966) and other authors (Starrett 1969). Their outstanding feature is the introduction of technological assumptions that eliminate those 'complicated patterns of price-movement with several ups and downs' (Sraffa 1960, p. 37) on which the very possibility of reswitching is founded.

As shown above, the possibility of reswitching in the comparison between alternative states of the economy is associated with differences in the proportions between labour and intermediate inputs for any given pair of production techniques.

This implies that reswitching can be observed only if the economic system is represented in such a way as to bring in view the 'ups and downs' of relative prices. This property was implicitly recognized by John Hicks (1973), when he noted that the possibility of reswitching arises when techniques are 'no longer capable of being distinguished by a single parameter', so that any switch along the technological frontier 'will be a matter of balance between advantages and disadvantages, a balance which itself is affected by prices' Hicks, (1973, pp. 44–5). This consideration is at the basis of Hicks's 'simple profile', in which any given technique is described by a single parameter (the ratio of construction labour to utilization labour) and reswitching is excluded (Hicks 1973, pp. 44–6). Joseph Stiglitz took a different view when he distinguished between reswitching as a possibility relative to the comparison among steady states, and 'recurrence of techniques' as an outcome for an economy 'on its optimal development trajectory' (Stiglitz 1973, p. 138). In particular, Stiglitz noted that 'recurrence of techniques may occur in technologies which do not allow reswitching', and that 'in technologies in which there is reswitching there may be no recurrences' (Stiglitz 1973, p. 139). John Wright (1975, p. 22) examined a related issue showing that reswitching can be avoided if one assumes an appropriate 'rate of fall of discount rate through time'. A few years later, Edwin Burmeister and Peter Hammond (1977) explored a related issue, and suggested that reswitching can be excluded as soon as we allow economies to 'jump' over intermediate states (techniques) along a given optimal adjustment path. Recent literature has examined the likelihood of reswitching from the computational or the empirical point of view. In particular, Stefano Zambelli (2004) has shown that a discrete production model is significantly likely to (computationally) generate a reswitching economy, whereas Zonghie Han and Bertram Schefold (2006) have found the empirical likelihood of reswitching to be significant but not very high. Another strand of literature investigated the relationship between the possibility of reswitching and the stability of optimal paths. In this connection, John Barkley Rosser further explored the problem set-up examined in Burmeister and Hammond (1977) and noted the existence of a trade off between the observability of reswitching and the smoothness of optimal adjustment paths (Rosser 1983, 2000). More recently, Michael Mandler (2005) and Bertram Schefold (2005) have discussed alternative conditions under which reswitching may or may not be associated with unstable economic dynamics.

Synthesis and Appraisal

The capital controversy of the 1960s has conclusively shown that the logical possibility of reswitching is of a general nature. Disagreement about the implications of reswitching for economic theory as a whole does not conceal the fact that a crucial discovery in the theory of technical choice was made. In particular it was shown that choice of technique is related to income distribution in a much more complex way than it was once thought to be, and that the rate of interest (or the rate of profit) cannot provide an unambiguous ranking of different technical alternatives as the distribution of income is varied.

The discussion of reswitching called attention to a paradox that had long been overlooked. This is that rational choice, in its classical formulation, presupposes not only agents capable to rank alternatives in a consistent way, but also objective states of the world making such a consistent ranking feasible (see Urmson 1950, pp. 154–9; Scazzieri 1982). The reswitching debate has shown that a 'granular' representation of production techniques leads to a complex pattern of interaction such that any given technique may be associated with two or more different positions on the profitability ranking of techniques (see above). This discovery was made possible by the consideration of price movements in a capital-using economy (see above). Its most immediate implication has been to cast doubt upon the representation of capital structure in terms of simple aggregate parables. However, reswitching also called attention to another, perhaps more fundamental, feature of technical choice. This is the dual nature of the grading procedure associated with choice. For grading situations express not only the agent's ability to rank states of the world in a consistent way, but also the possibility to rank those states in terms of 'objective' characteristics independent of the agent's preferences and choices. The reswitching debate has proved that the latter prerequisite may be a will-o'-the-wisp as soon as we consider the complex interactions that take place in a production economy.

See Also

- ▶ Capital Theory
- ▶ Capital Theory (Paradoxes)
- ▶ Preference Reversals
- ▶ Reverse Capital Deepening
- ▶ Technical Change

Bibliography

Bharadwaj, K. 1970. On the maximum number of switches between two production systems. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 106: 409–429.

Bruno, M., E. Burmeister, and E. Sheshinski. 1966. The nature and implications of the reswitching of techniques. *Quarterly Journal of Economics* 80: 526–553.

Burmeister, E., and P. Hammond. 1977. Maximin paths of heterogeneous capital accumulation and the instability of paradoxical steady states. *Econometrica* 45: 853–870.

Champernowne, D. 1953. The production function and the theory of capital: A comment. *Review of Economic Studies* 21: 112–135.

Garegnani, P. 1966. Switching of technique. *Quarterly Journal of Economics* 80: 554–567.

Han, Z., and B. Schefold. 2006. An empirical investigation of paradoxes: Reswitching and reverse capital deepening in capital theory. *Cambridge Journal of Economics* 30: 737–765.

Hicks, J. 1965. *Capital and growth*. Oxford: Clarendon Press.

Hicks, J. 1973. *Capital and time: A neo-austrian theory*. Oxford: Clarendon Press.

Levhari, D. 1966. A nonsubstitution theorem and switching of techniques. *Quarterly Journal of Economics* 79: 98–105.

Levhari, D., and P. Samuelson. 1966. The nonswitching theorem is false. *Quarterly Journal of Economics* 80: 518–519.

Mandler, M.A. 2005. Well-behaved production economies. *Metroeconomica* 56: 477–494.

Morishima, M. 1964. *Equilibrium, stability and growth: A multi-sectoral analysis*. Oxford: Clarendon Press.

Morishima, M. 1966. Refutation of the nonswitching theorem. *Quarterly Journal of Economics* 80: 520–525.

Pasinetti, L. 1966. Changes in the rate of profit and switches of techniques. *Quarterly Journal of Economics* 80: 503–517.

Robinson, J. 1956. *The accumulation of capital*. London: Macmillan.

Rosser, J.B. Jr. 1983. Reswitching as a cusp catastrophe. *Journal of Economic Theory* 31: 182–193.

Rosser, J.B. Jr. 2000. *From catastrophe to chaos: A general theory of economic discontinuities: Mathematics, microeconomics, macroeconomics, and finance*. Vol. 1. 2nd ed. Boston/Dordrecht/London: Kluwer Academic Publishers.

Samuelson, P.A. 1966. A summing up. *Quarterly Journal of Economics* 80: 568–583.

Scazzieri, R. 1982. Scale and efficiency in models of production. In *Advances in economic theory*, ed. M. Baranzini. Oxford: Basil Blackwell.

Schefold, B. 2005. Reswitching as a case of instability of intertemporal equilibria. *Metroeconomica* 56: 438–476.
Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.
Starrett, D. 1969. Switching and reswitching in a general production model. *Quarterly Journal of Economics* 80: 673–687.
Stiglitz, J.E. 1973. Recurrence of techniques in a dynamic economy. In *Models of economic growth*, ed. J.A. Mirrlees and N.H. Stern. London/Basingstoke: Macmillan.
Urmson, J.O. 1950. On grading. *Mind* 59: 145–169.
Wright, J.F. 1975. The dynamics of reswitching. *Oxford Economic Papers*, n.s. 27: 21–46.
Zambelli, S. 2004. The 40% neoclassical aggregate theory of production. *Cambridge Journal of Economics* 28: 99–120.

Retention Ratio

A. Cosh

The retention ratio of a corporation in any period may be defined as the ratio of retained earnings to the sum of retained earnings and dividend payments. The retention ratio is one factor in the decision concerning the optimal level of investment and the manner in which this investment is financed. The related variable, the dividend payout ratio, is defined as the proportion of available earnings paid out as dividends. In principle the sum of the retention ratio and the payout ratio should be unity. A major part of the economic debate concerning the retention ratio has mirrored the debate surrounding the debt-to-equity ratio chosen by a firm.

The subject of debate has been whether the market valuation of a firm is dependent upon its retention ratio. Miller and Modigliani (1961) conclude that, given the production and investment strategy of a firm, which determine its future earnings, the financing decision has no impact on its market value. First assume a world of perfect certainty and perfect capital markets in which all participants are price takers with costless access to all relevant information and in which there are no taxes or transactions costs. Under these conditions the choice of retention ratio will affect only the division of the return to the shareholder between dividend and capital gain and not the market valuation. When uncertainty is introduced, Lintner (1962) argues, different subjective assessments of a firm's prospects by investors will undermine the view that dividend policy is irrelevant. Furthermore, Gordon (1963) proposes that the discount rate will rise (and market value fall) with increases in the retention ratio due to the greater uncertainty of future returns. These attacks do not successfully undermine the argument in favour of the irrelevancy of dividend policy given the assumptions of perfect capital markets and the independence of the investment and financing decisions. However, the issue becomes more difficult to resolve when it is recognized that neither of these assumptions is likely to be true in reality.

Sources of capital market imperfections include transactions costs, taxes, lack of information and constraints on the supply of finance. Transactions costs include all charges concerned with the sale and purchase of shares and flotation of new shares. The existence of transactions costs limits the ability of the investor to create a 'home-made' payout ratio through dealing in shares. Taxation may influence retention ratios in a number of ways. In general, taxation of dividends is higher than that of capital gains. This will generally result in higher retentions being favoured, due to the lower rate of tax and tax deferral advantages of capital gains over dividends. However, different types of shareholders are affected in different ways, ranging from the charity or pension fund which is tax exempt to the wealthy private investor who may face a high marginal taxation of dividend income. This suggests that different types of shareholders will be attracted to different retention ratios. In equilibrium the range of retention ratios will reflect the range of shareholders, and no price advantage will be achieved by the firm through the choice of any particular retention ratio. Corporate taxation might itself be dependent upon the retention ratio, and a system which taxes dividends differentially will tend to reduce the payout ratio. Lack of information and risk aversion will tend to bias shareholders in favour of dividends. If there is a limited supply of finance, either in general or to specific companies, a higher retention ratio might result.

The residual theory of dividends suggests that the investment decision and the financing decision should be taken jointly. Providing the debt-to-equity ratio is optimal and given that taxes and transactions costs exist, the retention ratio will be determined by the availability and potential profitability of investment opportunities. Investment is taken to the point at which its prospective return is equal to the perceived opportunity cost to shareholders of dividends foregone. There are other equally important reasons for suggesting that the investment and financing decisions are not independent, which together imply that the market value may be influenced by the choice of retention ratio. Managerial models of the firm assume that management has discretion over the choice of business objectives and do not accord shareholder welfare-maximization a primary role in these objectives. In such models management may take investment beyond its optimal level financed by a higher retention ratio. This would generate a lower market valuation associated with a higher retention ratio. Less scrutiny by shareholders of investment financed by retentions would reinforce this effect. If cost-plus pricing is being employed and the margin is related to the firm's financial requirements, the retention ratio and the profitability (and market valuation) may be inversely related.

It might be hoped that empirical analysis would resolve the question of whether these multiple, and often conflicting, influences of dividend policy on market valuation yield a definite conclusion in practice. This is not the case. The evidence demonstrates that dividends are more stable then earnings and that efforts are made to avoid reductions in dividends. Dividends adapt to earnings changes over a period of time. This phenomenon means that there is an information content of dividends. Changes in dividends may provide the best guide to investors in a world of uncertainty to the future path of earnings. This results in considerable difficulty in distinguishing between the impact on share prices of changes in retention ratios themselves from the impact of the associated implications for future earnings and investment.

Therefore, the empirical evidence has not proved conclusive, but the observed dispersion of retention ratios across firms, even within the same industry, suggests that the market value is fairly insensitive to the choice of retention ratio. On the other hand, since different groups of shareholders may not be indifferent to the choice of retention ratio, and since changes in target retention ratios may be misinterpreted by shareholders, the analysis suggests that firms will not wish to change their target retention ratio.

See Also

► Dividend Policy
► Finance

Bibliography

Gordon, M. 1963. Optimal investment and financing policy. *The Journal of Finance* 18: 264–272.

Lintner, J. 1962. Dividends, earnings, leverage, stock prices and the supply of capital to corporations. *The Review of Economics and Statistics* 44: 243–269.

Miller, M.H., and F. Modigliani. 1961. Dividend policy, growth and the valuation of shares. *The Journal of Business* 34: 411–433.

Retirement

Michael Hurd

Abstract

This article uses a simple life-cycle economic model of retirement to characterize the optimal retirement age and the effects of the wage rate, wealth, and the time horizon on that age. The model is then extended to include pensions, both public and private, which can produce non-convexities in the lifetime budget constraint. The model is further extended to include health effects on retirement, uncertainty and joint retirement (the coordination of retirement dates by husband and wife). The

chapter concludes with a discussion of retirement in the context of behavioural economics.

Keywords
Age discrimination; Annuities; Bequest motive; Collective models of household; Dynamic programming; Health care expenditure; Health insurance; Labour supply; Leisure; Life expectancy; Pensions; Precautionary savings; Probability distributions; Retirement; Retirement hazard; Risk aversion; Social Security in the United States; Subjective probability; Uncertainty

JEL Classifications
J14

The common-sense definition of retirement is leaving employment of a substantial nature by a worker in his or her fifties, sixties or older with no intention of returning to work. However, this definition has no empirical counterpart because we do not observe intentions in the data. Rather, empirical work typically measures retirement in one of two ways. First, a worker is said to retire when he or she leaves the labour force in his or her fifties or older for a 'considerable' period of time. The 'considerable' period may be limited by the length of the observation period in panel data, but it is meant to distinguish retirement from normal job change by workers in their fifties or sixties. The second definition is an affirmation by the worker that he is retired. This definition aims to address right-censoring in panel data by using the individual's own assessment of retirement status. Because many workers state that they are retired after they have left a career job yet continue to work, this definition often adds the requirement of departure from the labour force. Which definition should be preferred will depend on the empirical analysis and the objective of the research. For some research questions, the definition can make a substantial difference, for example in the study of 'unretirement'. In this article I think of retirement as the transition from being in the labour force to not being in the labour force by people in their fifties or older.

Historical Trends in Labour Force Participation

In 1957 the labour force participation rate of men aged 60–64 in the United States was about 83 per cent; by 1987 it had fallen to 55 per cent, and since then has risen to about 58 per cent. The participation rate of women aged 60–64 rose over this time period because of the historical increase in the participation rate of women of younger ages: an increasing rate of retirement by older women was offset by an increasing number of women reaching age 60 and still in the labour force. Although the levels and rates of decline are somewhat different, participation rates of older men fell sharply in nine European countries and Canada. What caused these very large declines? In the United States and in many European countries the generosity of the public pension system increased sharply in the late 1960s and 1970s. For example, a good measure of the generosity of the system in the United States is the monthly Social Security benefit for men were they to retire at age 65. The average of those Social Security benefits was $307 in 1957 and $649 in 1987 (both in 1987 dollars), for an annual growth rate of 2.5 per cent. Since 1987 the real growth rate has been just 1.1 per cent and that growth has been due to wage growth, not to changes in the programme rules which have been stable. The coincidence of the decline in labour force participation with the increase in Social Security benefits suggests that Social Security was at least partly responsible for the decline, but there were changes in other determinants of labour supply as well. The private pension system expanded, and real household income increased both because of a rise in earnings and an increase in dual earner households. One objective of research on retirement has been to quantify the contributions of these and other sources to the decline in labour force participation, to predict the future course of labour force participation of the older population, and to understand the response to policy change such as alterations in the structure and generosity of Social Security.

The leading edge of the baby-boom generation will begin to retire in substantial numbers in about

2008, leading to a worsening of the financial health of the Social Security and Medicare systems in the United States. For example, the ratio of the population 65 or over to the working age population (ages 20–64) is a commonly used measure of demographic aging. In 2000 this ratio was 0.21; it is forecast to increase to 0.36 by 2030, an increase of 72 per cent. The retirement of the baby-boom generation will affect the Social Security and Medicare trust funds, requiring adjustments to those programmes. What will be the effect of those changes on retirement? In particular could policy delay retirement without unduly harming workers while improving fiscal balance? To make a good assessment of the effects of policy requires a model of retirement behaviour.

Data

Since the 1970s the most important advance in our ability to study retirement behaviour has been the development of the Health and Retirement Study (HRS). The HRS is a longitudinal data collection on about 20,000 people aged 51 or over in the United States. The HRS was fielded in 1992 with the express purpose of providing data with which to study retirement and health, and their interactions. As such it contains data on all the relevant economic variables that affect retirement, many health variables and many other non-economic variables that have additional effects. The HRS is a biennial longitudinal survey, and as of 2006 it had fielded eight waves. The original cohort was initially 51–61, so that by wave 8 it was 65–75 and had mostly retired. New cohorts aged 51–61 were added in 1998 and in 2004, and they were re-interviewed in successive waves. Based on the success of the HRS, the English Longitudinal Study of Ageing was modelled on HRS and was fielded in England in 2002. It is also a biennial panel. The Survey of Health, Ageing and Retirement in Europe was fielded in 2004 in 11 European countries, and a second wave with an expanded roster of countries followed in 2006. It is modelled on the HRS and ELSA with the aim of providing data that will permit international comparative studies.

Economic Models of Retirement

Retirement is an aspect of labour supply, and so the same general framework applies. However, in a number of ways it is easier to study retirement than hours worked: most of the retirement incentives are well measured; typically (although not always) retirement can be freely chosen whereas the choice of hours may be constrained by the demands of employers. As a consequence the response of retirement to incentives is substantial whereas the response of hours to the wage rate, at least among males, is small. Although some models of retirement are very complex, many of the ideas can be illustrated with a simplified version of a retirement model which, nonetheless, incorporates most of the important aspects of economic model of retirement.

Retirement must be placed in a life-cycle context because the gain from additional work is an addition to lifetime economic resources, and its value depends on life expectancy. Consider a worker who will live another N years and who is contemplating whether to retire. Should he work another year he would lose a year of leisure which has utility of U and which initially I assume is constant no matter what the age of the worker. He would gain a year's income which he could add to his stock of wealth. The increase in utility from the income is $V' \times wage$: the marginal utility of wealth multiplied by the annual wage. To maximize utility the worker should not work when $U > V' \times wage$. Under the universal assumption that the marginal utility of consumption declines in consumption, V' will be smaller at older ages with wealth held constant: at older ages the fixed amount of wealth would have to be consumed over fewer years so that per period consumption would be greater than at younger ages. Greater consumption would cause the marginal utility of consumption to be lower and therefore the marginal utility of wealth to be lower. At some age V' declines enough that $V' \times wage < U$, and at that age the worker would leave the labour force.

In a complete life-cycle model, consumption and, therefore, saving would be chosen by the worker as well as retirement. Yet we would like to think of a 'wealth' effect on retirement.

Variation in wealth across workers and the accompanying variation in retirement ages can be generated by variation in wages to which the worker reacts both in the choice of consumption and the retirement age. In this example, wealth is endogenous; but we might think of some variation in wealth that is exogenous to the model. Examples would be variation in initial wealth or through inheritances, variation in rates of return on assets or variation in required expenditures during the working life such as the number of children. Having in mind some exogenous variation in wealth across individuals, I will speak of a 'wealth effect' on retirement, but it should be understood that its estimation is difficult because it is endogenous in a complete life-cycle model of retirement behaviour.

These ideas are illustrated in a model of retirement choice. In this model the worker's problem is to choose the retirement age R and consumption level c to maximize lifetime utility

$$\max_{R,c}\left\{\int_0^R (u(c)+L)dt+\int_R^N (u(c)+L+U)dt\right\}$$

where c is consumption which is assumed to be constant. (In this model with fixed lifespan, consumption will be constant if the interest rate and the subjective time rate of discount are the same.) L. is baseline utility from leisure and U is the additional utility from leisure that someone gets when retired. The lifetime budget constraint is $Nc = Rw$ where w is the fixed wage. A corner solution is possible when someone places little value on leisure: he will work his entire life. But in the more usual case of retirement before age N, the solution satisfies the lifetime budget constraint and the first-order condition

$$u'(c) = U/w$$

where u' is marginal utility. Then some manipulation will show that $dR = dU < 0$, and $dc = dU < 0$: an increased value of leisure in retirement will reduce the retirement age and the budget constraint will require a reduction in consumption. Also, $dR = dN > 0$: increases in life expectancy will increase the retirement age. Because those in good health have greater life expectancy, healthy people will work longer, independent of any health effect on productivity or on the disutility of work.

An increase in the wage will increase consumption: $dc = dw > 0$. The effect of w on R is indeterminate because of the income and substitution effects whose relative magnitudes depend on the utility function. For example, if utility is constant relative risk aversion so that $u'(c) = c^{-\gamma}$, then $dR = dw > 0$ if $\gamma < 1$ and $dR = dw < 0$ if $\gamma > 1$.

In the context of this model and other models that allow consumption to be chosen, wealth will be an object of choice. We can observe co-variation in wealth and R across individuals due to variation in w, but that co-variation will not show how R would change were we to add additional wealth to someone's wealth holdings. In this model such an addition would reduce R whereas the observed variation in assets at retirement associated with variation in R could either be positive or negative depending on the details of the utility function.

Because u' is constant in age, the model predicts that once retired, no one will 'unretire'.

The Retirement Hazard

A common object of study in retirement research is the retirement hazard: the probability of retirement at age t given working at t. The retirement hazard can be found from the simple retirement model by considering only the part of the population still working at age t. Among those workers find those who will chose $R = t + 1$; the ratio of the number of those workers to the number of workers at t is the retirement hazard.

Estimation of a hazard model requires panel data where the hazard would be expressed in discrete time: the probability of retirement between t and $t + 1$ conditional on working at t. Retirement hazard models are a rather natural way to think about the retirement process particularly in the context of time-varying covariates such as the wage rate or health – just as in the simple model, an increase in wealth will increase the retirement hazard because of the reduction in V'.

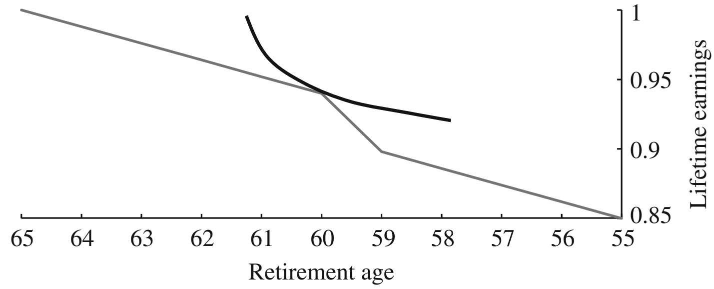
The model does not make a prediction about the variation in the retirement hazard with age, which will depend on the rate at which V' declines with wealth.

Other predictions depend on whether we are thinking of long-run comparisons across individuals, or short-run reactions by an individual to a change in the environment. For example, if we compare the retirement behaviour of two individuals, one who has a high wage rate and one who has a low wage rate, we cannot predict who will retire earlier because their saving rates would have been different and so their wealth levels would be different: in the comparison of U with $u'w$, u' and w move in opposite directions.

A good deal of the work on retirement comes from extensions of this simple model to take account of complexities in the budget set, changes in U with age and uncertainty about the future. A leading example is the study of the effects of private pensions on retirement. Private pensions are either defined contribution (DC) pensions or defined benefit (DB) pensions. In a DC plan, the employer and/or the employee puts money as specified by the plan into an investment account usually at each pay period. The amount is a small fraction of pay (say, six per cent) which implicitly increases the wage by a small per cent. The account grows at the rate determined by the portfolio held in the account. At retirement the funds are available to the retired worker typically to spend as he wishes. Thus the plan is defined by the contribution rules (hence, DC). What the worker actually receives will depend on the performance of the portfolio.

In a DB plan the worker will receive a pension or annuity at retirement which is based on the years of service with the employer, on the age at retirement and on a measure of earnings in the last few years of employment. Thus the pension plan is defined by the benefit that a worker will receive on retirement. Most DB plans have the curious feature that the benefit will depend on the age of retirement in a highly nonlinear way. If PV_a is the expected present value of lifetime pension benefits (pension wealth) conditional on retiring at a, $PV_{a+1} - PV_a$ is the addition to pension wealth (additional compensation) from working from a to $a + 1$. DB plans often have a critical age, say A, at which a full or unreduced pension benefit is paid to a worker who retires at that age. Workers who retire before A may have their pension benefit reduced substantially. Then the apparent compensation from working from $A - 1$ to A is the wage plus $PV_A - PV_{A\text{-}1}$. It is not hard to find examples where the total compensation is more than twice the wage. Said differently, the pension is reduced sufficiently for early retirement so that the gain in pension wealth from a year's work exceeds the wage. Furthermore, often pensions are not adjusted upward if a worker retires past A even though for a given pension level the expected present value declines with age. For example, for a single male aged 60 under the assumption that the pension is not indexed and that the nominal interest rate is five per cent, the pension should increase by about eight per cent per year of delayed retirement after age 60 to keep pension wealth constant. If the pension is not adjusted upward at all, the implicit wealth loss from delaying retirement for a year would be about eight per cent of the expected present value of the pension. Assuming the pension replaces 50 per cent of the pre-retirement wage and assessing pension wealth at 12.2 times annual pension income (which is PV_A in this example) means that the loss in pension wealth from delaying retirement is about 50 per cent of the wage. Said differently, the worker would be working for just half of the apparent wage.

The large gain before age 60 in pension wealth creates a large gain in compensation for working from 59 to 60, and the large loss reduces compensation substantially should the worker not retire at 60. These changes in DB pension wealth modify the money wage to produce net compensation. Net compensation for working from age 59 to 60 would be large and net compensation for working from 60 to 61 would be small. It is likely that a worker aged 59 would calculate that $U < V' \times wage$ because *wage*, which is understood to be the net wage, would be large. A worker aged 60 would calculate $U > V' \times wage$ because the net wage would be small. Thus we would observe many retirements at age 60. More generally spikes in compensation induced by DB plans cause



Retirement, Fig. 1 Lifetime earnings and stylized indifference curve

correspondingly large spikes in the retirement hazard at critical ages such as *A*. An important part of research on retirement is to obtain data on the details of DB plans so as to relate retirement to these spikes.

DC plans matter, but mainly as an addition to wealth. Typically DC plans do not have special ages at which the implicit compensation is very large or small: rather they add a small percentage to the implicit wage. (However, DC plans can have early withdrawal penalties which may affect retirement among those who have no private savings.)

This simple model of retirement has a number of advantages: estimation would show the effect of changing the net wage or changing wealth on the retirement hazard, and the estimations follow in a straightforward way from what is directly observable in data. It is clear what variation in the data produces the results. The model has considerable flexibility: the retirement hazards can be age specific so that the wage and wealth effects are different for each age. Because the estimation only requires the net wage and wealth at t and the retirement outcome at $t + 1$, just two waves of panel data are needed. Data far out of sample are not needed: for example, to study the retirement hazard of 59 year-olds, one does not need data on what their wage would be should they work until, say, 65.

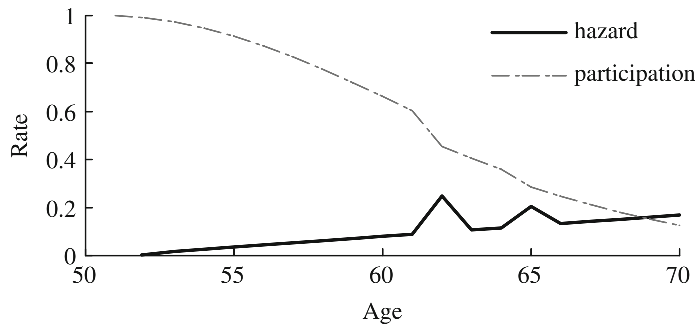
However, the simple model has a number of disadvantages. Sometimes DB pensions can induce non-convexities in the lifetime budget constraint as shown in Fig. 1. In that figure the vertical axis is lifetime earnings on an arbitrary scale, and the horizontal axis is age at retirement, inverted to show increasing years of retirement. The lifetime budget constraint has a slope equal to the net wage. In our example the large implicit net wage from working while 59 causes the slope of the budget constraint to steepen, and the small implicit net wage from working while 60 causes it to flatten. We would expect that normal shaped indifference curves would cause large numbers to retire at age 60, and, indeed, this is what is observed in data when we have good information about DB plans. The simple model would replicate this clustering at 60. But the simple model would not replicate the prediction that very few would retire at 58: the apparent gain from working while 57 is about the same as the gain from working at 55, 56 and 58, so the simple model would predict that about the same number would retire at each of those ages. But workers can foresee that if they work until they are 59 they will have the option of working from age 59 to age 60, resulting in considerable financial gain. Of course the reason the simple model would not replicate the data is that it makes only a two period comparison: retirement at t compared with retirement at $t + 1$. It does not make global utility comparisons.

The simple model does not take account of uncertainty. For example, if a possible decline into bad health will require considerable health care expenditures at some future date, a worker may consider retiring later to build up precautionary saving. In this simple model any such tendency would show up in other estimated parameters.

A second type of model is designed to handle non-convexities in the budget set. It is the option value model. In a simplified form it specifies a utility function in which utility depends on years of leisure and on lifetime earnings (Stock and

R

Retirement, Fig. 2 Stylized labour force participation and retirement hazard rates



Wise 1990). A worker will chose the retirement that maximizes utility which is shown at age 60 in the fig. A worker aged 58 would observe that the gain from working another year would be relatively modest, but that the gain from working two more years would be substantial. He would work another year so as to have the option of working the year after that.

The main advantage of the option value model over the simple model is that it can account for non-convexities in the budget set. If properly specified and estimated, it can simulate a greater range of policy options than the simple model. For example, an expansion of the budget set at age 63 by the introduction of a work bonus could only locally affect predicted retirement in the simple model: a worker would have to remain in employment until age 63 in the absence of the alteration. But in the option value model workers who had been contemplating retirement at 60, 61 or 62 could be affected.

A disadvantage of the option value model is that it is dependent on the specification of the utility function. Also it requires the construction of the budget set even at ages where the worker is not observed to work. In the extreme, it requires the construction of the budget set for all future ages. For example, if a worker continues to work at age 57, it may be that he has strong tastes for work or it may be that he has a DB plan with a large incentive to work until age 60. To study his retirement behaviour at age 57 we have to construct the budget set that he perceives at age 57.

In the same way as the simple model, the option value model does not account for uncertainty. This creates some tension because the model assumes that at age t the worker uses information about the environment at t and has expectations about what the environment will be at $t + 1$, at $t + 2$, at $t + 3$ and so forth. Based on this information he may decide to retire at, say, $t + 4$. At $t + 1$ he will use information about the environment at $t + 1$ which will usually be different from the information that was used at t about the environment at $t + 1$. This new information along with new expectations about the future environment may cause an alteration in the intended retirement age. Yet the model does not allow the decision at t to be influenced by the knowledge that new information will be arriving and that the (tentative) decision could be changed.

Social Security and Retirement

Social Security, the public pension system in the United States, is a DB pension but it differs from private (employer provided) pensions in at least two ways. First, it is almost universal so that its empirical effects on retirement are difficult to study due to the lack of programme variation across individuals. Second, at critical retirement ages Social Security is approximately actuarially fair; that is, $PV_{a+1} - PV_a \approx 0$ for most workers so that it does not generate the strong retirement incentives of private DB plans. Nonetheless, it is clear that Social Security has an important influence on retirement. First, the retirement hazard is much greater at 62 than at 61 or at 63, as illustrated in Fig. 2. Age 62 is the age at which a worker can first claim Social Security benefits, and there is no other explanation for the elevated hazard. Until year 2003, 65 was the normal retirement age under Social Security, and, in addition, the age of entitlement to Medicare, the health care insurance plan for the elderly. Second, this pattern is found in international comparisons where there is programme variation that can help identify programme effects (Gruber and Wise 1999).

Despite the empirical evidence of the effect of Social Security on retirement, its influence is difficult to explain in an economic model: why should a worker retire at 62 and claim Social Security benefits rather than at some other age, when there are no economic gains from doing so? One possible explanation is based on a liquidity constraint: low-wage workers have been forced by the Social Security system to save more than they would desire and so they do not engage in any private saving. Thus they reach their early sixties with greater-than-desired retirement resources but access to those resources is conditional on retirement. As a consequence, they retire at 62 when they are first able to access them. A difficulty with this explanation is that many workers who have wealth (demonstrating that they were not forced to over-save by the Social Security system) also retire at 62. A second possible explanation concerns the rate of return (about three per cent real) used in the calculation of $PV_{a+1} - PV_a$. If some individuals believe they can obtain a higher rate on their investments, they can increase their lifetime resources by taking Social Security payments and investing them at a higher rate of return than the implicit rate associated with delayed claiming. It is difficult to evaluate this explanation because we do not know the rates of return people expect. A third explanation is that most people believe their life expectancies to be less than average; in that case they are better off taking Social Security benefits early because they may die before they have received substantial benefits. While individual variation in subjective survival may explain the desire of some to retire and claim Social Security benefits at 62, the average subjective survival as measured by subjective survival in the HRS is close to life-table survival rates. Thus, these factors may explain some (small) part of the excessive retirements at age 62, but they are inadequate for explaining the major part of the excess.

We expect that greater wealth will lead to earlier retirement, and, indeed a contributory factor to the decline in labour force participation in the older population is probably increases in wealth. However, a wealth effect is difficult to show empirically, for several reasons. Wealth is measured with substantial error, and this tends to reduce any estimated effects. Taste variation for retirement can mean that observed wealth is not causative for retirement but rather the result of a desired retirement age. For example, those who place little value on leisure will want to retire late in life and so will save at a low rate. Thus when they reach normal retirement age they will be observed to have low wealth and not to retire. But the delay in retirement is not the result of the lack of wealth: rather the lack of wealth is the result of wanting to retire late. In this model it is necessary to find an instrumental variable to correct for the endogeneity and observation error of wealth. As always this is difficult.

The wage rate measures the price of leisure. In the simple model which says that retirement will occur when $U > V' \times wage$ an increase in the wage would lead to delayed retirement. It is, of course, necessary to control for wealth in an estimation aimed at finding a wage effect: those with high wages in the past will have accumulated more wealth, reducing V', the marginal utility of wealth. If we do not account for wealth, we may observe little relationship between the wage and retirement in data.

Health

The HRS as well as the international data gathering efforts collected many non-economic variables that are likely to influence retirement. The leading class of additional variables measures health. In the evaluation of whether $U > V' \times wage$, U is understood as the utility of leisure relative to the utility of working. It is likely that worsening health increases U because it reduces the utility of working more than it reduces the utility from leisure. Thus a first-order effect of poor health is earlier retirement, both in cross-person comparisons (comparing those in poor health with those in good health) and within person comparisons (comparing those suffering a decline in health with those having stable health). If health declines on average with age, the retirement hazard will tend to increase in age. However, health also influences V', the marginal utility of wealth, although the direction of that influence is not solidly established. If the institutional setting exposes individuals to considerable health spending risk, V' will be higher among those in poor health because of the high productivity of (private) spending on health. If individuals are fully insured against health spending risk, V' may be lower among those in worse health: health could prevent individuals from fully enjoying what money can purchase. Of course, in simple cross-person comparisons other influences on retirement vary systematically with health and they must be controlled. For example, those in worse health have less wealth, causing V' to be larger; they have reduced life expectancy, which reduces V'; and they have lower wages. Thus the relationship between health and $V' \times wage$ is ambiguous. However, as an empirical matter, we observe in panel data that those in worse health retire earlier than those in better health.

The response to an unexpected decline in health (a health shock) is easier to understand because some factors that vary across persons are constant. If exposure to health care expenditure risk is relatively small, V' would decline because of reduced life expectancy and (possibly) because of a reduced ability to spend wealth. If, in addition, the health shock caused a decline in the wage, $V' \times wage$ would decline leading to an increased likelihood of retirement. Indeed, empirically health shocks such as a heart attack are associated with elevated retirement hazard rates.

The availability of health insurance on the job and of employer-sponsored retiree health insurance should affect retirement before the age of 65 because it will change both the expected value of out-of-pocket health care costs and the variance in those costs (Blau and Gilleskie 2001, 2006). For couples the situation is more complex because in retirement one spouse can be covered by the health insurance of the other. This variation in the provision of health spending insurance provides opportunities for the identification of an insurance effect on retirement.

Accounting for Uncertainty

The effects of uncertainty are put in sharpest perspective under the assumption that it is costly to return to work once retired. If it is not costly, an individual can simply return to work as new information arrives in the future, buffering the effects of any negative shocks. This means the decision to 'retire' has less consequence. It is undoubtedly true that it is costly to return to work once retired, although the magnitude of the cost varies substantially across persons because of differences in specific human capital by occupation.

A worker contemplating retirement should be thinking about uncertainties that he would face should he continue to work and uncertainties should he retire. The first type would include wage growth, the likelihood of job displacement, the likelihood of a health event that would limit work, the evolution of his pension entitlement and other job characteristics. Remaining on the job gives an option to experience these outcomes both positive and negative. Retiring means both forgoing these options and forgoing the option of continuing to work both in the coming year and in subsequent years. The second type of uncertainty, that associated with the state of retirement, includes the rates of return on assets, health expenditures in the health insurance environment associated with retirement, survival or life expectancy, uncertainty about one's own utility function especially about one's ability to enjoy wealth, and the utility associated with full-time, uninterrupted leisure.

It is obvious that decision-making under these kinds of uncertainty is difficult. For example, economic resources have to last for many years on average; yet a typical survival curve shows that there are significant chances of dying shortly following retirement and significant chances of dying many years after retirement. In 2003 a 65-year-old man had a life expectancy of 16.8 years, or, stated differently, he could expect to die at age 81.8. However, he had an 11 per cent chance of dying before age 70 and an 11 per cent chance of surviving to age 93. To find the optimal or even satisfactory consumption path is difficult: on the one side he would need to guard against running out of resources should he survive to 93; on the other side excessively low consumption is likely to lead to his dying with considerable wealth, which, if he has no bequest motive, is wasted.

The market solution to this problem is annuities. However, the private purchase of annuities is minimal: among 65–69-year-olds just three per cent receive any income from privately purchased annuities; among those aged 70–74, six per cent receive income from annuities. There are a number of possible explanations for the lack of privately purchased annuities. The rate of return is low because of the profit of the sellers. The price of annuities is actuarially unfair to most people because the typical purchaser lives longer than average, which increases the seller's break-even price. Finally, some people may have a bequest motive: complete annuitization would eliminate any bequests.

In my view, none of these explanations is adequate to explain the lack of annuitization. Profits are not unusually high compared with profit margins in other financial products. While annuities are actuarially unfair for many people, they are not unfair for people who expect to be long lived; yet, annuity purchase is low among such people. Even people with a bequest motive should find it advantageous to annuitize partially. A possible explanation that has been little explored concerns the actual insurance that annuities provide. Privately purchased annuities are not indexed, so people will be concerned about the real value of consumption an annuity would be able to finance at advanced old age. Even a fairly moderate level of inflation will reduce substantially the real value of an annuity over 25 years. Also, the appropriate time horizon is 25–30 years: what is the probability the annuity provider will still be in business?

Estimation of the effects of uncertainty on retirement is conducted in the context of a dynamic programming (DP) model (Rust and Phelan 1997). The model will specify all possible future states of the world and assign utilities to them conditional on economic resources. Then by well-established backward solution methods the algorithms will find the expected utility associated with continued work and associated with retiring, where the expectation is taken with respect to the joint distribution of stochastic elements in the model. Thus, the analyst will supply the probability distribution of the age of death, the probability distributions of rates of return on assets, the probability distribution of health shocks and associated spending, and so forth. The model predicts continued work when the expected utility from work is greater than the expected utility from retiring, and it predicts retirement at the age when the reverse becomes true. For the reasons discussed in the context of the certainty retirement model, expected utility associated with continued work declines with age and expected utility associated with retirement increases with age, so a worker will eventually be predicted to retire. The model is adjusted with respect to parameters and specification until the predicted retirement ages match most closely those observed in the data.

The data requirements of such a DP model are immense: the analyst needs to assign probabilities to all future exogenous outcomes such as mortality, asset returns and so forth, but the probabilities are not observable. The appropriate probabilities are subjective probabilities: those used by the respondent when making the retirement decision under uncertainty. In particular, the probabilities need not be the same as any observable probabilities of the corresponding events. For example, a population life table displays an estimate of the population mortality risk at each age. Even if the population subjective survival probabilities match those in the life table, individuals should have subjective survival probabilities that deviate from the life table because each person has risk factors that will alter the objective survival probabilities. People with above-average health will survive longer than people with below-average health; people with less education die earlier than people with more education. However, the subjective survival probabilities of people with those characteristics need not correspond even to the objective survival probabilities conditional on those observable characteristics. First, people undoubtedly have private information about their true survival probabilities. Second, they may well have biased subjective survival probabilities in the sense that the probabilities on which people base their retirement decisions are not good predictors of their actual survival. In a similar manner, people have subjective probability distributions over rates of return on assets that may not correspond at all to historic market rates of return or to rates of return predicted by any model based on rational expectations. In this situation, there are no objective data from which the analyst can find the probabilities of the stochastic events required by the dynamic programming model.

Because subjective probabilities are so important in the study of intertemporal decision making, including retirement, the HRS asks respondents to state their probabilistic beliefs about important stochastic events such as survival. Although the model requires survival probabilities to each future age, knowing even the subjective survival probability to a single age is a considerable improvement over using life tables: based on a model of the relationship between subjective survival, actual survival and a life table, one can estimate individualized subjective survival curves (Gan et al. 2005).

The advantages of DP models are that they incorporate uncertainty in a formal manner and in principle they can provide an estimate of the effects of uncertainty on retirement. For example, they could predict the response to a mean-preserving spread in survival such as a 20 per cent change of dying before age 70 and a 20 per cent change of surviving to age 93. DP models produce estimates of utility function parameters and so they are capable of out-of-sample prediction. For example, they could predict retirement patterns were the normal retirement age under Social Security increased to age 70.

The disadvantages of dynamic programming models of retirement include the data requirements. Because of the complexity and data requirements, dynamic programming models of retirement are able to account for only a limited number of stochastic events. A significant problem is that the data are not subject to validation: thus a model failure could be due to an incorrectly specified model or to invalid data, particularly the probability distributions of the stochastic events. A second disadvantage is that it is difficult to understand what in the data is causing observed model outcomes because of the complexity of the model.

Joint Retirement

Most people are married when they reach retirement age, and, because of increases in the labour force participation of wives, often both spouses work. Among working husbands aged 55–59 in 2004, 74 per cent of their wives also work. On average, husbands are about three years older than their wives, so that, when the husbands reach age 62, their wives will be about 59. A husband may be influenced by his Social Security benefits to retire but it may be disadvantageous for the wife to retire at 59. Nonetheless, we observe in data some coordination of retirement dates; that is, the probability a wife will retire given the retirement of the husband is greater than the unconditional probability that a wife will retire, and similarly for the conditional probability a husband will retire (Blau 1998; Gustman and Steinmeier 2000). A way to model this is to assume a household utility function in which the value of leisure of one spouse is increased by the leisure of the other spouse. That is, their leisures are complements. In a reduced form, the retirement of the husband will be influenced by the incentives he faces such as his wage rate and pension provisions, but also by the incentives his wife faces. Notice that these effects are in addition to any operating through the lifetime budget constraint: if their leisures are not complements, we should observe the early retirement of the husband balanced by the late retirement of the wife in compensation for the loss of earnings of the husband.

Joint retirement offers an arena for the study of household decision-making when there may be conflict between husband and wife as in a collective model of household utility. A typical empirical implementation of the collective model studies the demand for various purchased goods; but it is difficult to know which spouse benefits from a particular purchase. For example, suppose we observe that in households where the husband earns substantially more than the wife the household spends relatively little on clothing. Is this evidence that the power allocation in the household is related to relative earnings? The answer would depend on the assumption that wives benefit more from clothing purchases than husbands. In the case of early retirement, however, we observe who the primary beneficiary of the leisure is, providing sharper identification of the collective model versus the unitary model.

We should anticipate that models of joint retirement will become more important and useful. First, an increasing number of wives reach traditional retirement ages while working, so the quantitative importance of joint retirement will increase. Second, the strong influence of DB pension plans on retirement incentives has made the cost of coordinating retirement high when there are conflicts between optimal retirement dates induced by the plans. With the shift to DC plans, these conflicts will become quantitatively less important in the population because DC plans do not have the sharp retirement incentives of DB plans. We should expect an increase in the amount of coordinated retirements.

Behavioural Economics and Retirement

The discussion of the determination of retirement has assumed rational decision-making in the context of the life-cycle model: individuals and couples are assumed to maximize expected lifetime utility conditional on beliefs about the probabilities of future events. In that set-up we are still very far from testing important aspects of the model because of our limited measures of those beliefs: an apparent failure of the model could be due to its being an incorrect characterization of decision-making but it could also be due to our lack of valid measures of those beliefs. Nonetheless, there is evidence at least in saving behaviour that the forward-looking model does not apply to all people. Some people strongly prefer the present to the future; they lack the ability to process relevant information; they apparently use rules of thumb; they are heavily influenced by defaults; they do not take actions that would result in fairly large financial gains even though the cost of those actions seems to be small. These examples are about saving behaviour and portfolio choice. It is more difficult to find evidence of non-optimizing behaviour in retirement choice, possibly because it is more difficult to understand what the optimal decision is. For example, job characteristics including distaste for work, perceived discrimination, health and how it interacts with job characteristics could all have large influences on retirement; yet they are mostly unobserved.

In view of these difficulties it is worthwhile asking what would constitute evidence against the rational retirement model. One type of evidence would be an empirically important ‘normative’ retirement age: a high rate of retirement at an age that may be economically disadvantageous (or at least not economically advantageous) where that age is determined by social norms or convention. Thus, a substantial number of individuals who retire at that age would have better economic outcomes had they retired at some other age. However, to find this empirically faces considerable difficulty. At the population level in the United States excessive retirements occur at 62 and 65, ages of importance under Social Security and Medicare. To be certain that these retirements are due to convention rather than to rational economic choice, we need a great deal of information about expectations and personal tastes, and we need to have confidence that our models are complete. In my view we are not in a position to assert that individuals are making a mistake when they retire at those ages.

A somewhat different category of evidence concerns economic preparation for retirement. Do we observe large numbers of workers retiring with inadequate resources to finance their consumption in retirement? If the answer is ‘yes’, retirement is suboptimal in the sense that the marginal utility of wealth is too high for the retirement age chosen. While there is controversy in the literature about the empirical facts concerning preparation for retirement, the main assertion in the literature about a lack of maximizing behaviour has been that saving is suboptimal, not that retirement is suboptimal. Abstracting from spending on health shocks, this argument has considerable validity because saving is under the control of the individual whereas the individual has somewhat less control over retirement. For example, unemployment often leads to retirement because of the difficulty for older workers to find re-employment. Or unmeasured job characteristics such as discrimination against older workers may make continued employment uncomfortable.

Future Course of Research

The economic models of retirement that have been discussed here and estimated in the literature assume that the decision to work at a given wage is made by the worker. Yet, there is a large literature on the desire of employers to shed older workers; furthermore, there is unemployment at older ages, implying that sometimes retirement is not completely voluntary. Although there are laws against age discrimination in employment, it is likely that employers are able to put some pressure on older workers to retire. However, this observation about the demand side of the labour market for older workers has been formed during an era of increasing number of workers in their forties and fifties when firms may, indeed, have felt they had too many older workers. But as the baby-boom generation begins to retire, the attitude of firms towards older workers should change: firms will want to retain them. Thus, an important research question concerns the evolution of the demand side. How will employers accommodate older workers who may not want to work full time or with the intensity of younger workers? Connected to this question is the long-run productivity of older workers. Apparently firms have wanted to shed older workers because their productivity relative to their costs (including the cost of health care) declined with age. With changing technological requirements on the job, will this unfavourable age-related decline in productivity worsen or improve over time?

We have witnessed a long-run improvement in the health of the older population both in terms of life expectancy and in terms of disability. According to economic theory, increased longevity should cause an increase in the retirement age because of the necessity of financing increased years in retirement. A decline in disabilities will allow more workers to remain in the workforce, and better health is likely to lead to greater productivity at any given age. An important research objective is to quantify these effects and to forecast how changes in retirement will affect important policy concerns. For example, in the United States working past age 65 reduces Medicare expenditures because employer-provided health insurance pays for health care before Medicare. Should enough workers remain in the labour force, the financing difficulties with Medicare will be partially solved, requiring less vigorous policy intervention.

Research on the interaction between health and retirement will likely increase in importance. For example, the sanguine scenario of later retirement because of a reduction in the rate of disability depends on continuing improvements in health. Yet there is considerable uncertainty in forecasts of health, particularly because of the high levels of obesity in the working-age population. We do not have a good understanding of how obesity leads to disability.

Although we have made some progress in our understanding of intertemporal decision-making, a great deal remains to be done. We need to understand how people make such decisions, what information they use, what expectations or probability distributions they have and how they form them, how those expectations evolve as they approach retirement, and so forth. These investigations would be helped were we to have methods of estimating at the individual level preference parameters such as risk aversion and the subjective time rate of discount independently of actual choices. A good example is portfolio choice, where we would like to estimate risk aversion. Lacking data on expected rates of return, we have to make assumptions about those expectations so that risk aversion is conditional on those assumptions. If we knew something about risk aversion we could estimate beliefs about expected rates of return.

These objectives will likely lead to a greater use of subjective data combined with objective data. For example, rather than just studying the determinants of actual retirement, we can study the determinants of the subjective probability of retirement at some target age, say, 62. In panel data, this method can control for unobserved heterogeneity in a straightforward way because we can observe the change in the subjective probability of retirement as the environment changes (Chan and Stevens 2004).

Because of the continued evolution in survey methods and the ongoing data collection by the HRS, we will have greater sample sizes with which to estimate retirement models. We should be able to observe the effects of natural experiments that can help identify our models. For example, the HRS was in the field in 2002 for the beginning of the natural experiment of the increase in the normal retirement age under Social Security. We will be able to find directly any movement in the retirement spike associated with that age. The continued data collection, natural experiments, innovations in survey design, and the greater use of subjective data should lead to considerable progress in modelling and retirement decision and in quantifying the determinants of retirement.

See Also

- ► Labour Supply
- ► Pensions
- ► Social Security in the United States

Acknowledgment Financial support from the National Institute on Ageing is gratefully acknowledged. Many thanks to Susann Rohwedder for her valuable advice and suggestions.

Bibliography

Blau, D.M. 1998. Labor force dynamics of older married couples. *Journal of Labor Economics* 16: 595–629.

Blau, D., and D. Gilleskie. 2001. Retiree health insurance and the labor force behavior of older men in the 1990's. *Review of Economics and Statistics* 83: 64–80.

Blau, D., and D. Gilleskie. 2006. Health insurance and retirement of married couples. *Journal of Applied Econometrics* 21: 935–953.

Chan, S., and A.H. Stevens. 2004. Do changes in pension incentives affect retirement? A longitudinal study of subjective retirement expectations. *Journal of Public Economics* 88: 1307–1333.

Gan, L., M. Hurd, and D. McFadden. 2005. Individual subjective survival curves. In *Analyses in the economics of aging*, ed. D. Wise. Chicago: University of Chicago Press.

Gruber, J., and D. Wise. 1999. *Social security and retirement around the world*. Chicago: University of Chicago Press.

Gustman, A.L., and T.L. Steinmeier. 2000. Retirement in a family context: A structural model for husbands and wives. *Journal of Labor Economics* 18: 503–545.

Rust, J., and C. Phelan. 1997. How Social Security and Medicare affect retirement behavior in a world of incomplete markets. *Econometrica* 65: 781–831.

Stock, J., and D. Wise. 1990. Pensions, the option value of work and retirement. *Econometrica* 58: 1151–1180.

Returns to Scale

John Eatwell

Keywords

Constant returns; Decreasing returns; Dynamic increasing returns; Increasing returns; Indivisibilities; Returns to scale; Substitution; Technical change

JEL Classifications

D2

The technique of production of a commodity y may be characterized as a function of the required inputs x_i:

$$y = f(x_1, x_2, \ldots x_n)$$

If all inputs are multiplied by a positive scalar, t, and the consequent output represented as $t^s y$, then the value of s may be said to indicate the magnitude of returns to scale.

If $s = 1$, then there are constant returns to scale: any proportionate change in all input results in an equiproportionate change in output. If $s > 1$, there are increasing returns to scale. If $s < 1$ (though not

R

less than zero, given the possibility of free disposal) then there are decreasing returns to scale.

These mathematical definitions suggest a symmetry between the three classifications of returns to scale. This appearance of symmetry is entirely spurious.

The original arguments from which is derived the economic rationale underlying the various categories of returns to scale are to be found in the works of the classical economists. Yet there, as Sraffa (1925) pointed out, each category is derived from quite different economic phenomena. Increasing returns derived from the process of accumulation and technological change, associated as they were with the division of labour attendant upon the extension of the market. Decreasing returns were held to derive from the limited availability of land, and were an important component of the theory of income distribution, being the foundation of the theory of rent.

Yet it was from these disparate origins that Marshall (1890) attempted to formulate a unified, symmetric, analysis of returns to scale which would provide the rationale for the construction of the supply curve of a competitive industry, derived in turn from the equilibria of the firms within the industry. Marshall himself recognized the incompatibility of the assumption of competition and presence of increasing returns (1890, Appendix H). Piero Sraffa (1925, 1926) exposed the entire exercise as ill-founded by demonstrating that neither increasing nor decreasing returns to scale are compatible with the assumption of perfect competition in the theory of the firm or of the partial-equilibrium industry supply curve – a result which, although prominently published and debated, has apparently escaped the notice of those who still draw that bogus U-shaped cost curve whilst purporting to analyse the equilibrium of the competitive firm.

The difficulties identified by Sraffa rest upon the *economic* rationales for variable returns to scale.

The idea of constant returns to scale derives essentially from the proposition that a given set of production conditions may be replicated so long as all the requisite inputs may be varied in the same proportion. Indivisibilities in the production process may limit exact replication to particular levels of output. But the concept, though less precise, is not in any way diminished by the presence of indivisibilities, particularly if the optimal scale of operation of a given technique is small relative to the overall level of output.

The presence of decreasing returns *to scale* would suggest that replication is, for some reason, impossible. Yet if all inputs are correctly enumerated and all increased in the same proportion, then, barring indivisibilities, there can be no barrier to replication. Decreasing returns can derive only from a fixed input (or an input which cannot be increased in the same proportion as others) which prevents replication. In other words, there is no such thing as decreasing returns *to scale*. Decreasing returns derives from *substitution*, from the necessity of changing input proportions.

Whilst decreasing returns to scale do not exist, increasing returns to scale are typically based on propositions so general as to defy precise clarification.

There are some examples in which outputs are an increasing function of inputs for purely technical reasons. The capacity of a pipeline, for example, is defined by the area of its cross-section, πr^2 whereas the circumference of that cross-section is equal to $2\pi r$. If it were possible to increase capacity merely by increasing the circumference (if the walls of the pipe did not require strengthening), then a quadrupling of capacity could be achieved simply by doubling the material inputs.

There is one odd symmetry in this ‘technical’ case of increasing returns. Whereas decreasing returns can derive only from substitution and not from scale, increasing returns can derive only from scale, not from substitution! Choice of optimal proportions of inputs (with free disposal and no indivisibilities) will always ensure *at least* constant returns.

Such technical examples are not, however, the examples which typically come to mind in the discussion of increasing returns to scale. More typical are examples of mass production, of production lines, or, today, of production integrated by means of sophisticated information systems. Yet these examples, which are akin to Adam Smith’s analysis of increasing returns, are associated more with technological change, and with the possibilities for change inherent in a larger, or more rapidly growing, market, than with a simple increase in the scale of identical inputs. Generalization of the concept to 'dynamic increasing returns' (Young 1928; Kaldor 1966) increasing returns associated with growth of output further distances the idea of increasing returns from the formal characteristics of scale.

These arguments suggest that the concept of 'returns to scale' is not merely a very limited means of characterizing technology, but it is also a very *limiting* concept. None of the interesting characteristics of the relationship between scale of production and method of production are captured by the idea of returns to scale. Indeed, the only really satisfactory formal characterization of returns to scale is that of constant returns – and this only because replication is formally a precise notion, however empty empirically.

Bibliography

Kaldor, N. 1966. *Causes of the slow rate of economic growth in the United Kingdom*. Cambridge: Cambridge University Press.

Marshall, A. 1890. *Principles of economics*. 9th (Variorum) edn. London: Macmillan, 1961.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*. London: Methuen, 1961.

Sraffa, P. 1925. Sulla relazioni fra costo e quantità prodotta. *Annali di Economia* 2: 277–328.

Sraffa, P. 1926. The laws of returns under competitive conditions. *Economic Journal* 36: 535–550.

Young, A.A. 1928. Increasing returns and economic progress. *Economic Journal* 38: 527–542.

Returns to Scale Measurement

Susanto Basu

Abstract

If output grows faster than inputs, holding technology constant, the production function exhibits increasing returns to scale. Increasing returns in the aggregate production function may be due to overhead (fixed) costs, diminishing marginal cost, positive spillovers from aggregate activity, the entry of new varieties of inputs or changes in the distribution of inputs across heterogeneous firms. Each channel has significant implications for models of growth, trade and business cycles. Returns to scale are hard to estimate and even difficult to define, since the definition may depend on the degree of aggregation and the time horizon under study.

Keywords

Capital utilization; Cobb–Douglas function; Cost functions; Imperfect competition; Markups; Measurement error; Production functions; Productivity growth; Research and development; Returns to scale; Returns to scale measurement; Scale economies; Shadow pricing; Spillovers; Sunspot equilibrium; Technology; Transmission problem

JEL Classifications

E23

Knowing the degree of returns to scale (RTS) in a firm, or its average in an industry or economy, is important for a variety of economic questions. First, it is important for assessing the plausibility of models of endogenous growth, which typically require at least constant returns to *reproducible* inputs, and thus increasing returns overall. Second, the size of scale economies is an important determinant of the gains from trade. Third, knowing the RTS is important for assessing the plausibility of certain business-cycle models, which often rely on the existence of substantial increasing returns to scale (IRS). Fourth, the RTS is a lower bound on the size of the markup of price over marginal cost, which is a quantity of great interest in industrial organization. Finally, a basic tenet of productive efficiency requires that the value marginal product of an input be equalized across the uses to which that input could be devoted. Knowing the RTS in each use is important for checking that this condition holds.

Assume that firms have a production function for gross output:

$$Y = F\left(\tilde{K}, \tilde{L}, M, T\right). \tag{1}$$

Firms use capital services $\tilde{K}$, labour services $\tilde{L}$, and intermediate inputs of materials and energy, M. T indexes 'technology', not directly observed, which is defined to include any inputs that affect firm-level production but are not compensated by the firm (including, for example, Marshallian externalities as well as exogenous technical change). All variables are functions of time.

Assume $\tilde{L} = EHN$, where the number of employees, N, and hours worked per employee, H, are observed, but the effort of each employee, E, is unobserved. Capital services are the product of the observed capital stock, K, and its unobserved utilization rate, Z (for example, the number of shifts the machine is operated): $\tilde{K} = ZK$. The capital stock and the number of employees may be quasi-fixed (costly to adjust). The adjustment cost can be modeled explicitly in the production function (see Berndt 1986).

F is (locally) homogeneous of arbitrary degree γ in the priced inputs. Constant returns implies $\gamma = 1$: RTS equals the sum of output elasticities:

$$\gamma = \frac{F_1 \tilde{K}}{Y} + \frac{F_2 \tilde{L}}{Y} + \frac{F_3 M}{Y}, \tag{2}$$

where F_J is the marginal product of input J. Assuming firms minimize cost, we can denote the firm's cost function by $C(Y)$. γ also equals the inverse of the elasticity of cost with respect to output:

$$\gamma(Y) = \frac{C(Y)}{YC'(Y)} = \frac{C(Y)/Y}{C'(Y)} = \frac{AC}{MC}, \tag{3}$$

where AC equals average cost and MC equals marginal cost. IRS may reflect overhead (fixed) costs or decreasing marginal cost; both imply that average cost exceeds marginal cost. If increasing returns take the form of overhead costs, then $\gamma(Y)$ is not a constant structural parameter, but depends on the level of output the firm produces. To make this point more clear, consider a special case of Eq. (1):

$$Y = G(K, L, M, T) - \Phi, \tag{4}$$

where Φ is a *flow* (per-period) fixed cost and G is homogeneous of degree ρ in K, L and M. In this case, $\gamma = \rho(Y + \Phi) / Y$. Thus, RTS, γ, may strictly exceed ρ. Some papers use empirical estimates of γ to calibrate ρ. Since this procedure is not generally correct, some of the results in these papers (for example, the existence of sunspot equilibria), do not follow from the existence of IRS per se. Indeed, IRS is compatible with increasing marginal costs ($\rho < 1$), as in the standard Chamberlinian model of imperfect competition.

Even if firms are identical, the RTS of the *aggregate* production function (either of an industry or an economy) is not necessarily the γ of every firm; it also depends on the dynamics of firm entry and exit. Suppose that in the long run all changes in aggregate output are accommodated by changes in the number of firms, with firm-level output remaining constant. Then the aggregate production function has constant returns to scale in the long run, but increasing returns when the number of firms is fixed in the short run. (However, if the new firms produce new varieties of goods, then the aggregate function may exhibit a form of increasing returns through a 'love of variety' in production, as in Ethier 1982.)

Firms may charge a price P with a markup, μ, where $\mu \equiv P / MC$. RTS is a technical property of the production function, whereas the markup is a behavioural parameter. However, from Eq. (3), the two are linked:

$$\gamma = \frac{C(Y)}{YC'(Y)} = \frac{P}{C'(Y)} \frac{C(Y)}{PY} = \mu(1 - s_\pi), \tag{5}$$

where s_π is the share of pure economic profit in gross revenue. As long as pure economic profits are small, as most estimates suggest, Eq. (5) shows that μ approximately equals γ. Large markups thus require large increasing returns. Since most studies estimate low profit rates, and since $\mu \geq 1$, Eq. (5) shows that firm-level RTS must either be approximately constant or increasing. Internal IRS also *requires* that firms charge a markup, to avoid losses.

One can estimate RTS from either the production function or the cost function, using the implications of Eq. (2) or Eq. (3) (see Berndt 1986, for an exposition of the cost approach). The two literatures have developed to have different aims. The costfunction literature typically takes second-order approximations to the underlying production function, which allows it to estimate elasticities of substitution between inputs, but pays little attention to the issue that observed factor prices may not be allocative, especially at high frequencies. The production-function literature takes first-order approximations, but devotes more attention to correcting biases from unobserved right-hand side variables (the quantity analogue of unobserved true factor prices). Neither literature has found a good solution for dealing with issues of endogenous regressors (for example, the presence of output in the firm cost function when one allows for non-constant returns to scale).

Taking logs of both sides of Eq. (1) and differentiating with respect to time gives:

$$dy = \frac{F_1 K}{Y} d\tilde{k} + \frac{F_2 L}{Y} d\tilde{l} + \frac{F_3 M}{Y} dm + dt. \tag{6}$$

Small letters denote growth rates (so dy, for example, equals $\dot{Y}/Y$); the output elasticity with respect to technology is normalized to one.

Cost minimization puts additional structure on Eq. (6). (The advantage of the cost minimization framework is that it is unnecessary to specify the potentially very complicated, dynamic profit maximization problem that gives rise to P or μ.)

Suppose that firms take the price of all J inputs, P_J, as given by competitive markets. The first-order conditions for cost-minimization then imply that

$$PF_J = \mu P_J. \tag{7}$$

If firms make pure economic profits, these appear in the data as factor payments (most often to capital, sometimes to labour). In order for Eq. (7) to hold, the prices of capital and labour must be defined as the rental price (or shadow rental price) of capital and the competitive wage rate for labour. The relationship still holds if some factors are quasi-fixed (costly to adjust), as long as we define the input price of the quasi-fixed factors as the appropriate *shadow* prices, or implicit rental rates.

Using Eqs. (5) and (7), we can write each output elasticity as the product of RTS multiplied by total expenditure on each input divided by total *cost* (not revenue). Thus, for example,

$$\frac{F_1 ZK}{Y} = \gamma \frac{P_K K}{\Sigma P_J J} \equiv \gamma c_K. \tag{8}$$

c_J are the *cost shares* of each type of input, and sum to 1.

Substitute these output elasticities into Eq. (6) and use the definition of input services:

$$\begin{aligned} dy &= \gamma\left[c_K d\tilde{k} + c_L d\tilde{l} + c_M dm\right] + dt \\ &= \gamma[c_K(dk + dz) + c_L(dn + dh + de) + c_M dm] \\ &\quad + dt \\ &= \gamma[c_K dk + c_L(dn + dh) + c_M dm] \\ &\quad + \gamma[c_K dz + c_L de] + dt \equiv \gamma dx + \gamma du + dt \end{aligned} \tag{9}$$

Defining dx as a share-weighted average of conventional (observed) input growth, and du as a weighted average of unobserved variation in capital utilization and effort, we obtain our basic estimating equation for γ, the last line of Eq. (9). Note that to create the cost shares c_J one needs to construct an estimate of pure profits, as in Hall (1990). Alternatively, one can assume zero economic profit on average and use the observed revenue shares.

Regarding Eq. (9) as an estimating equation, one immediately faces three issues.

First, the econometrician usually does not observe utilization du directly. In this case, the regression suffers from measurement error. Unlike classical measurement error, variations in utilization du are likely to be (positively) correlated with changes in the measured inputs dx, leading to an upward bias in the estimated γ.

Second, should one take the output elasticities as constant (appropriate for a Cobb–Douglas production function or for a first-order log-linear approximation), or time-varying? That is, should one allow γ and the share-weights in Eq. (9) to change over time? If the elasticities are not truly constant over time, then treating them as constant may introduce bias.

Third, even if the output elasticities are constant and all inputs are observable, one faces the 'transmission problem': The technical change term, dt, is likely to be correlated with a firm's input choices, leading to biased OLS estimates of γ. In principle, one can solve this problem by instrumenting the right-hand-side variables, or by using a proxy for dt, following Olley and Pakes (1996).

Approaches to controlling for du also involve the use of proxies. One method builds on the intuition that firms view all inputs (whether observed by the econometrician or not) identically. For example, a firm should equate the marginal cost of obtaining more services from the observed intensive margin (for example, working current workers longer hours) and from the unobserved intensive margin (working them harder each hour). If the costs of increasing hours and effort are convex, firms will choose to use both margins. Thus, changes in an observed input – for example, hours per worker – provide a measure of unobserved changes in the intensity of work. This suggests a regression of the form:

$$dy = \gamma dx + \kappa dh + dt, \tag{10}$$

where dh is the growth rate of hours per worker. Basu and Fernald (2001) summarize research showing that regression Eq. (10) controls for variable effort. In addition, if the cost of varying the workweek of capital takes the form of a shift premium – for example, one needs to pay workers more to work at night – then this regression corrects for variations in utilization of capital as well as labour. (If the cost of varying capital's workweek is 'wear and tear' – that is, capital depreciates in use – then the regression is somewhat more complicated, but theory still suggests appropriate proxies.)

In principle, allowing for time-varying factor shares in an estimating equation like Eq. (10) is always preferable to having constant shares, since using time-varying shares approximates the true function to a second order. However, attempting to estimate the time-varying shares requires observing (or estimating) the true shadow cost of inputs at each point in time. If observed factor payments at each point in time do not correspond to the factor's true cost each period – for example, if firms smooth wage payments by offering workers insurance through an implicit contract – then treating the observed prices as allocative may introduce larger biases (see also Carlton 1983, on intermediate goods prices).

Since one is unlikely to observe allocative factor prices period-by-period, one probably should take a first-order approximation and assume constant, not time-varying, elasticities. For estimating the RTS a first-order approximation may suffice, and it will be accurate as long as the true average factor price is the mean of the observed prices over the sample period.

So far, the discussion has concerned the estimation of internal returns to scale. However, a number of interesting models assume the existence of spillovers between competitive firms with internal constant returns, leading to *external* increasing returns in the aggregate production function. The empirical literature searching for such spillovers follows two sharply divergent tracks. The search for high-frequency spillovers is usually atheoretical, and amounts to augmenting disaggregated estimating equations like Eq. (10) with measures of aggregate activity. However, since most such exercises do not attempt to control for unobserved changes in utilization (omitting, for example, the κdh term in Eq. (10)), they are vulnerable to the charge that the putative externalities are actually proxies for unobserved changes in internal inputs. Furthermore, Basu (1995) presents a model where apparent external effects are actually driven by a different economic mechanism, and shows that his model can be distinguished from true technological spillovers by examining gross-output data, as opposed to the commonly used value-added data. Performing the test, apparent externalities are found in value-added but not in gross output, suggesting they are not true spillovers.

However, the search for long-run external effects is based firmly on the economic insight that knowledge creation has built-in increasing returns, since knowledge is non-rival. Thus, there is a long tradition of searching for externalities to R&D, summarized by Griliches (1998). R&D spillovers appear to be a fact, but their exact magnitude is still an issue subject to debate. And there is no consensus at all on whether the magnitude of the spillover is large enough to permit fully endogenous long-run growth.

So far, the discussion has been couched in terms of firm-level output, or aggregation over identical firms. For some applications, one wants to know the RTS for an industry or a sector but allow – plausibly – for the possibility that firms have heterogeneous characteristics, including different γ's. It turns out that, in this realistic scenario, there is not even an unambiguous definition of increasing returns to scale. Basu and Fernald (1997) show that industry output growth equals:

$$d\bar{y} = \bar{\gamma}d\bar{x} + d\bar{u} + R + d\bar{t}. \tag{11}$$

$\bar{\gamma}$ is the average RTS across firms; $d\bar{y}$, $d\bar{x}$ and $d\bar{u}$ are appropriately – weighted averages of firm-level output and input growths; R represents various reallocation (or aggregation) effects; and $d\bar{t}$ is an appropriately – weighted average of firm-level technology.

The intuition for 'R' is that γ need not be the same across firms within an industry (or the economy). Output growth therefore depends on the *distribution* of input growth as well as on its mean: if inputs grow faster in firms where they have above-average marginal products (γ is higher), industry output grows more rapidly as well. Thus, aggregate productivity growth is not just firm-level productivity growth writ large; comparing Eqs. (9) and (11) shows that there are qualitatively new effects at the aggregate level. Is the RTS of an industry just the average of firm-level RTS, $\bar{\gamma}$, or does it include the aggregation effects, R, which are also the result of deviations from constant returns and perfect competition? The answer will depend on the economic question being asked (see Basu and Fernald 1997, section V), but empirically the magnitudes are often quite different.

See Also

- ▶ Capital Utilization
- ▶ Cyclical Markups
- ▶ External Economies
- ▶ Multiple Equilibria in Macroeconomics
- ▶ Production Functions
- ▶ Returns to Scale
- ▶ Technical Change

Bibliography

Basu, S. 1995. Intermediate goods and business cycles: implications for productivity and welfare. *American Economic Review* 85: 512–531.
Basu, S., and J.G. Fernald. 1997. Returns to scale in U.S. production: Estimates and implications. *Journal of Political Economy* 105: 249–283.
Basu, S., and J.G. Fernald. 2001. Why is productivity procyclical? Why do we care? In *New developments in productivity analysis*, ed. C. Hulten, E. Dean, and M. Harper. Cambridge, MA: NBER.
Berndt, E. 1986. *The practice of econometrics: Classic and contemporary*. New York: Addison-Wesley.
Carlton, D.W. 1983. Equilibrium fluctuations when price and delivery lags clear the market. *Bell Journal of Economics* 14: 562–572.
Ethier, W.J. 1982. National and international returns to scale in the modern theory of international trade. *American Economic Review* 72: 389–405.
Griliches, Z. 1998. *R&D and productivity*. Chicago: University of Chicago Press.
Hall, R.E. 1990. Invariance properties of Solow's productivity residual. In *Growth/Productivity/Unemployment: Essays to celebrate Bob Solow's birthday*, ed. P. Diamond. Cambridge, MA: MIT Press.
Olley, G.S., and A. Pakes. 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64: 1263–1297.

R

Returns to Schooling

David Card

Abstract

The returns to schooling represent the incremental increase in earnings associated with an increase in schooling. Under assumptions first spelled out by Jacob Mincer in 1958, each

additional year of schooling will lead to a percentage gain in earnings that is equal to the interest rate. More recent research has treated the return to schooling as a causal parameter that can vary across people, and by the level of education.

Keywords

Ability bias; Control functions; Degree effects; Equalizing differences; Human capital earnings function; Instrumental variables; Mincer, J.; Publication bias; Returns to schooling; Self-selection bias; Treatment effect; Wage differentials

JEL Classifications

I2; J3

Introduction

The return to schooling is the internal rate of return on an additional year of schooling: the discount rate at which the present value of the gains associated with the investment equals the costs. The notion of treating education as a capital investment – and calculating the return accordingly – was proposed by Walsh (1935) in an aptly titled article, 'Capital Concept Applied to Man'. Subsequent contributions (Mincer 1958, 1974; Becker 1962, 1964, 1967) have elaborated the theoretical underpinnings of this exercise, while advances in data availability and econometric methods have led to refinements in the empirical procedures used to calculate the return to schooling (see Griliches 1977; Card 2001; Harmon et al. 2003, for surveys).

Following Mincer (1974), the term return to schooling also refers to the coefficient of years of schooling in a linear regression of log earnings on years of schooling and controls for labour market experience. Under certain simplifying assumptions this coefficient is approximately equal to the internal rate of return to an additional year of schooling (see section "The Internal Rate of Return and Equalizing Differences" below). More generally, however, applied economists use the term return to schooling to denote the causal effect of additional schooling on log earnings, holding constant experience (or in some cases age). In this sense, which I will adopt below, the return to schooling is a structural parameter that may vary with the level of schooling, personal characteristics, and the economic environment. Moreover, the observed *ex post* returns to schooling can differ from the *ex ante* returns that were anticipated when the schooling decision was made (Cunha and Heckman 2006).

Theoretical Framework

The Internal Rate of Return and Equalizing Differences

The internal rate of return is an accounting concept that can be implemented without reference to a particular theory of wages and schooling. As was recognized by Walsh (1935), however, if there is free entry into different schooling options, and if increases in the supply of workers with a given schooling level reduce relative wages for the group, internal rates of return to different choices will be driven down to a common level.

(Walsh 1935, p. 284, wrote: 'Investment in training ... tends to be made as long as the returns promise to cover the cost of that training with an ordinary commercial profit. And this of course is the fundamental characteristic of the competitive, equalizing market . . .').

Using this 'equalizing differences' framework, Mincer (1958) showed that the equilibrium wage differential between two occupations requiring differing amounts of schooling will equal the difference in years of schooling multiplied by the discount rate.

Willis (1986) considers the choice of an optimal schooling level S (measured in units of time) under four assumptions: (1) individuals maximize the discounted present value of earnings using a common interest rate r; (2) earnings are zero while in school, and equal to $f(S)g(t - S)$ at age t (where age is measured in units of time since the completion of compulsory schooling); (3) the duration of work life is independent of S; (4) the only cost of schooling is the opportunity cost of forgone earnings. Under these assumptions, the internal rate of return for a marginal increase in schooling from an initial level S_0 is $f'(S_0)/f(S_0)$ – that is, the proportional earnings differential per year of education between people with schooling S_0 and those with a little more (or less), holding constant work experience. (Under the assumptions specified, the internal rate of return r equates $V(S_0, r)$ and $V(S_0 + \varepsilon, r)$, where $V(S,r) = \int_S^{S+n} f(S)g(t-S)e^{-rt}dt = f(S)e^{-rS} \int_0^n g(x)e^{-rx}dx$. Equality implies that $e^{r\varepsilon} = f(S_0 + \varepsilon)/f(S_0) \approx 1 + \varepsilon f'(S_0)/f(S_0)$ and taking the limit as $\varepsilon \to 0$ gives $r = f'(S_0)/f(S_0)$.)

If people can choose freely between schooling opportunities, in equilibrium log earnings will be a linear function of years of schooling (with slope r), and the internal rate of return for any schooling choice will equal r. Consistent with this insight, one of the most important regularities in labour economics is that a regression of log earnings on years of schooling and controls for experience yields a coefficient that is comparable to a discount rate for a risky investment – of the order of 5–15 per cent per year. Though the precise magnitude of such a coefficient varies over time and across labour markets, the predictability of the magnitude of the estimated return to schooling is unmatched in any other area of empirical microeconomics.

An Extended Model

While a simple equalizing differences framework provides a useful starting point for understanding the relationship between earnings and schooling, it does not explain why different people choose different levels of schooling. In fact, children's education choices are strongly correlated with their parent's schooling and socio-economic status, and with their own test scores in early grades. (See Card 1999, and Solon 1999.)

These correlations raise a fundamental question: to what extent do people with more education have other attributes – like ability or privileged family background – that would cause them to earn more even in the absence of extra schooling? In the literature this possibility is known as ability bias. A closely related question is whether people who acquire additional schooling have higher returns than those who do not – a sorting or self-selection bias of the type identified by Roy (1951).

Becker (1967) presented a simple model of earnings and schooling determination that can be used to address these issues. In this model, an individual faces a market opportunity locus $y(S)$ that gives the level of earnings y associated with different schooling choices S, and chooses a level of schooling by equating the marginal benefit of schooling with the marginal cost. Following Card (1995a), it is convenient to assume the individual chooses S to maximize a utility function $U(S, y) = \log y - h(s)$, where h is an increasing convex function. An optimal schooling choice satisfies the first-order condition

$$h'(S) = y'(S)/y(S).$$

Note that, because the objective function is linear in log y, the optimal choice of schooling is independent of factors that generate a parallel shift in the log $y(S)$ function. Griliches (1977) presented a more general model of preferences with the feature that a uniform upward shift in log earnings for all levels of schooling leads to a lower schooling choice.

Individual heterogeneity in the optimal schooling outcome arises from two sources: differences in the costs of (or tastes for) schooling, represented by heterogeneity in $h(S)$; and differences in the economic benefits of schooling, represented by heterogeneity in the marginal return $y'(S)/y(S)$. A tractable assumption is that both functions are linear in S, with additive heterogeneity components:

$$y'(s)/y(S) = b_i - k_1 s, h'(S) = r_i + k_2 S.$$

Here b_i and r_i are random variables with means $\overline{b}$ and $\overline{r}$ and some joint distribution across individuals (indexed by i), and k_1 and k_2 are non-negative constants. This specification implies that the optimal schooling choice is *linear* in the individual-specific heterogeneity terms:

R

$$S_i^* = (b_i - r_i)/k,$$

where $k = k_1 + k_2$.

The assumed model for the marginal returns to schooling implies that log earnings are generated by a model of the form

$$\log y_i = \alpha_i + b_i S_i - \tfrac{1}{2} k_1 S_i^2,$$

where α_i is a person-specific constant of integration. This is a generalization of the semi-logarithmic functional form adopted in Mincer (1974) and hundreds of subsequent studies. In particular, individual heterogeneity potentially affects both the *intercept* of the earnings equation (via α_i) and the *slope* of the earnings-schooling relation (via b_i). In general the optimal schooling choice will be positively correlated with b_i, leading to a 'self-selection bias' that arises because people with higher returns to schooling acquire more schooling. If α_i is also positively correlated with S_i (via a positive correlation with b_i or a negative correlation with r_i) the relationship between earnings and schooling will also include an 'ability bias', that is, a bias that arises because people with a higher level of earnings for each level of schooling have characteristics that lead them to acquire more schooling.

A particularly simple version of this model has only two schooling choices (Willis and Rosen 1979). In this case the model reduces to a discrete choice model for the longer schooling option, and an earnings equation with a random intercept and random coefficient on a dummy representing the longer schooling option. A more general version arises if one relaxes the linearity assumptions for the marginal costs and marginal returns, but maintains additive heterogeneity: that is, $y'(S)/y(S) = b_i + \lambda(S)$, $h'(S) = r_i + \mu(S)$. In this case, Rau-Binder (2006) shows the optimal schooling is $S = \theta^{-1}(b_i - r_i)$, where $\theta(S) = \mu(S) - \lambda(S)$, and log earnings are generated by a model of the form $\log y_i = \alpha_i + \phi(S) + b_i S_i$, where $\varphi'(S) = \lambda(S)$. What does this class of models imply about the return to schooling? For individual i, the marginal return to the last unit of schooling is:

$$\beta_i = b_i - k_1 S_i^* = b_i(1 - k_1/k) + r_i k_1/k,$$

which varies across people unless one of two conditions is satisfied: either $b_i = \overline{b}$ for all i and $k_1 = 0$ (so each additional unit of schooling has the same proportional effect on earnings for everyone); or $r_i = \overline{r}$ for all i and $k_2 = 0$ (so everyone uses the same discount rate and invests in schooling until the return on their last unit of schooling is driven down to $\overline{r}$). Even if one of these conditions is satisfied and β_i is constant across the population, it is not necessarily true that one can obtain an unbiased estimate of the average marginal return to schooling $\overline{\beta} = E[\beta_i]$ from observational data on earnings and schooling. In the first case (homogeneous returns) the implied earnings model is

$$\log y_i = \alpha_i + \overline{b} S_i.$$

Only if α_i and S_i are uncorrelated will an ordinary least squares (OLS) regression yield a consistent estimate of $\overline{b}$. In the second case (homogeneous interest rates) the implied earnings model is:

$$\log y_i = \alpha_i + \overline{r}\, S_i + \tfrac{1}{2}\, k_1 S_i^2.$$

Since people with higher values of b_i invest in more schooling, the implied relationship between earnings and schooling is convex, leading to an upward bias in the OLS estimator relative to the true marginal return to schooling, $\overline{r}$ (Mincer 1997). Any correlation between α_i and S_i will confound the situation even further.

For the general case where marginal returns vary across the population, Card (1999) shows that an OLS regression of earnings on schooling yields a coefficient b_{ols} that has probability limit

$$\text{plim } b_{ols} = \overline{\beta} + \lambda + \psi \overline{S},$$

where $\lambda = \text{cov}[\alpha_i, S_i]/\text{var}[S_i]$ represents an ability bias term and $\psi = \text{cov}[b_i, S_i]/\text{var}\,[S_i]$ represents a self-selection or sorting bias term. (This expression assumes that the heterogeneity terms have symmetric distributions – see Card 1999.) Since people with higher returns at each level of schooling will tend to acquire more schooling, the sorting bias term should be positive, although the magnitude may be small. The sign of the ability bias term is less clear: several studies – including the seminal paper by Willis and Rosen (1979) – have obtained negative estimates of λ. In any case, observed pay differences between people with different levels of education may imply rates of return that are above or below $\overline{\beta}$, the average marginal return to education in the population.

The Mincer–Willis equalizing differences model is a long-run general equilibrium model in which wage differentials across education groups are determined by a free entry condition on the supply side. Becker's (1967) model, in contrast, is a partial equilibrium model describing the schooling decisions made by different individuals in a given cohort, taking the earnings generating function as given. Once these decisions are made, shifts in the demand and supply for different education groups can lead to realized returns that are higher or lower than were originally anticipated *ex ante*. Moreover, the fraction of a cohort that acquires higher education can affect their *ex post* returns – a general equilibrium effect. In the mid-1970s for example, the college-high school premium in the United States was relatively low, and analysts described an 'oversupply' of college-educated labour (Freeman 1976). Within 15 years, however, the premium bounced back, and it now appears that cohorts born in the 1950s have enjoyed higher returns to education than they expected *ex ante*.

Dynamic Models of Schooling

A more realistic alternative to the static Becker (1967) model is one in which young people make a series of decisions about whether to enrol in school (for example, Keane and Wolpin 1997, 2001). If they do, their education increments by an amount which may depend on effort and ability, and they then become eligible to enter a higher level of schooling the next period. Individuals also choose a level of savings or borrowing which can depend on tuition costs, earnings, family transfers, and access to loans and grants. This class of models sheds light on a number of features that are inconsistent with (or simply ignored by) a static framework. For example, a dynamic model can be used to formally address the question of how students learn about their potential returns to different levels of schooling (Arcidiacono 2004), and how schooling choices are affected by risk aversion and access to credit markets (Keane and Wolpin 2001).

A dynamic framework is also helpful for understanding the distribution of observed education choices in the presence of 'sheepskin' or 'degree' effects that create non-concavities in the earnings–schooling relationship. In the United States, for example, people with three years of college education have about the same earnings as those with only two years of college (Park 1994). (Likewise, people with three years of high school earn about the same as those with only two years of high school; see Hungerford and Solon 1987.) From a static modelling perspective it is unclear why anyone would ever plan to leave college after three years. From a dynamic perspective, however, the outcome of three years of college can be explained by noting that the true return to the third year of college is the option value of entering the fourth year (Altonji 1993). Students begin their third year of college knowing it is a necessary step to graduation, but may receive some information – for example, about their ability to complete the programme – that causes them to re-evaluate the costs and benefits of enrolment and drop out without graduating.

A dynamic perspective suggests that one should calculate the distribution of final education outcomes conditional on starting a specific education programme, and use this distribution, in combination with the estimated costs and earnings for each outcome, to measure the *ex ante* return to programme entry. (In fact, such a calculation is explicitly built into dynamic optimization models like the ones estimated by Keane and Wolpin 1997 and Eckstein and Wolpin 1999.) An interesting case in point is entry to a junior college, which has three main outcome possibilities: early dropout, completion of an Associates (AA) degree, or entry (with two years of college credit) to a four-year college programme. The third node creates an option value to entering an academic programme at junior college that is ignored in simple *ex post* comparisons of earnings between those who are observed holding an AA degree and those with only high school education.

Evidence on the Returns to Schooling

Mincerian Studies

Most of the existing evidence on the returns to schooling is based on Mincer's (1974) 'human capital earnings function': an OLS regression of log earnings on years of completed schooling and a polynomial of post-schooling experience (that is, current age minus an estimate of age at the completion of schooling). As noted in section "The Internal Rate of Return and Equalizing Differences", under certain simplifying assumptions the coefficient of schooling can be interpreted as an estimate of the internal rate of return to alternative education choices. Though the empirical validity of these assumptions varies from application to application, Mincer's model has fitted in hundreds of studies of earnings determination around the world.

Several issues arise in the specification of the human capital earnings function (HCEF) that affect the magnitude of the estimated returns to schooling. One is the choice of earnings measure. Since better-educated people tend to work more hours per week and weeks per year than those with less education, the estimated returns to schooling are usually larger for annual earnings than for weekly or hourly earnings (Card 1999).

Arguably, earnings should also include the cash value of work-related benefits like health insurance and pensions, leading to an additional source of 'returns' to schooling. A related issue is the treatment of taxes and transfer income during periods of nonwork, for example, from unemployment insurance and welfare programmes. From the perspective of an individual investor, the return to a given schooling choice presumably depends on the expected net incomes associated with the choice (that is, taking into account expected taxes and transfers). Interestingly, the earnings measures available in conventional surveys for many European countries (for example, France and Spain) are net of social security and income taxes, whereas the earnings measures available for other countries (in particular the United States) exclude taxes. Thus, there is some adjustment for taxes built-in to conventional human capital earnings functions estimated for many European countries, but not for the United States. Finally, schooling may affect longevity or health, leading to another indirect effect on earnings.

A second issue is functional form. Mincer's equalizing differences framework implies that log earnings are related to the opportunity cost of a given schooling choice, measured in years of forgone earnings, plus an additive term in years of post-schooling experience. Mincer (1974) assumed a linear path for on-the-job investments in human capital after the completion of schooling and showed that earnings would then depend on a *quadratic* function of years of post-schooling experience. Unless the assumptions underlying this derivation are correct, however, the conditional expectation of earnings, given education and age, will differ from this highly restrictive functional form. Empirically, the model adopted by Mincer (1974) is probably too restrictive (Lemieux 2006). For example, Murphy and Welch (1990) conclude that a model with a third- or fourth-order polynomial in experience provides a significant improvement in fit.

Researchers have generalized the HCEF by including dummies for degrees (or a complete set of dummies for all possible schooling choices), by including interactions between schooling and experience (or cohort), and by including interactions between schooling and characteristics like gender, cognitive ability, family background, and school quality. Estimation results from such models can be used to calculate 'returns' to schooling that vary by the level or type of schooling and by individual characteristics. Although the resulting estimates cannot be strictly interpreted as internal rates of return, it is conventional to refer to the implied marginal effects as *returns to schooling*.

Related to the issue of functional form is the question of whether post schooling choices – like occupation or industry – should be added as controls to HCEF. From the perspective of calculating the returns to alternative schooling choices, the answer is 'no', since some of the return to additional schooling is the increased chance of working in a more highly paid occupation or industry (Becker 1964). A more subtle issue is region or urban location, since some part of the wage differential associated with these choices is caused by differences in the cost of living (which presumably should be subtracted from earnings to calculate the return to schooling).

Recent surveys of the returns to schooling based on the Mincerian HCEF (Psacharopoulos 1994; Psacharopoulos and Patrinos 2004; Harmon et al. 2001, 2003) suggest that returns are in the range of 5–15 per cent for most OECD countries, and somewhat higher in developing countries, on average. In Europe, returns appear to be relatively low in Scandinavia (around 5 per cent) and relatively high in the United Kingdom and Ireland (10 per cent or more). Estimated returns in the United States are comparable to those in the United Kingdom, with evidence of a positive trend in both countries over the 1980–2000 period (Katz and Murphy 1992; Card and Lemieux 2001; Gosling and Lemieux 2003). Using meta-analytic techniques, Harmon et al. (2001) conclude that estimated returns are on average 1–2 points lower when the sample is limited to the public sector, when the earnings model includes controls for occupation or 'ability' measures, and when allowances are made for taxes. They also conclude that returns are slightly higher for women than men. In the United States, estimated returns for the mid-1990s from a conventional HCEF based on hourly earnings were about 10 per cent for men and 11 per cent for women (Card 1999, Table 1).

Causal Studies

In his pioneering study Walsh noted: 'No doubt the students who go on from high school to college are, on average, richer in natural endowments than those who are left behind. They are a selected lot . . .' (Walsh 1935, pp. 272–3). Two main methods have been developed to control for the potential selection biases that confound simple earnings comparisons between people with different levels of schooling: (1) comparisons of siblings or twins; (2) comparisons based on interventions or exogenous factors that affected the education choices of one group relative to another. Detailed discussions of these methods are presented in Card (1995a, 1999, 2001), Krueger and Lindahl (2001), Harmon et al. (2003), and Blundell et al. (2004). This section presents a brief overview of some of the main methodological issues – and some of the associated findings – without attempting a comprehensive review.

Gorseline (1932) first proposed the use of sibling comparisons to control for selection biases between different education groups. The basic idea can be illustrated using a variant of the 'homogeneous returns' model discussed in section "An Extended Model". Letting y_{ij} and S_i denote the earnings and schooling of sibling j ($j = 1,2$) from family i, the homogeneous returns model posits:

$$\log y_{ij} = \alpha_{ij} + \overline{b} S_{ij},$$

where α_{ij} represents the level of earnings that sibling j would receive in the absence of schooling. One possible assumption is that $\alpha_{i1} = \alpha_{i2}$: that is, that the two siblings have equal 'ability'. In this case, one can obtain an unbiased estimate of the true return to schooling from a within-family regression, since

$$\log y_{i1} - \log y_{i2} = \overline{b}(S_{i1} - S_{i2}).$$

Chamberlain and Griliches (1975) re-analysed Gorseline's sample of Indiana brothers and obtained a within-family estimate of $\overline{b}$ equal to 0.080 – only slightly below the estimate of 0.082 obtained from a conventional earnings model estimated by OLS on the same data. (Chamberlain and Griliches 1975, also included the sibling's differences in age and age-squared as added regressors.) Of course siblings may not have identical abilities, and if they don't, a within-family estimator b_w can be worse (that is, more biased) than the corresponding OLS estimator b_{ols}. Assuming a homogenous returns model is correct, the bias in the OLS estimate is

R

$$\text{plim } b_{ols} - \overline{b} = \text{cov}\left[\alpha_{ij}, S_{ij}\right]/var\left[S_{ij}\right],$$

while the bias in the within-family estimator is

$$\text{plim } b_w - \overline{b} = \begin{array}{l}\text{cov}\left[\alpha_{ij} - \alpha_{i2}, S_{i1} - S_{i2}\right] \\ /\text{var}[S_{i1} - S_{i2}].\end{array}$$

Although differencing eliminates the shared component of α_{i1} and α_{i2}, it is possible that the remaining within-family difference in ability is large relative to the within-family variance of schooling, implying a larger bias in b_w than b_{ols}. (A similar analysis can be conducted when both the slope and intercept of the earnings function have person-specific components; see Card 1999.)

One approach to the concern over ability differences between siblings is to focus on identical (monozygotic) twins. Unfortunately, schooling differences are small among identical twins, and even a little measurement error in reported education can lead to large attenuation bias in the within-twin estimate of the return to schooling. Ashenfelter and Krueger (1994) proposed an innovative solution based on asking each twin about its own and its sibling's education. Their method has been widely adapted in the literature (for example, Ashenfelter and Rouse 1998; Miller et al. 1995; Bonjour et al. 2003) and leads to estimated returns within twins that are comparable to the corresponding OLS estimates, or only slightly smaller. (As noted by Bound and Solon 1999, if the measurement error in schooling is mean-reverting, the Ashenfelter–Krueger approach 'over-corrects' and leads to an upward bias in the resulting estimator.)

Despite the intuitive appeal of identical twins to some researchers, others (for example, Bound and Solon 1999) have questioned whether twins who choose different schooling levels are really 'identical' or whether the small differences in upbringing and experience that lead them to choose different schooling also contribute to heir different earnings. Fundamentally, the problem is that the source of the differential schooling choices is unobserved, so different observers can argue that the choice was driven by factors that are either correlated or uncorrelated with earnings. A similar problem arises in 'matching' estimates of the return to schooling (see Blundell et al. 2004), which attempt to compare earnings between people who are very similar in all dimensions except their choice of schooling. Indeed, a perfect matching algorithm applied to a sample of twins would presumably match twins to each other, leading to a within-family estimate of the return to schooling.

A second approach to the issue of selection bias is the use of instrumental variables (IV) methods. Specifically, the researcher posits the existence of a variable Z that is exogenous to individuals but affects their schooling choices. As shown by Heckman (1978) in a different context, the equation relating S to Z need not represent a well-specified model, only a linear projection. For example, assume:

$$S_i = Z_i\pi + \xi_i.$$

If earnings are generated by the homogeneous returns model:

$$\log y_i = \alpha_i + \overline{b}S_i,$$

and Z_i is orthogonal to α_i, then a consistent estimate of $\overline{b}$ can be obtained by IV, using Z_i as an instrument for S_i. Individual-level instruments that have been proposed include quarter of birth (Angrist and Krueger 1991), the sex composition of one's siblings (Butcher and Case 1994), and distance to the nearest college (Card 1995b). Other IV studies use school system reforms such as changes in the minimum school-leaving age (Harmon and Walker 1995; Oreopoulos 2006; Meghir and Palme 2005), changes in tuition at local state colleges (Kane and Rouse 1993; Fortin 2006), and expansions in local infrastructure (Duflo 2001).

Many IV studies yield estimated returns to schooling that are as large as or slightly larger than the corresponding IV estimates (see for example, Card 2001; Harmon et al. 2003). Since the IV approach was motivated by the concern that OLS leads to an *overestimate* of the returns to schooling, this is potentially puzzling, and three explanations have been offered. First, OLS estimates are downward-biased by measurement error in education, and the measurement error bias may offset any upward selectivity bias (Griliches 1977). Second, the search for IV designs that yield statistically significant estimates may create a 'publication bias' in favour of samples and specifications with relatively large IV coefficients (Ashenfelter et al. 1999). Third, if the returns to education vary across the population, certain instruments may identify returns for subgroups with relatively high marginal returns to schooling (Card 1995a).

The third explanation can be most easily understood in the context of a social experiment with a randomly assigned intervention (indexed by $Z_i = 1$). Let (S_{i0}, y_{i0}) represent the schooling and earnings outcomes for person i if he or she were assigned to the control group, and (S_{i1}, y_{i1}) denote the outcomes if he or she was assigned to treatment (note that only one of these pairs is observed). The treatment effect on schooling for person i is $\Delta S_i = S_{i1} - S_{i0}$, while the effect on log earnings is $\Delta \log y_i = \log y_{i1} - \log y_{i0}$. Assuming that individual i's marginal return to schooling in the absence of the intervention is β_i, and that the intervention only affects earnings through its effect on schooling, $\Delta \log y_i = \beta_i \Delta S_i$. An IV estimate of the return to schooling based on assignment status is numerically equal to the difference in mean log earnings between the treatment and control groups, divided by the corresponding difference in their average schooling, and has probability limit

$$\begin{aligned} \text{plim } b_{IV} &= \frac{E[\log y_i | Z_i = 1] - E[\log y_i | Z_i = 0]}{E[S_i | Z_i = 1] - E[S_i | Z_i = 0]} \\ &= \frac{E[\beta_i \Delta S_i]}{E[\Delta S_i]}. \end{aligned}$$

If $E[\beta_i \Delta S_i] = E[\beta_i] E[\Delta S_i]$, then the IV estimator gives a consistent estimate of the average marginal return to education $\overline{\beta} = E[\beta_i]$. This will be true if the intervention induces the same change in schooling for everyone, or more generally if $E[\Delta S_i | \beta_i]$ is independent of β_i. Otherwise, provided that $\Delta S_i \geq 0$ for all i (that is, no one reduces schooling because of the intervention) the IV estimate is a weighted average of the β_i's, with the weight for person i equal to $\Delta S_i / E[\Delta S_i]$.

An intervention that induces larger gains in schooling for people with high values of β_i can lead to an IV estimate that overstates $\overline{\beta}$. Card (1995a) argued that this might be true for interventions – like the increases in the minimum school leaving age studied by Harmon and Walker (1995) and Oreopoulos (2006) – that mainly affect children from disadvantaged family backgrounds who stop school early because of high marginal costs rather than because of low marginal benefits. An alternative explanation for the finding that an IV estimate exceeds the corresponding OLS estimate is that the assumptions underlying the particular instrumental variable are invalid. In particular, in the absence of a true experiment, one can never 'prove' that the instrument is as good as randomly assigned. Even in an experimental setting, it is also possible that an intervention has an independent causal effect on earnings, confounding the interpretation of the IV estimate. (For example, Willis and Rosen 1979, and Heckman and Li 2004, use parental education as instruments for schooling, though others have argued that parental education has an independent effect on earnings.)

A generalization of IV that is useful when a researcher believes there may be random payoffs to schooling is a control function approach, first used in the schooling context by Garen (1984). (Other recent applications include Conneely and Uusitalo 1997, Blundell et al. 2004, and Rau-Binder 2006.) This method relies on assumptions about the relationship between the error component in the equation relating schooling to the instrument(s) Z, and the random slope and intercept in the earnings equation. Assuming these are satisfied, a control function approach can recover unbiased estimates of the average marginal return to schooling, as well as useful information on how the returns to education vary with the unobserved factors driving the choice of schooling (Rau-Binder 2006).

Summary

The idea of treating schooling as an investment that yields internal rates of return comparable to other investments in the economy has proven extremely useful, and has led to an unusually coherent body of research that combines theoretical modelling and detailed empirical analysis. Much of the existing empirical work is conducted in the framework of Mincer's (1974) human capital earnings function, which relates the logarithm of earnings to completed schooling – measured in years to reflect the opportunity cost of the investment – and a control for post-schooling experience. In a strict equalizing differences framework the coefficient of schooling is the internal rate of return to schooling. In a more general framework that recognizes the endogenous nature of the schooling decision, and the importance of ability differences that partially determine the choice of schooling, observed differences in earnings across different education groups will not necessarily reveal the rate of return to schooling for any one person, or for the population as a whole. Nevertheless, existing evidence from studies of siblings and twins, and from studies that focus on arguably exogenous sources of variation in education choices, suggest that the return to schooling is in the range of 5–15 per cent, and not too different from the value implied by the simple Mincerian approach.

See Also

- ▶ Becker, Gary S. (Born 1930).
- ▶ Control Functions
- ▶ Griliches, Zvi (1930–1999)
- ▶ Human Capital
- ▶ Mincer, Jacob (1922–2006)
- ▶ Rosen, Sherwin (1938–2001)
- ▶ Schultz, T. W. (1902–1998)
- ▶ Selection Bias and Self-Selection

Bibliography

Altonji, J.J. 1993. The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics* 11: 48–83.

Angrist, J.D., and A.B.. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106: 979–1014.

Arcidiacono, P. 2004. Ability sorting and the returns to college major. *Journal of Econometrics* 121: 343–375.

Ashenfelter, O., and A.B.. Krueger. 1994. Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84: 1157–1173.

Ashenfelter, O., and C. Rouse. 1998. Income, schooling, and ability: Evidence from a new sample of twins. *Quarterly Journal of Economics* 113: 869–895.

Ashenfelter, O., C. Harmon, and H. Oosterbeek. 1999. A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics* 6: 453–470.

Becker, G.S. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70: 9–49.

Becker, G.S. 1964. *Human capital: A theoretical and empirical analysis, with special reference to education*. New York: Columbia University Press.

Becker, G.S. 1967. *Human capital and the personal distribution of income*. Ann Arbor: University of Michigan Press.

Blundell, R., Dearden, L. and Sianesi, B. 2004. Evaluating the impact of education on earnings in the U.K.: Models, methods, and results from the NCDS. Working Paper No. 03/20. Institute for Fiscal Studies.

Bonjour, D., L. Cherkas, J. Haskel, D. Hawkes, and T. Spector. 2003. Returns to education: Evidence from UK twins. *American Economic Review* 93: 1799–1812.

Bound, J., and G. Solon. 1999. Double trouble: On the value of twins-based estimation of the returns to education. *Economics of Education Review* 18: 169–182.

Butcher, K.F., and A. Case. 1994. The effect of sibling composition on women's education and earnings. *Quarterly Journal of Economics* 109: 531–563.

Card, D. 1995a. Earnings, schooling, and ability revisited. In *Research in labor economics*, ed. S. Polachek, vol. 14. Greenwich: JAI Press.

Card, D. 1995b. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of labour market behaviour: Essays in honour of John Vanderkamp*, ed. L.N. Christofides, E.K. Grant, and R. Swidinsky. Toronto: University of Toronto Press.

Card, D. 1999. The causal effect of education on earnings. In *Handbook of labor economics*, ed. O. Ashenfelter and D. Card, vol. 3A. Amsterdam: North-Holland.

Card, D. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69: 1127–1160.

Card, D., and T. Lemieux. 2001. Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *Quarterly Journal of Economics* 116: 705–746.

Chamberlain, G., and Z. Griliches. 1975. Unobservables with a variance–covariance structure: Ability, schooling, and the economic success of brothers. *International Economic Review* 16: 422–449.

Conneely, K. and Uusitalo, R. 1997. Estimating heterogeneous treatment effects in the Becker schooling model. Discussion Paper, Industrial Relations Section, Princeton University.

Cunha, F. and Heckman, J.J. 2006. Identifying and estimating the distributions of *ex post* and *ex ante* returns to schooling: A survey of recent developments. Working Paper, University of Chicago.

Duflo, E. 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* 91: 795–813.

Eckstein, Z., and K.I. Wolpin. 1999. Why youths drop out of high school: The impact of preferences, opportunities and abilities. *Econometrica* 67: 1295–1339.

Fortin, N. 2006. Higher education policies and the college premium: Cross-state evidence from the 1990s. *American Economic Review* 96: 959–987.

Freeman, R.B. 1976. *The over-educated American*. New York: Academic Press.

Garen, J. 1984. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica* 52: 1199–1218.

Gorseline, D.E. 1932. *The effect of schooling upon income*. Bloomington: University of Indiana Press.

Gosling, A., and T. Lemieux. 2003. Labor market reforms and changes in wage inequality in the United Kingdom and the United States. In *Seeking a premier economy: The economic effects of British economic reforms, 1980–2000*, ed. D. Card, R. Blundell, and R.B. Freeman. Chicago: University of Chicago Press.

Griliches, Z. 1977. Estimating the returns to schooling: Some econometric problems. *Econometrica* 45: 1–22.

Harmon, C., and I. Walker. 1995. Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* 85: 1278–1286.

Harmon, C., I. Walker, and N. Westergaard-Nielson. 2001. Introduction. In *Education and earnings in Europe*, ed. C. Harmon, I. Walker, and N. Westergaard-Nielson. Aldershot: Edward Elgar.

Harmon, C., I. Walker, and N. Westergaard-Nielson. 2003a. The returns to education: Microeconomics. *Journal of Economic Surveys* 17: 115–156.

Harmon, C., H. Oosterbeek, and I. Walker. 2003b. The returns to education: Microeconomics. *Journal of Economic Surveys* 17: 115–156.

Heckman, J.J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959.

Heckman, J.J., and X. Li. 2004. Selection bias, comparative advantage, and heterogeneous returns to education: Evidence from China in 2000. *Pacific Economic Review* 9: 155–171.

Hungerford, T., and G. Solon. 1987. Sheepskin effects in the return to education. *Review of Economics and Statistics* 69: 175–177.

Kane, T.J. and Rouse, C.E. 1993. Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter? Working Paper No. 4268. Cambridge: NBER.

Katz, L.F., and K.M. Murphy. 1992. Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics* 107: 35–78.

Keane, M.P., and K.I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105: 473–521.

Keane, M.P., and K.I. Wolpin. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review* 42: 1051–1103.

Krueger, A.B.., and M. Lindahl. 2001. Education for growth: Why and for whom? *Journal of Economic Literature* 39: 1101–1136.

Lemieux, T. 2006. The ‘Mincer equation’ thirty years after *Schooling, experience, and earnings*. In *Jacob Mincer: A pioneer of modern labor economics*, ed. S. Grossbard-Shechtman. Heidelberg: Springer.

Meghir, C., and M. Palme. 2005. Educational reform, ability and parental background. *American Economic Review* 95: 414–424.

Miller, P.W., C. Mulvey, and N. Martin. 1995. What do twins studies reveal about the economic returns to education? A comparison of Australian and US findings. *American Economic Review* 85: 586–599.

Mincer, J. 1958. Investment in human capital and personal income distribution. *Journal of Political Economy* 66: 281–302.

Mincer, J. 1974. *Schooling, experience, and earnings*. New York: NBER (distributed by Columbia University Press).

Mincer, J. 1997. Changes in wage inequality, 1970–1990. *Research in Labor Economics* 16: 1–18.

Murphy, K.M., and F. Welch. 1990. Empirical age–earnings profiles. *Journal of Labor Economics* 8: 202–229.

Oreopoulos, P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* 96: 152–175.

Park, J.H. 1994. Returns to schooling: A peculiar deviation from linearity. Working Paper No. 335. Industrial Relations Section, Princeton University.

Psacharopoulos, G. 1994. Returns to investment in education: A global update. *World Development* 22: 1325–1343.

Psacharopoulos, G., and H.A. Patrinos. 2004. Returns to investment in education: A further update. *Education Economics* 12: 111–134.

Rau-Binder, T. 2006. Semiparametric estimation of microeconometric models with endogenous regressors and sorting. Manuscript, Department of Economics, University of California Berkeley.

Roy, A.D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146.

Solon, G. 1999. Intergenerational mobility in the labor market. In *Handbook of labor economics*, ed. O.-C. Ashenfelter, D. Card, and R. Layard, vol. 3-A. Amsterdam: North-Holland.

Walsh, J.R. 1935. Capital concept applied to man. *Quarterly Journal of Economics* 49: 255–285.

Willis, R. 1986. Wage determination: A survey and reinterpretation of human capital earnings functions. In

Handbook of labor economics, ed. O. Ashenfelter and R. Layard, vol. 1. Amsterdam: North-Holland.
Willis, R., and S. Rosen. 1979. Education and self-selection. *Journal of Political Economy* 87: S7–S36.

Revealed Preference Theory

Marcel K. Richter

JEL Classifications
D1

Economists do not observe preferences. They may, however, observe demand behaviour – the choices made by consumers. Is there a way for economists to tell whether the observed behaviour is generated through the maximization of a preference relation or utility function? Since most economic theories are ultimately based on a consumer who maximizes a preference or utility, the question is clearly important for developing and testing theories.

Revealed preference theory answers this question by characterizing choice behaviour that is generated by preference or utility maximization. Relating choice behaviour and preference maximization is also a goal of integrability theory. What distinguishes the theories from each other, and from the other parts of rationality theory, is the special nature of their tools: integrability theory uses mathematical integration in its proofs, and usually states its hypotheses in differential form; revealed preference theory uses a variety of mathematical tools for its proofs, and its hypotheses are usually in a discrete 'revelation' form. The distinctions are not always sharp, however, and we shall see areas in which the theories overlap.

Samuelson invented revealed preference theory in 1938. The basic idea, much of the terminology, and some of the axioms are due to him. In the following outline, a useful paradigm is the one that guided the first three decades: a consumer with a finite-dimensional euclidean commodity space, facing 'competitive' budgets determined by fixed positive prices, and satisfying a budget equality constraint.

The Problem of Rationality

From the economist's point of view, unobservable preferences generate observable choices. Since many preference relations may generate the same choice correspondence, the map from preferences to choice correspondences is many-one: We cannot hope to find *the* preference generating choices, but only *some* preferences – a set of 'equivalent' preferences. For example:

It is well known in preference theory that a lexicographic preference on the plane does not admit a real-valued utility function (Debreu 1954). A hasty conclusion might be that there is no hope of representing a lexicographic-maximizing consumer as a utility-maximizing consumer. Too hasty! For her behaviour clearly maximizes this function g on the non-negative plane (for positive prices):

$$g(x_1, x_2) = x_1 \qquad (1)$$

Even if her 'intention' is to maximize a lexicographic preference, she acts *as if* her intention were to maximize g. In fact, even if the choices were made by a committee, a machine, or any other mindless decision maker, we can still say the actions are *as if* the intent were g-maximization.

This example shows the distinction between a typical question in utility and preference theory ('Does *this* preference have a utility function?'), and the basic question in revealed preference and integrability theories ('Is this demand generated by *some* preference?'). It also demonstrates the need for precise definitions (Our notation will follow the glossaries of Richter 1966, 1971).

To describe choices, the theory requires an underlying set X and a family $\mathcal{B}$ of subsets $B \subset X$ (Often X is the non-negative orthant of n-space and each B is a 'competitive' budget determined by positive prices and income). We call any $B \in \mathcal{B}$ a *budget*. A *choice* or *demand correspondence* h is a function assigning to each

$B \in \mathcal{B}$ a subset $h(B) \subset B$ interpreted as the set of elements chosen from B. And any binary relation on X is called a *preference.* Rationality theory relates choices h on $(X, \mathcal{B})$ to preferences R on X in two ways.

(i) If we start with a preference relation R we can ask what kind of choice it generates. There are two obvious senses in which it could generate a choice h. First, we might have, for all $B \in \mathcal{B}$

$$h(B) = \left\{x \in B : \forall y_{y \in B} xRy\right\}, \tag{2}$$

i.e., the set of elements chosen from B is the set of R-most preferred elements in B. Then we say that R *rationalizes* h (Richter 1971).

Alternatively, we might have, for all $B \in \mathcal{B}$

$$h(B) = \left\{x \in B : \forall y_{y \in B} yRx\right\}, \tag{3}$$

i.e., the set of elements chosen from B is the set of elements in B for which nothing in B is R-more preferred. Then we say that R *motivates* h (Kim and Richter 1986).

Definition (2) is appropriate if we think of R as a 'weak' (i.e. reflexive) relation, while (3) is appropriate if we think of R as a 'strict' (asymmetric) relation.

(ii) Conversely, if we start with a choice h we can ask whether any preference R generates, or 'explains' h. If there exists some R generating h in the sense of (2) (Richter 1966, 1971), then we say that h is *rational.* Often we are interested in *reflexive*-rationality (rationalization by a reflexive preference), *transitive*-rationality (rationalization by a transitive preference), *regular*-rationality (rationalization by a reflexive, transitive, and total preference), etc. For example, *utility–rationality* requires the existence of a function $f: X \to R^1$ satisfying

$$h(B) = \left\{x \in B : \forall y_{y \in B} f(x) \geqq f(y)\right\} \tag{4}$$

for all $B \in \mathcal{B}$ – i.e., the set of elements chosen from B is the set of those elements in B with the highest utility.

If there exists some R generating h in the sense of (3), then we say that h is *motivated* (Kim and Richter 1986). Again, we are often interested in *asymmetric*- motivation (motivation by an asymmetric preference), etc. In fact, h is rational if and only if it is motivated (Clark 1985; Kim and Richter 1986). Of course, the example (1) makes it clear that such a rationalizing or motivating R will not usually be unique: there will be a whole equivalence class of such relations generating the same choice (Kim and Richter 1986).

It is important to note that rationality and motivation have been defined as properties of demand, not of preference. We do not say, for example, that a particular preference is rational or irrational. Instead, the definitions relate demand and preferences.

An economist who derives comparative statics results from preference maximization is answering qsts of type (i). The issue arises – for both theoretical development and empirical testing – whether any further results can be derived, or whether all the (independent) consequences of preference maximization have been found. This is usually a much more difficult problem. A major task of both revealed preference and integrability theory is to address this issue, by answering qsts of type (ii). The two qsts are parts, then, of the fundamental Problem of Rationality: give necessary (i) and sufficient (ii) conditions for a demand to be rational (of a particular type), or motivated (of a particular type). Revealed preference theory solves the problem through axioms with a unique flavour.

Revealed Preference Solutions

It is important to distinguish revealed preference definitions from revealed preference axioms, and these in turn from revealed preference theorems.

Revelation Definitions

If consumer (i.e. choice) h selects alternative $x \in B$ from B – i.e., if $x \in h(B)$ – when alternative y could have been selected – i.e., if $y \in B$ – then we write xVy. And it is natural to say that x is *revealed as good as* y. If also $x \neq y$ then we write xSy, and it is natural to say that x is *revealed preferred to* y. This terminology of Samuelson's is very suggestive, because if $\succcurlyeq$ is any rationalization, then xVy implies $xx \succcurlyeq y$ as does xSy. In fact, if $x \in h(B)$ & $y \in (B)h(B)$ and if $\succcurlyeq$ is regular, then its asymmetric part $\succ$ also satisfies $x \succ y$. So an observer of h can deduce properties common to all rationalizations. But beware: xSy is a statement about *choice*, not about a particular preference.

Unlike the psychologist, who may be able to present an individual with binary choices, and thereby uncover a total ordering, the economist will typically observe S as only a partial ordering. This is one of the challenging features of revealed preference theory. It is why, mathematically, revealed preference theory is a study of partial orders, in contrast to the classical theory of preference, which is a theory of total orders. It is also why there is generally more than one preference in the equivalence class of preferences that rationalize or motivate a given choice.

Revelation Axioms

We describe four revealed preference axioms. Samuelson proposed the *asymmetry* of S as a basic axiom of consumer theory: for all $x, y \in X$

$$xSy \Rightarrow ySx. \tag{5}$$

In other words, if x is revealed preferred to y (under some budget), then y is never (under any budget) revealed preferred to x. As Samuelson noted, this is a property of any single-valued demand function maximizing a regular preference. This is now called the *Weak Axiom of Revealed Preference.*

Houthakker noted other necessary consequences of regular-rationality, for single-valued demand functions: there can be no cycles of the form

$$xSy_1Sy_2S, \quad \ldots, \quad Sy_kSx. \tag{6}$$

In other words, x is never, even indirectly, revealed preferred to itself. Houthakker proposed this as a new axiom, now called the *Strong Axiom of Revealed Preference*. If we define xHy to mean that xSy or $xSv_1Sv_2S, \ldots Sv_kSy$, then we can rephrase Houthakker's axiom as saying that H is asymmetric. In other words, if x is (even indirectly) revealed preferred to y, then y is never (even indirectly) revealed preferred to x.

Richter noted still another consequence of regular-rationality. For this it is convenient to define xWy to mean either xVy or $xVu_1\ V, \ldots Vu_kVy$. Clearly regular-rationality implies: for all $x, y \in X$ & $B \in \mathcal{B}$

$$x \in h(B) \& y \in B \& yWx \Rightarrow y \in h(B). \tag{7}$$

In other words, if x is chosen from B, and if y is also available in B and is revealed (even directly) as good as x, then y is also chosen from B. This is the *Congruence Axiom of Revealed Preference.* He also noted a behavioural consequence of any rationality: for all $x, y \in X$ & $B \in$

$$x \in B \forall y_{y \in B} xVy \Rightarrow x \in h(B). \tag{8}$$

In other words, if x is in B and is revealed as good as everything in B, then x is chosen from B. This is the *V-Axiom.*

We will use these axioms to discuss the main solutions to the Problem of Rationality.

Revelation Theorems

(a) *Weak Axiom.* Samuelson proposed the Weak Axiom in 1938, as a foundation for all consumer theory (1938a, b). He did not name it, but he suggested that (for single-valued demand functions) it followed from maximizing a utility function (cf. also Samuelson (1955, pp. 110–11). In the opposite direction, his idea of founding consumer theory on it was implicitly a conjecture that it implied utility-rationality, or at least regular-rationality. Indeed, after preliminary work by I.M.D. Little, Samuelson succeeded in showing that, for two commodities and Lipschitz-continuous demand functions, the Weak Axiom implied regular-rationality (Samuelson 1948).

(b) *Strong Axiom.* Then in 1950 Houthakker (1950) proposed the Strong Axiom (by a different name) as a basis for consumer theory, and showed that, for any number of commodities, it implied utility-rationality for Lipschitz-continuous demand functions. Samuelson (1950) then gave the Weak and Strong Axioms their modern names.

In 1959, Uzawa (1960, 1971) developed a more precise analogue of Houthakker's result, showing that the Strong Axiom and a Lipschitzian hypothesis on the demand implied irreflexive-transitive-monotone-convex-lower semi continuous-motivation. His proof was along the lines of the Samuelson–Little–Samuelson–-Houthakker analytic methods.

Although the Strong Axiom implied the Weak, it was still not clear whether the Weak implied the Strong. Indeed, Rose (1958) showed that the Weak Axiom does imply the Strong Axiom, when there are only two commodities and prices are positive (needed!). Then Gale (1960) constructed an example with three commodities, showing that the Weak Axiom did not imply the Strong. And Kihlstrom et al. (1976) showed how to obtain very easily many examples, for any number of commodities greater than two. And Shafer (1977b), affirming a conjecture of Samuelson (1953), showed that the full strength of the Strong Axiom is needed: even for three goods, there is no upper bound on the length of S-cycles that must be ruled out. In the opposite direction, several authors have discussed special conditions under which the Weak Axiom does imply the Strong (Arrow 1959; Uzawa 1960, 1971).

Richter (1966) used set-theoretic methods – very different from the analytic methods of Samuelson, Little, Houthakker and Uzawa – to simplify the proofs, eliminate extraneous assumptions, and strengthen the rationality results. In a framework of abstract budget spaces, and without the technical assumptions required by the earlier analytical approaches, he showed that the Strong Axiom is equivalent to regular-rationality for demand functions. Thus the Strong Axiom completely exhausts the theory of demand functions maximizing a regular preference.

Richter (1966) also showed that, if a competitive demand satisfies the Strong Axiom, then it is utility-rational if its range is well behaved, but it may not be utilityrational otherwise (Richter 1971).

Extensions

There have been many extensions. Richter (1966) showed that the V-Axiom characterized rationality, and the Congruence Axiom characterized regular-rationality, for demand correspondences (Hansson (1968) gave an alternative criterion for regular-rationality). Other extensions have obtained stronger properties of the rationalization under special hypotheses (Hurwicz and Richter 1971; Mas-Colell 1978, Theorem 1; Richter 1986; Matzkin and Richter 1986); uniqueness of the rationalization within certain classes (Mas-Colell 1977); revealed preference axioms characterizing more general rationality types (Richter 1971; Kim and Richter 1986; Kim 1987); dual axioms (Sakai 1977; Richter 1979); and axioms for stochastic rationality (McFadden and Richter 1970).

Applications

Several applications have supported Samuelson's original idea that revealed preference could provide an alternative to preference theory as a foundation for consumer theory. Revealed preference techniques have been applied to prove the existence of competitive equilibrium (Wald 1936, 1951); to prove the stability of competitive equilibrium (Arrow and Hurwicz 1958, 1960); to prove the Hicks Composite Commodity Theorem (Richter 1970; Calsamiglia 1978); to analyse and characterize aggregate excess demand functions (Debreu 1974; McFadden et al. 1974); to prove aggregation properties for correspondences (Shafer 1977a); to prove properties of measurable demand correspondences (Yamazaki 1984); to prove theorems about social choice functions (Plott 1973); etc.

Revealed Preference and Integrability

With the same rationality goal as revealed preference theory, integrability theory uses axioms on the Slutsky or Antonelli matrices to characterize rational choice (cf. Hurwicz 1971; see also integrability of demand). Under some smoothness assumptions on the demand function, the basic theorems state that symmetry and negative semidefiniteness of these matrices is necessary and sufficient for (upper- semicontinuous-) regular-rationality.

Samuelson established a link between revealed preference theory and integrability theory by showing that his Weak Axiom implied negative semidefiniteness of the matrices (Samuelson 1938b, 1955, pp. 111–14). Later Kihlstrom et al. (1976) demonstrated that negative semidefiniteness was equivalent to a Weak Weak Axiom.

This left open the question of finding a revealed preference axiom equivalent to the symmetry. The Strong Axiom was clearly too strong, since it already implied regular-rationality, and therefore both symmetry and negative-semidefiniteness. Then Hurwicz and Richter (1979a, b) showed that a differential axiom of Ville (1946, 1951) provided the exact strength needed. Although it does not even imply the Weak Axiom, it is similar in spirit to the Strong Axiom and can be given a revealed preference inpt. It thus serves, like Kihlstrom, Mas-Colell and Sonnenschein's Weak Weak Axiom, as a bridge between the Revealed Preference and Integrability approaches to consumer rationality. Richter (1979) discussed these bridges from the viewpoint of duality.

Other Notions of Rationality

Many economists have used notions of rationality different from Richter's notion (2).

Sometimes the term 'rational' has been applied to preference, rather than demand (In such applications it is often a synonym for 'transitive'). In Uzawa (1957) and Arrow (1959), on the other hand, it was applied to demand, but only in terms of axioms on demand behaviour. By contrast, (2) is applied to demand, but in terms that relate both demand and preferences.

Some economists have used weaker notions of rationality than (2), requiring only: for all $x, y \in X$ & $B \in \mathcal{B}$

$$h(B) \subset \left\{x \in B : \forall y_{y \in B} xRy\right\}. \tag{9}$$

In other words, every element chosen from B is R-most preferred in B, but B may contain other R-most preferred elements that are not chosen. We will call this *subsemi-rationality,* although it has often been referred to as rationality.

A drawback of this concept is its loose linkage of preference and demand. Any constant function, for example, satisfies (9). On the other hand, if one interprets $h(B)$ as a set of incomplete observations, then one might wonder whether, with more observations of choices from B, the set $h(B)$ of chosen elements might grow. Then one might want to find a preference R satisfying just (9), rather than insisting (as does (2)) that R explain *precisely* the observed set $h(B)$.

Afriat (1967) gave conditions on a demand function, over a finite set of budgets, that are necessary and sufficient for it to be subsemi-rationalized by a continuous monotone concave function. His work was clarified by Diewert (1973) who gave a criterion for continuous-monotone-concave-subsemirationality in terms of a linear programming problem. Varian (1983) restated Afriat's finite-budgets result in terms of a Generalized Axiom of Revealed Preference – weaker than the Strong Axiom.

Matzkin and Richter (1986) obtained full rationality by replacing the Generalized Axiom with the Strong Axiom, which they proved was necessary and sufficient for continuous-monotone-strictly concave-utility-rationality in the finite case. No revealed preference criterion for concave-regular-rationality is known for the not-necessarily finite case.

See Also

▶ Demand Theory
▶ Integrability of Demand
▶ Samuelson, Paul Anthony (1915–2009)

Bibliography

Afriat, S.N. 1967. The construction of utility functions from expenditure data. *International Economic Review* 8: 67–77.

Arrow, K.J. 1959. Rational choice functions and orderings. *Economica* N.S. 26: 121–127.

Arrow, K.J., and L. Hurwicz. 1958. On the stability of competitive equilibrium, I. *Econometrica* 26: 522–552.

Arrow, K.J., and L. Hurwicz. 1960. Some remarks on the equilibria of economic systems. *Econometrica* 28: 640–646.

Calsamiglia, X. 1978. Composite goods and revealed preference. *International Economic Review* 19: 395–404.

Clark, S.A. 1985. A complementary approach to the strong and weak axioms of revealed preference. *Econometrica* 53: 1459–1463.

Debreu, G. 1954. Representation of a preference ordering by a numbering function. In *Decision processes*, ed. R.M. Thrall, C.H. Coombs, and R.L. Davis, 159–165. New York: Wiley.

Debreu, G. 1974. Excess demand functions. *Journal of Mathematical Economics* 1: 15–21.

Diewert, W.E. 1973. Afriat and revealed preference theory. *Review of Economic Studies* 40: 419–425.

Gale, D. 1960. A note on revealed preference. *Economica* NS 27: 348–354.

Hansson, B. 1968. Choice structures and preference relations. *Synthese* 18: 443–458.

Houthakker, H.S. 1950. Revealed preference and the utility function. *Economica* NS 17: 159–174.

Hurwicz, L. 1971. On the problem of integrability of demand functions. In *Preferences, utility and demand*, ed. J.S. Chipman, L. Hurwicz, M.K. Richter, and H.F. Sonnenschein. New York: Harcourt, Brace, Jovanovich, ch. 9.

Hurwicz, L., and M.K. Richter 1971. Revealed preference without demand continuity assumptions. In *Preferences, utility and demand*, ed. J.S. Chipman, L. Hurwicz, M.K. Richter, and H.F. Sonnenschein. New York: Harcourt, Brace, Jovanovich, ch. 3.

Hurwicz, L., and M.K. Richter. 1979a. An integrability condition with applications to utility theory and thermodynamics. *Journal of Mathematical Economics* 6: 7–14.

Hurwicz, L., and M.K. Richter. 1979b. Ville axioms and consumer theory. *Econometrica* 47: 603–619.

Kihlstrom, R., A. Mas-Colell, and H. Sonnenschein. 1976. The demand theory of the weak axiom of revealed preference. *Econometrica* 44: 971–978.

Kim, T. 1987. Intransitive indifference and revealed preference. *Econometrica.*

Kim, T., and M. Richter. 1986. Nontransitive-nontotal consumer theory. *Journal of Economic Theory* 38: 324–363.

Mas-Colell, A. 1977. The recoverability of consumers' preferences from market demand behavior. *Econometrica* 45: 1409–1430.

Mas-Colell, A. 1978. On revealed preference analysis. *Review of Economic Studies* 45: 121–131.

Matzkin, R., and M.K. Richter. 1986. *Testing concave rationality.* Minneapolis: Department of Economics, University of Minnesota.

McFadden, D., and M.K. Richter. 1970. Stochastic rationality and revealed stochastic preference. Presented to the 1970 Winter Meetings of the Econometric Society.

McFadden, D., and M.K. Richter. 1988. Stochastic rationality and revealed stochastic preference. In *Uncertainty, preferences and optimality. Essays in honor of Leonid Hurwicz*, ed. J.S. Chipman, D. McFadden, and M.K. Richter. New York.

McFadden, D., A. Mas-Colell, R. Mantel, and M.K. Richter. 1974. A characterization of community excess demand functions. *Journal of Economic Theory* 9: 361–374.

Plott, C.R. 1973. Path independence, rationality, and social choice. *Econometrica* 41: 1075–1091.

Richter, M.K. 1966. Revealed preference theory. *Econometrica* 34: 635–645.

Richter, M.K. 1971. Rational choice. In *Preferences, utility, and demand*, ed. J.S. Chipman, L. Hurwicz, M.K. Richter, and H.F. Sonnenschein. New York: Harcourt, Brace, Jovanovich, ch. 2.

Richter, M.K. 1979. Duality and rationality. *Journal of Economic Theory* 20: 131–181.

Richter, M.K. 1986. *Continuous demand functions*. Minneapolis: Department of Economics, University of Minnesota.

Rose, H. 1958. Consistency of preference: The two-commodity case. *Review of Economic Studies* 25: 124–125.

Sakai, Y. 1977. Revealed favorability, indirect utility, and direct utility. *Journal of Economic Theory* 14: 113–129.

Samuelson, P.A. 1938a. A note on the pure theory of consumer's behaviour. *Economica* NS 5: 61–71.

Samuelson, P.A. 1938b. A note on the pure theory of consumer's behaviour: An addendum. *Economica* NS 5: 353–354.

Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Samuelson, P.A. 1948. Consumption theory in terms of revealed preference. *Economica* NS 15: 243–253.

Samuelson, P.A. 1950. The problem of integrability in utility theory. *Economica* NS 17: 355–385.

Samuelson, P.A. 1953. Consumption theorems in terms of overcompensation rather than indifference comparisons. *Economica* NS 20: 1–9.

Shafer, W.J. 1977a. Revealed preference and aggregation. *Econometrica* 45: 1173–1182.

Shafer, W.J. 1977b. Revealed preference cycles and the Slutsky matrix. *Journal of Economic Theory* 16: 293–309.

Uzawa, H. 1957. Note on preference and axioms of choice. *Annals of the Institute of Statistical Mathematics* 8: 35–40.

Uzawa, H. 1960. Preference and rational choice in the theory of consumption. In *Mathematical methods in the social*

sciences, 1959, ed. K.J. Arrow, S. Karlin, and P. Suppes. Stanford: Stanford University Press, ch. 9.
Uzawa, H. 1971. Preference and rational choice in the theory of consumption. In *Preferences, utility, and demand*, ed. J.S. Chipman, L. Hurwicz, M.K. Richter, and H.F. Sonnenschein. New York: Harcourt, Brace, Jovanovich, ch. 1.
Varian, H.R. 1983. Non-parametric tests of consumer behaviour. *Review of Economic Studies* 50: 99–110.
Ville, J. 1946. Sur les conditions d'existence d'une ophélimité totale et d'un indice du niveau des prix. *Annales de l'Université de Lyon* 9, Sec. A(3): 32–39.
Ville, J. 1951. The existence conditions of a total utility function. *Review of Economic Studies* 19: 123–128.
Wald, A. 1936. Über einige Gleichungssyteme der mathematischen ökonomie. *Zeitschrift für Nationalökonomie* 7: 637–670.
Wald, A. 1951. On some systems of equations of mathematical economics. *Econometrica* 19: 368–403.
Yamazaki, A. 1984. The critical set of a demand correspondence in the price space and the weak axiom of revealed preference. *Hitotsubashi Journal of Economics* 25: 137–144.

Revelation of Preferences

J. J. Laffont

Competitive rational consumers reveal their preferences through their market behaviour as was made clear by Samuelson's (1947) revealed preference approach and by the literature on demand theory. Any bundle of commodities less costly than his chosen bundle must be less appreciated by a rational consumer than his chosen bundle.

However, in various circumstances collective decision processes must be used to mitigate market failures (public goods, externalities etc.). To what extent these processes can truthfully elicit agents' preferences, i.e., overcome the decentralization of information, is the issue raised here.

Revelation

Consider an agent who has preferences, represented by a preordering parameterized by $\theta^i \in \Theta^i$, $R(\theta^i)$ over a set of social states A. It is not difficult to convince oneself that any mechanism, $x : \Theta^i \to A$ which induces truthful revelation by agent i is equivalent to giving to the agent a subset $B \subset A$ and letting him maximize over this set. The sufficiency is obvious; the necessity is shown by choosing

$$A = \underset{\theta^i \in \Theta^i}{\cup} x\left(\theta^i\right).$$

When I agents are present, a mechanism is a mapping $x : \Theta \equiv \Pi_{t=1}^{I} \Theta^i \to A$.

It is reasonable in most circumstances to assume that agent i does not know the characteristics θ^j of the other agents. The revelation of preferences is therefore imbedded necessarily in a game of imperfect information for which several solution concepts are possible. A game which induces truthful revelation of preferences is said to be incentive compatible.

Implementation

Consider a social choice function, i.e. a mapping $f : \Pi_{t=1}^{I} \mathrm{H}^i \to A$. A social choice function f is said to be *implementable* if there exist message spaces M^i, $i = 1, \ldots, I$, and an outcome function $g : \Pi_{t=1}^{I} M^i \to A$ for which the equilibrium messages $m^i(\theta^i)$ $i = 1, \ldots, I$ are such that:

$$g\left(m^1\left(\theta^1\right), \ldots, m^I\left(\theta^I\right)\right) \quad \forall \theta \in \Theta$$

The equilibrium messages depend on the chosen solution concept for the game of imperfect information. The strongest notion of implementation is implementation in dominant strategies. Then, m^i (θ^i) is the best message of agent i whatever the messages of the other agents, for any i. A weaker notion of implementation is Bayesian implementation. Consider common knowledge prior expectations $\Psi^i(\theta^{-i}/\theta^i)$ describing agent i's expectations about the other agents' characteristics

$$\theta^{-i} = \left(\theta^1, \ldots, \theta^{i-1}, \theta^{i+1}, \ldots, \theta^I\right), i = 1, \ldots, I.$$

For any $\theta^i \in \Theta^i, m^{*i}(\theta^i)$ is the best message for agent i in the sense of his expected utility computed by using his prior $\Psi^i(\theta^{-i}/\theta^i)$ and by assuming that the others are using the response functions $m^{*j}(\theta^j), j \neq i$. Then f is implementable in Bayesian equilibrium if there exists a Bayesian equilibrium such that $f(\theta) = g(m^{*1}(\theta^1),\ldots, m^{*I}(\theta^I))$.

There are many other notions of implementation.

Implementation in Dominant Strategy

The *revelation principle* says that any f which is implementable in a dominant strategy can be implemented by a mechanism in which messages are identified with characteristics spaces Θ^i – direct mechanisms – and truthful revelation is a dominant strategy equilibrium.

In other words, it is not useful to consider more complex mechanisms than revelation mechanisms. (This neglects problems due to multiple equilibria.) We will therefore concentrate in the sequel on direct revelation mechanisms.

A fundamental result due to Gibbard (1973) and Satterthwaite (1975) tells us that, for more than two states, when no a priori information is available about individuals' preorderings, the only deterministic social choice function implementable in dominant strategies are dictatorships. To obtain positive results we must either introduce a priori information or weaken the notion of incentive compatibility.

The Vickrey Auction and the Clarke–Groves Mechanisms

To fight non-competitive behaviour, Vickrey (1961) proposed an auction which has the remarkable property that each bidder should announce his true willingness to pay for the auctioned object as a dominant strategy. The auction gives the object to the agent who makes the highest bid, but the payment is only the second highest bid.

The solution to the Wicksell–Samuelson free rider problem of public goods provided by Clarke (1971) and Groves (1973) can be viewed as an adaptation of this result. Preferences for public goods are assumed to be restricted to the class of quasi-linear utility functions which permits to go away from the negative result of the Gibbard–Satterthwaite theorem.

Consider the simple case of a costless indivisible project ($d = 0$ or 1) and call v_i the willingness to pay of agent i, $i = 1, \ldots, I$. The Pareto optimal decision under perfect information is

$$d = 1 \Leftrightarrow \sum_{i=1}^{I} v^i \geq 0$$

The Clark mechanism chooses to realize the project if the sum of the answers $\Sigma_{i=1}^{I} w^i$ is positive and agent i must pay a transfer $\Sigma_{j \neq i} w^j$ if he is pivotal, i.e. his answer changes the sign of the sum. He must pay the cost he imposes on the rest of society, just as in the Vickrey auction, an agent must pay the cost he imposes on the society which is the second willingness to pay. Groves mechanisms are obtained by adding to the Clarke transfer of agent i an arbitrary function of the answers of the others.

The first best public project decision is implemented. However, the incentive compatible transfers do not sum to zero in general so that a Pareto optimal allocation is not achieved. This should not come as a surprise. The decentralization of information imposes a cost on allocation rules.

Preferences can be elicited but at the cost of some distortions in allocations rules.

Large Numbers

The problem of revelation of preferences for private goods is not a serious problem in large economies. Indeed, as a ‹negligible› agent cannot affect prices he cannot affect his budget set and therefore the competitive equilibrium is incentive compatible in dominant strategies.

With public goods the problem becomes more and more severe with the number of agents since everyone can hope to have the others finance the public good. Despite the fact that, as the number of agents increases, the imbalance of transfers in the Groves mechanisms can be made negligible in various senses, the question of the strength of incentives must be raised in such circumstances.

Historical Note

The free rider problem was recognized by Wicksell (1896) and emphasized by Samuelson (1954). The positive results by Groves (1973) and Clarke (1971) and by Aspremont and Gerard-Varet (1979) using Bayesian equilibria have shown that positive results are achievable when prior information is available. These results have played a major role in opening new avenues in the economics of information. The reason is that generalizations of these mechanisms have provided a precise way of evaluating transaction costs due to asymmetric information. Industrial organization, macroeconomics, and public economics have been considerably renewed recently by the possibility of taking seriously into account the decentralization of information.

See Also

- ▶ Bidding
- ▶ Incentive Compatibility
- ▶ Lindahl Equilibrium
- ▶ Organization Theory
- ▶ Public Economics
- ▶ Public Goods

Bibliography

d' Aspremont, C., and L.A. Gerard-Varet. 1979. Incentives and incomplete information. *Journal of Public Economics* 11: 25–45.

Clarke, E.H. 1971. Multipart pricing of public goods. *Public Choice* 2: 19–33.

Gibbard, A. 1973. Manipulation of voting schemes. A general result. *Econometrica* 41: 487–601.

Groves, T. 1973. Incentives in teams. *Econometrica* 41: 617–631.

Samuelson, P.A. 1947. *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Samuelson, P.A. 1954. The pure theory of public expenditure. *Review of Economics and Statistics* 37: 350–356.

Satterthwaite, M. 1975. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10: 187–217.

Vickrey, W. 1961. Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance* 16: 1–17.

Wicksell, K. 1896. *Finanztheoretische Untersuchungen und das Steuerwesen*. Jena: Schweders.

Revelation Principle

Roger B. Myerson

Abstract

In any economic institution, individuals must be given appropriate incentives to share private information or to exert unobserved efforts. The revelation principle is a technical insight that allows us to make general statements about what allocation rules are feasible, subject to incentive constraints, in economic problems with adverse selection and moral hazard. The revelation principle tells us that, for any general coordination mechanism, any equilibrium of rational communication strategies for the economic agents can be simulated by an equivalent incentive-compatible direct-revelation mechanism, where a trustworthy mediator maximally centralizes communication and makes honesty and obedience rational equilibrium strategies for the agents.

Keywords

Adverse selection; Aumann, R.; Bayesian games; Capitalism; Centralization; Communication; Correlated equilibrium; Decentralization; Direct-revelation mechanisms; Hayek, F. von; Honesty; Incentive compatibility; Incentive constraints; Moral hazard; Nash, J.; Obedience; Principal and agent; Private information; Revelation principle; Sequential equilibrium; Socialism; Strategic-form games; Trust

JEL Classifications
D43; D89; C72

Communication is central to the economic problem (Hayek 1945). Opportunities for mutually beneficial transactions cannot be found unless individuals share information about their preferences and endowments. Markets and other economic institutions should be understood as mechanisms for facilitating communication. However, people cannot be expected to reveal information when it is against their interests; for example, a seller may conceal his willingness to sell at a lower price. Rational behaviour in any specific communication mechanism can be analysed using game-theoretic equilibrium concepts, but efficient institutions can be identified only by comparison with all possible communication mechanisms. The revelation principle is a technical insight that allows us, in any given economic situation, to make general statements about all possible communication mechanisms.

The problem of making statements about all possible communication systems might seem intractably complex. Reports and messages may be expressed in rich languages with unbounded vocabulary. Communication systems can include both public announcements and private communication among smaller groups. Communication channels can have noise that randomly distorts messages. A communication mechanism may also specify how contractually enforceable transactions will depend on agents' reports and messages. So a general communication mechanism for any given set of agents may specify (*a*) a set of possible reports that each agent can send, (*b*) a set of possible messages that each agent can receive from the communication system, and (*c*) a probabilistic rule for determining the messages received and the enforceable transactions as a function of the reports sent by the agents. However, the revelation principle tells us that, for many economic purposes, it is sufficient for us to consider only a special class of mechanisms, called 'incentive-compatible direct-revelation mechanisms'.

In these mechanisms, every economic agent is assumed to communicate only with a central mediator. This mediator may be thought of as a trustworthy person or as a computer at the centre of a telephone network. In a direct-revelation mechanism, each individual is asked to report all of his private information confidentially to the mediator. After receiving these reports, the mediator then specifies all contractually enforceable transactions, as a function of these reports. If any individual controls private actions that are not contractually enforceable (such as efforts that others cannot observe), then the mediator also confidentially recommends an action to the individual. A direct-revelation mechanism is any rule for specifying how the mediator determines these contractual transactions and privately recommended actions, as a function of the private-information reports that the mediator receives.

A direct-revelation mechanism is said to be 'incentive compatible' if, when each individual expects that the others will be honest and obedient to the mediator, then no individual could ever expect to do better (given the information available to him) by reporting dishonestly to the mediator or by disobeying the mediator's recommendations. That is, the mechanism is incentive compatible if honesty and obedience is an equilibrium of the resulting communication game. The set of incentive-compatible direct-revelation mechanisms has good mathematical properties that often make it easy to analyse because it can be defined by a collection of linear inequalities, called 'incentive constraints'. Each of these incentive constraints expresses a requirement that an individual's expected utility from using a dishonest or disobedient strategy should not be greater than the individual's expected utility from being honest and obedient, when it is anticipated that everyone else will be honest and obedient.

The analysis of such incentive-compatible direct-revelation mechanisms might seem to be of limited interest, because real institutions rarely use such fully centralized mediation and often generate incentives for dishonesty or disobedience. For any equilibrium of any general communication mechanism, however, there exists an incentive-compatible direct-revelation mechanism that is essentially equivalent. This proposition is the revelation principle. Thus, the revelation principle tells us that, by analysing the set of incentive-compatible direct-revelation mechanisms, we can derive general properties of all equilibria of all coordination mechanisms.

The terms 'honesty' and 'obedience' here indicate two fundamental aspects of the general economic problem of communication. In a general communication system, an individual may send out messages or reports to share information that he knows privately, and he may also receive messages or recommendations to guide actions that he controls privately. The problem of motivating individuals to report their private information honestly is called 'adverse selection', and the problem of motivating individuals to implement their recommended actions obediently is called 'moral hazard'. To describe the intuition behind the revelation principle, let us consider first the special cases where only one or the other of these problems exists.

Pure Adverse Selection

First, let us formulate the revelation principle for the case of pure adverse selection, as developed in Bayesian social choice theory. In this case we are given a set of individuals, each of whom has some initial private information that may be called the individual's 'type', and there is a planning question of how a social allocation of resources should depend on the individuals' types. Each individual's payoff can depend on the resource allocation and on the types of all individuals according to some given utility function, and each type of each individual has some given probabilistic beliefs about the types of all other individuals. A general communication system would allow each individual i to send a message m_i in some rich language, and then the chosen resource allocation would depend on all these messages according to some rule γ-$(m_1, \ldots m_n)$. In any equilibrium of the game defined by this communication system, each individual i must have some strategy σ_i for choosing his message as a function of his type t_i, so that $m_i = \sigma_i(t_i)$.

For the given equilibrium $(\sigma_1, \ldots, \sigma_n)$ of the given social-choice rule γ, the revelation principle is satisfied by a mediation plan in which each individual is asked to confidentially report his type t_i to a central mediator, who then implements the social choice

$$\mu(t_1, \ldots t_n) = \gamma(\sigma_1(t_1), \ldots, \sigma_n(t_n)).$$

So the mediator computes what message would be sent by the reported type of each individual i under his or her strategy σ_i, and then the mediator implements the resource allocation that would result from these messages under the rule γ. It is easy to see that honesty is an equilibrium under this mediation plan μ. If any individual could gain by lying to this mediator, when all others are expected to be honest, then this individual could have also gained by lying to himself when implementing his equilibrium strategy σi under the given mechanism γ, which would contradict the optimality condition that defines an equilibrium. So μ is an incentive-compatible direct-revelation mechanism that is equivalent to the given general mechanism γ with the given equilibrium $(\sigma_1, \ldots, \sigma_n)$.

In this case of pure adverse selection, the revelation principle was introduced by Gibbard (1973), but for a narrower solution concept (dominant strategies, instead of Bayesian equilibrium). The revelation principle for the broader solution concept of Bayesian equilibrium was recognized by Dasgupta, Hammond and Maskin (1979), Harris and Townsend (1981), Holmstrom (1977), Myerson (1979), and Rosenthal (1978).

Pure Moral Hazard

Next let us formulate the revelation principle for the case of pure moral hazard, as developed in Aumann's (1974) theory of correlated equilibrium. In this case we are given a set of individuals, each of whom controls some actions, and each individual's payoff can depend on the actions $(c_1, \ldots, c_n)$ that are chosen by all individuals, according to some given utility function $u_i(c_1, \ldots c_n)$. That is, we are given a game in strategic form. In this case of pure moral hazard, nobody has any private information initially, but a communication process could give individuals different information before they choose their actions. In a general communication system, each individual i could get some message m_i in some rich language, with these messages $(m_1, \ldots, m_n)$ being randomly drawn from some joint probability distribution ρ. In any equilibrium of the game generated by adding this communication system, each individual i has some strategy σ_i for choosing his action c_i as a function of his message m_i, so that $c_i = \sigma_i(m_i)$.

For the given equilibrium $(\sigma_1, \ldots, \sigma_n)$ of the game with the given communication system ρ, the revelation principle is satisfied by a mediation plan in which the mediator randomly generates recommended actions in such a way that the probability of recommending actions $(c_1, \ldots, c_n)$ is the same as the probability of the given communication system ρ yielding messages $(m_1, \ldots, m_n)$ that would induce the players to choose $(c_1, \ldots, c_n)$ in the σ equilibrium. That is, the probability μ-$(c_1, \ldots, c_n)$ of the mediator recommending $(c_1, \ldots, c_n)$ is

$$\mu(c_1, \ldots, c_n) = \rho(\{(m_1, \ldots, m_n) | \sigma_1(m_1) = c_1, \ldots, \sigma_n(m_n) = c_n\}).$$

Then the mediator confidentially tells each individual i only which action c_i is recommended for him. Obedience is an equilibrium under this mediation plan μ because, if any individual could gain by disobeying this mediator when all others are expected to be obedient, then this individual could have also gained by disobeying himself in implementing his equilibrium strategy σ_i in the given game with communication system ρ. So μ is an incentive-compatible direct-revelation mechanism that is equivalent to the given mechanism ρ with the given equilibrium $(\sigma_1, \ldots, \sigma_n)$.

General Formulations

Problems of adverse selection and moral hazard can be combined in the framework of Harsanyi's (1967) Bayesian games, where players have both types and actions. The revelation principle for general Bayesian games was formulated by Myerson (1982, 1985). A further generalization of the revelation principle to multistage games was formulated by Myerson (1986). In each case, the basic idea is that any equilibrium of any general communication system can be simulated by a maximally centralized communication system in which, at every stage, each individual confidentially reports all his private information to a central mediator, and then the mediator confidentially recommends an action to each individual, and the mediator's rule for generating recommendations from reports is designed so that honesty and obedience form an equilibrium of the mediated communication game.

The basic assumption here is that, although the motivations of all economic agents are problematic, we can find a mediator who is completely trustworthy and has no costs of processing information. Asking agents to reveal all relevant information to the trustworthy mediator maximizes the mediator's ability to implement any coordination plan. But telling any other agent more than is necessary to guide his choice of action would only increase the agent's ability to find ways of profitably deviating from the coordination plan.

For honesty and obedience to be an equilibrium, the mediation plan must satisfy incentive constraints which say that no individual could ever expect to gain by deviating to a strategy that involves lying to the mediator or disobeying a recommendation from the mediator. In a dynamic context, we must consider that an individual's most profitable deviation from honesty and obedience could be followed by further deviations in the future. So, to verify that an individual could never gain by lying, we must consider all possible deviation strategies in which the individual may thereafter choose actions that can depend disobediently on the mediator's recommendations (which may convey information about others' types and actions).

When we use sequential equilibrium as the solution concept for dynamic games with communication, the set of actions that can be recommended in a sequentially incentive-compatible mechanism must be restricted somewhat. In a Bayesian game, if some action d_i could never be optimal for individual i to use when his type is t_i, no matter what information he obtained about others' types and actions, then obedience could not be sequentially rational in any mechanism where the mediator might ever recommend this action d_i to i after he reports type t_i. Myerson (1986) identified a larger set of *co-dominated actions* that can never be recommended in any sequentially incentive-compatible mechanism. Suppose that, if any individual observed a zero-probability event, then he could attribute this surprise to a mistake by the trembling hand of the mediator. Under this assumption, Myerson (1986) showed that the effect of requiring sequential rationality in games with communication is completely characterized by the requirement that no individuals should ever be expected to choose any co-dominated actions. (See Gerardi and Myerson 2007.)

Limitations

The revelation principle says that each equilibrium of any communication mechanism is equivalent to the honest-obedient equilibrium of an incentive-compatible direct-revelation mechanism. But this direct-revelation mechanism may have other dishonest equilibria, which might not correspond to equilibria of the original mechanism. So the revelation principle cannot help us when we are concerned about the whole set of equilibria of a communication mechanism. Similarly, a given communication mechanism may have equilibria that change in some desirable way as we change the players' given beliefs about each others' types, but these different equilibria would correspond to different incentive-compatible mechanisms, and so this desirable property of the given mechanism could not be recognized with the revelation principle.

The assumption that perfectly trustworthy mediators are available is essential to the mathematical simplicity of the incentive-compatible set. Otherwise, if individuals can communicate only by making public statements that are immediately heard by everybody, then the set of equilibria may be smaller and harder to compute.

In principal-agent analysis we often apply the revelation principle to find the incentive-compatible mechanism that is optimal for the principal. If the principal would be tempted to use revealed information opportunistically, then there could be loss of generality in assuming that the agents reveal all their private information to the principal. But we should not confuse the principal with the mediator. The revelation principle can still be applied if the principal can get a trustworthy mediator to take the agents' reports and use them according to any specified mechanism.

There are often questions about whether the allocation selected by a mechanism could be modified by subsequent exchanges among the individuals. An individual's right to offer his possessions for sale at some future date could be accommodated in mechanism design by additional moral-hazard constraints.

For example, suppose the principal can sell an object each day, on days 1 and 2. The only buyer's value for such objects is either low \$1 or high \$3, low having probability 0.25. To maximize the principal's expected revenue with the buyer participating honestly, an optimal mechanism would sell both objects for \$3 if the buyer's type is high, but would sell neither if the buyer is low. But if no sale is recommended then the principal could infer that the buyer is low and would prefer to sell for \$1. Suppose now that the principal cannot be prevented from offering to sell for \$1 on either day. With these additional moral-hazard constraints, an optimal mechanism uses randomization by the mediator to conceal information from the principal. If the buyer reports low then the mediator recommends no sale on day 1 and selling for \$1 on day 2. If the buyer reports high, then with probability 1/3 the mediator recommends no sale on day 1 and selling for \$3 on day 2, but with probability 2/3 recommends selling for \$1.50 on both days. A no-sale recommendation on day 1 implies probability 0.5 of low, so that obedience yields the same expected revenue $0.5 \times (0 + 1) + 0.5 \times (0 + 3)$ as deviating to sell for \$1 on both days.

A proliferation of such moral-hazard constraints may greatly complicate the analysis, however. So in practice we often apply the revelation principle with an understanding that we may be overestimating the size of the feasible set, by assuming away some problems of mediator imperfection or moral hazard. When we use the revelation principle to show that a seemingly wasteful mechanism is actually efficient when incentive constraints are recognized, such overestimation of the incentive-feasible set would not weaken the impact of our results (as long as this mechanism remains feasible).

Centralized mediation is emphasized by the revelation principle as a convenient way of characterizing what people can achieve with communication, but this analytical convenience does not imply that centralization is necessarily the best way to coordinate an economy. For fundamental questions about socialist centralization versus free-market decentralization, we should be sceptical about an assumption that centralized control over national resources could not corrupt any mediator. The power of the revelation principle for such questions is instead its ability to provide a common analytical framework that applies equally to socialism and capitalism. For example, a standard result of revelation-principle analysis is that, if only one producer knows the production cost of a good, then efficient incentive-compatible mechanisms must allow this monopolistic producer to take positive informational rents or profits (Baron and Myerson 1982). Thus the revelation principle can actually be used to support arguments for decentralized multi-source production, by showing that problems of profit-taking by an informational monopolist can be just as serious under socialism as under capitalism.

Nash (1951) advocated a different methodology for analysing communication in games. In Nash's approach, all opportunities for communication should be represented by moves in our extensive model of the dynamic game. Adding such communication moves may greatly increase the number of possible strategies for a player, because each strategy is a complete plan for choosing the player's moves throughout the game. But if all communication will occur in the implementation of these strategies, then the players' initial choices of their strategies must be independent. Thus, Nash argued, any dynamic game can be normalized to a static strategic-form game, where players choose strategies simultaneously and independently, and Nash equilibrium is the general solution for such games.

With the revelation principle, however, communication opportunities are omitted from the game model and are instead taken into account by using incentive-compatible mechanisms as our solution concept. Characterizing the set of all incentive-compatible mechanisms is often easier than computing the Nash equilibria of a game with communication. Thus, by applying the revelation principle, we can get both a simpler model and a simpler solution concept for games with communication. But, when we use the revelation principle, strategic-form games are no longer sufficient for representing general dynamic games, because normalizing a game model to strategic form would suppress implicit opportunities for communicating during the game (see Myerson 1986). So the revelation principle should be understood as a methodological alternative to Nash's strategic-form analysis.

R

See Also

▶ Mechanism Design

Bibliography

Aumann, R. 1974. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* 1: 67–96.

Baron, D., and R. Myerson. 1982. Regulating a monopolist with unknown costs. *Econometrica* 50: 911–930.

Dasgupta, P., P. Hammond, and E. Maskin. 1979. The implementation of social choice rules: Some results on incentive compatibility. *Review of Economic Studies* 46: 185–216.

Gerardi, D., and R. Myerson. 2007. Sequential equilibrium in Bayesian games with communication. *Games and Economic Behavior* 60: 104–134.
Gibbard, A. 1973. Manipulation of voting schemes: A general result. *Econometrica* 41: 587–601.
Harris, M., and R. Townsend. 1981. Resource allocation under asymmetric information. *Econometrica* 49: 1477–1499.
Harsanyi, J. 1967. Games with incomplete information played by Bayesian players. *Management Science* 14: 159–182 .320–34, 486–502.
Hayek, F. 1945. The use of knowledge in society. *American Economic Review* 35: 519–530.
Holmstrom, B. 1977. *On incentives and control in organizations*. Ph.D. thesis, Stanford University.
Myerson, R. 1979. Incentive-compatibility and the bargaining problem. *Econometrica* 47: 61–73.
Myerson, R. 1982. Optimal coordination mechanisms in generalized principal-agent problems. *Journal of Mathematical Economics* 10: 67–81.
Myerson, R. 1985. Bayesian equilibrium and incentive compatibility: An introduction. In *Social goals and social organization*, ed. L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge: Cambridge University Press.
Myerson, R. 1986. Multistage games with communication. *Econometrica* 54: 323–358.
Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 54: 286–295.
Rosenthal, R. 1978. Arbitration of two-party disputes under uncertainty. *Review of Economic Studies* 45: 595–604.

Revenue, Gross and Net

Giorgio Gilibert

The term revenue was once used as equivalent to the modern 'income', which has now replaced it. Both the concept and the word came from France, where *revenu* is the past participle of *revenir,* to return.

Gross revenue is defined by Adam Smith as 'the whole annual produce of the land and labour' of a country. Net revenue is what the inhabitants are free to spend 'upon their subsistence, conveniences, and amusements' 'without encroaching upon their capital' (Smith 1776, pp. 286–7). The ambiguities of this definition have generated a two centuries old discussion: should we consider the subsistence fund of the wage-earners as circulating capital, therefore excluding it from the net revenue, or as final consumption, and then as a part of it?

A clear-cut answer was offered by the eighteenth-century French predecessors of Smith. Having in mind a specific picture of the circular process of production, they defined net revenue (the physiocratic *produit net*) as the annually produced wealth (*reproduction totale*) minus the advances required to repeat the process on the same scale. Workers' subsistences, to which wages were strictly limited, were an obvious part of the advances, on the same footing as the feed for the cattle. Net revenue was then the value of the surplus product, which remained available to be – in the words of Nicolas Isnard (1781, p. 37), a civil engineer – 'nobly enjoyed by the proprietors'.

The clear-cutness faded away with Smith, and *pour cause*: he was trying to take into account workers' consumption (and the employment level) in assessing the prosperity of a nation. Therefore he was led sometimes to include wages in the net revenue and sometimes to consider gross (instead of net) revenue as the crucial indicator for the evaluation of prosperity.

Ricardo (1821, ch. XXVI) reverted unambiguously to the original meaning of net revenue, identifying it with the sum of rents and profits alone. He criticized Smith for his preferring 'a large gross, rather than a large net income'. The rationale of this attitude was that the power of a country 'of supporting fleets and armies, and all species of productive labour' is in proportion to its net income, from which taxes are paid (for a similar argument, see Quesnay 1759). He also admitted, however, that if more than subsistence is allotted to wage-earners 'a part of the net produce of the country is received by the labourer'.

Marx followed the Ricardian definition, but stressed the importance of taking account of constant capital, that is of the value of raw materials and the depreciation of fixed capital, in evaluating the gross revenue.

In more recent times, the current notion of national income (inclusive of wages) definitely prevailed in applied economics; but the theoretical question remained unsettled. See, for instance: Edgeworth (1896), quoting Jevons's remark ('as the horse has to be clothed and stabled, so the productive labourer has to be clothed and housed') as an argument in favour of the exclusion of necessary consumption from the net income; Leontief (1941) who emphasized the arbitrary nature of the definition, depending on the type and level of aggregation of the input–output scheme (a completely consolidated table, reduced to a single box, would show no net revenue); and Sraffa (1960).

Piero Sraffa defines national income as the gross product minus the value of the commodities used up in all industries; while – à la Ricardo – the subsistences of the workers 'continue to appear with the fuel, etc., among the means of production' (§ 8). On the other hand, attention is focused on the movements of the 'surplus' part of the wage, which participates in net revenue.

See Also

▶ Net Product
▶ Produit Net

Bibliography

Edgeworth, F.Y. 1896. Income. In *Palgrave's dictionary of political economy,* vol. 2. London: Macmillan.
Isnard, A.N. 1781. *Traité des richesses*. Lausanne: Grasset.
Leontief, W. 1941. *The structure of American economy.* New York: Oxford University Press, 1951.
Quesnay, F. 1759. *Tableau économique*, ed. M. Kuczynski and R. Meek. London: Macmillan, 1972.
Ricardo, D. 1821. *Principles of political economy,* 3rd edn. In *Works and correspondence of David Ricardo,* vol. I, ed. P. Sraffa. Cambridge: Cambridge University Press, 1951.
Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*, ed. R.H. Campbell, A.S. Skinner and W.B. Todd. Oxford: Clarendon Press, 1976.
Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.

Reverse Capital Deepening

Roberto Scazzieri

Abstract

Reverse capital deepening is the property whereby it may be efficient to associate a lower (higher) rate of interest with a lower (higher) capital per worker. This property is inconsistent with the traditional belief that, by virtue of the substitution principle, production techniques that are more 'capital intensive' will become optimal as the rate of interest is lowered. Reverse capital deepening is an important instance of the apparent paradoxes associated with indirect effects in a production economy. It entails that technical choice cannot be considered a monotonic function of the rate of interest, and questions the widespread policy implications of the traditional view.

Keywords

Böhm-Bawerk, E. von; Clark, J. B.; Capital accumulation; Capital deepening; Capital theory; Hayek, F. A.; Reswitching of technique; Reverse capital deepening; Technical choice; Wage–profit relationship

JEL Classifications

E22

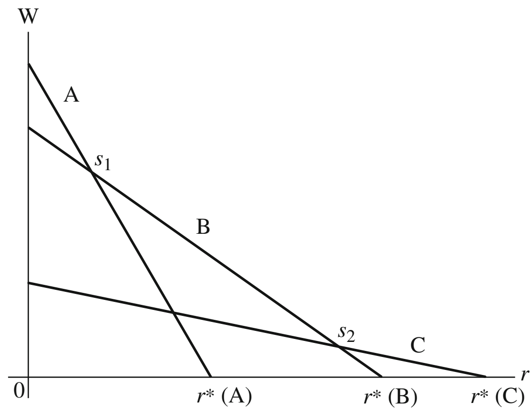
It has long been taken for granted that there is an inverse monotonic relationship between the rate of interest (or the rate of profit) and the quantity of capital per worker. This belief was founded on the principle of substitution, whereby 'cheaper' is substituted for 'more expensive' as the relative price of two inputs is changed.

In the field of capital theory, the principle of substitution persuaded many economists, such as E. von Böhm-Bawerk (1889), J.B. Clark (1899) and F.A. von Hayek (1941), that a lower rate of interest (which is equal to the rate of profit in equilibrium) is associated with the use of more 'capital intensive' techniques, and thus with the substitution of capital for other productive factors, such as labour or land. This process is called capital deepening.

Recent discussions have shown that this is not necessarily true, since a lower rate of interest might be associated with *lower*, rather than higher, capital per worker. This phenomenon is called *reverse capital deepening*.

This discovery was made at the same time as it was realized that it is not generally possible to order 'efficient' techniques in such a way that technical choice becomes a monotonic function of the rate of interest (and of the rate of profit).

It can be shown that both reverse capital deepening and reswitching of technique are related to the same fundamental property of the economic system: the possibility (in fact, the near generality) of nonlinear wage–profit relationships. To illustrate this proposition, it is useful to begin by considering the hypothetical case of linear wage–profit relationships (see Fig. 1).

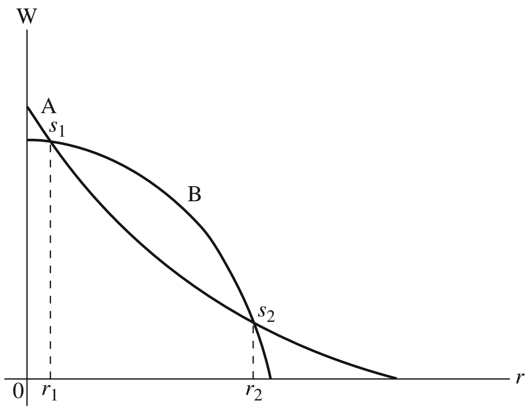


Reverse Capital Deepening, Fig. 1

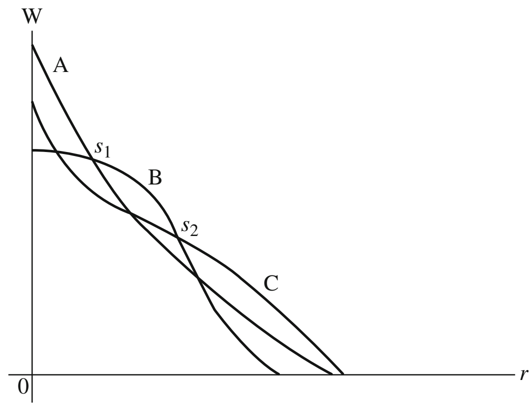
The linearity of the three wage–profit relationships makes reswitching impossible as r increases between 0 and r^*(C) (which is the maximum rate of profit with technology C). The reason is that no wage–profit line can ever be crossed more than once by another wage–profit line. In this special case, there is an inverse monotonic relationship between the rate of profit and the quantity of capital per worker. This may be shown as follows. We can read the net final output per worker on the w-axis of Fig. 1 at the point at which $r = 0$. (At that point the net final output per worker coincides with the maximum wage.) The net final output per worker associated with technology A is higher than the net final output per worker associated with technology B. The net final output per worker associated with technology B is higher than the net final output per worker associated with technology C. At switchpoints s_1 and s_2 the wage is the same for both technologies between which substitution takes place. It follows that, at switchpoint s_1, profit per worker is higher with technology A than with technology B. Similarly, at switchpoint s_2, profit per worker is higher with technology B than with technology C. Assuming that the rate of profit is uniform across technologies, we find that, at s_1, A is associated with higher capital per worker than B. A higher rate of profit (or rate of interest) is thus associated with substitution of 'less capital' for 'more capital'. In this particular case, the traditional approach to capital theory would seem to be well founded.

However, these properties disappear altogether once we drop the assumption of linear wage–profit relationships. (It might be interesting to inquire into the economic meaning of straight wage–profit relationships, which are possible only in the case of a technology characterized by a uniform proportion between labour and intermediate inputs in all production processes: only in this case a change in the rate of profit leaves relative prices unaffected.)

But in general wage–profit relationships are of the nonlinear type, which means that the proportion between labour and intermediate inputs is generally different from one production process to another. This feature of the wage–profit frontier makes it possible for wage–profit curves to intersect more than once, thus bringing about the possibility of multiple switching. Under the same circumstances it can be shown that the relationship between the rate of profit and capital per worker is no longer of the inverse monotonic type. This can be seen in the reswitching case (Fig. 2), but it can also be seen in the case in which the wage–profit curves never intersect



Reverse Capital Deepening, Fig. 2



Reverse Capital Deepening, Fig. 3

more than once on the efficiency frontier (Fig. 3). (See also Pasinetti 1966.)

In Fig. 2, reswitching is associated with reverse capital deepening. Technology A is the more profitable at levels of the rate of profit lower than r_1, it is 'overtaken' by technology B at rates of profit between r_1 and r_2, it becomes again the more profitable at rates of profit higher than r_2. At the same time, switchpoint s_1 is associated with the substitution of the technology with lower value of capital per worker (B) for the technology with the higher value of capital per worker (A), whereas at switchpoint s_2 the opposite happens: the technology with higher capital per worker (A) is substituted for the technology with lower capital per worker (B), in spite of the fact that the rate of profit is higher (reverse capital deepening).

In Fig. 3, there is no reswitching but we still have a reverse capital deepening. For no wage–profit curves cross one another more than once on the efficiency frontier, but at switchpoint s_2 an increasing rate of profit is associated with the substitution of a technology with higher capital per worker (C) for a technology with lower capital per worker (B).

Complementarity in production is often at the root of apparently perverse price behaviour (see Broome 1978). One reason for this had been noted by John Hicks, when he wrote that the 'net effect' of a change in the price of productive factor x upon the price of complementary factor y 'is . . . compounded out of two contrary tendencies, a direct effect tending to raise it, and indirect effect tending to reduce it; either may be dominant' (Hicks 1946, p. 107). Reverse capital deepening is an especially important instance of a widespread phenomenon associated with indirect effects in a production economy. This possibility is associated with other phenomena which are not compatible with traditional beliefs about capital and capital accumulation. Simple inspection of Fig. 2 or 3 shows that at a switchpoint associated with reverse capital deepening (s_2 in either figure) a technology with higher capital per worker and higher net final product per worker is substituted for a technology with lower capital per worker and lower net final product per worker. At such switchpoints a higher rate of profit (and rate of interest) could be associated with a higher ratio of capital per worker to net final product per worker, that is, with a higher capital–output ratio.

Figures 2 and 3 also alert us as to the possibility that a technology adopted at a high rate of interest is associated with higher maximum consumption per head than a technology adopted at a lower rate of interest. In addition, transition to a lower rate of interest may involve the switch to a lower maximum consumption per head. (This can be seen at switchpoint s_2 in either figure, where maximum consumption per head can be read on the w-axis at point $r = 0$.) This behaviour of consumption per head in relation to the rate of interest is clearly incompatible with the view that a higher rate of interest brings about a special type of exchange, in which less consumption in the current period is substituted against higher consumption in the future. Reverse capital deepening alerts us as to the possibility that a higher rate of interest might be associated with greater current consumption per head than the consumption per head feasible with the technology adopted at a lower rate of interest (see Bruno et al. 1966; Samuelson 1966).

The relevance of reverse capital deepening is that the foundation of the traditional view that technical choice is a monotonic function of the rate of interest is seriously questioned. Similarly, the widespread policy implications of the traditional view are also questioned. However, there is as yet no full agreement as to the main consequences of this result. For example, Christopher Bliss (1975, p. 279) has noted that reverse capital deepening makes it impossible to see the accumulation of capital as a process associated with 'a continuous increase in consumption per capita,. . . a continuous decline in the rate of interest and . . . a continuous increase in the real wage rate'. He also called attention to the fact that the 'extended accumulation history' of an economic system moving through real time is normally different from the hypothetical history we can tell by 'travelling' across steady states (1975, pp. 194, 280–1). In particular, he emphasized that the statement that the rate of interest may be expected to fall as capital deepening takes place 'cannot be interpreted' in the case of extended accumulation history, as we would have, in that case, 'a whole structure of interest rates . . . not a single rate of interest' (1975, p. 294). A different point of view has been expressed by Pierangelo Garegnani, who has maintained that capital paradoxes in general, and reverse capital deepening in particular, by making traditional beliefs untenable, suggest a 'correction' to traditional theory, which would make it reasonable to expect 'instabilities or tendencies to zero of wages, or of net returns on capital' (Garegnani 1990, p. 76). This author's view is that, rather than introducing such a correction, one should drop any idea of a causal connection from marginal products to the distribution of the social product, and further develop the conjecture that distribution is brought about by 'more complex economic and social forces like those envisaged by the old classical economists' (Garegnani 1990, pp. 76–7). As it is common in theoretical sciences (see Kuhn 1970, 2000), the discovery of an apparent anomaly has induced economists to look for a more general theory, or to switch to an altogether different framework.

See Also

- ▶ Capital Theory
- ▶ Capital Theory (Paradoxes)
- ▶ Production Functions
- ▶ Reswitching of Technique

Bibliography

Bliss, C.J. 1975. *Capital theory and the distribution of income*. Amsterdam/Oxford/New York: North-Holland Publishing Company/American Elsevier Publishing Company.

Broome, J. 1978. Perverse prices. *Economic Journal* 88: 778–787.

Bruno, M., E. Burmeister, and E. Sheshinski. 1966. The nature and implications of the reswitching of techniques. *Quarterly Journal of Economics* 80: 526–553.

Clark, J.B. 1899. *The distribution of wealth*. New York: Macmillan.

Garegnani, P. 1990. Quantity of capital. In *The New Palgrave: Capital theory*, ed. J. Eatwell et al. London/Basingstoke: Macmillan.

Hicks, J. 1946. *Value and capital*, 2nd ed. Oxford: Clarendon Press.

Kuhn, T.S. 1970. *The structure of scientific revolutions*, 2nd ed enlarged. Chicago: University of Chicago Press.

Kuhn, T.S. 2000. *The road since structure: Philosophical essays, 1970–1993, with an autobiographical interview*, ed. J. Conant and J. Haugeland. Chicago/London: Chicago University Press.

Pasinetti, L.L. 1966. Changes in the rate of profit and switches of techniques. *Quarterly Journal of Economics* 80: 503–517.

Samuelson, P. 1966. A summing up. *Quarterly Journal of Economics* 80: 568–583.

von Böhm-Bawerk, E. 1889. *Positive Theorie des Kapitales*. Zweite Ableitung: *Kapital und Kapitalzins*. Trans. as *The positive theory of capital*, London: Macmillan, 1891.

von Hayek, F.A. 1941. *The pure theory of capital*. London: Routledge.

Rhetoric of Economics

Donald N. McCloskey

Keywords
Rhetoric; Rhetoric of economics

JEL Classifications
B4

Rhetoric is the study and practice of persuasive expression, an alternative since the Greeks to the philosophical programme of epistemology. The rhetoric of economics examines how economists persuade – not how they say they do, or how their official methodologies say they do, but how in fact they persuade colleagues and politicians and students to accept one economic assertion and reject another.

Some of their devices arise from bad motives, and bad rhetoric is what most people have in mind when they call a piece of writing 'rhetorical'. An irrelevant and inaccurate attack on Milton Friedman's politics while criticizing his economics would be an example, as would a pointless and confusing use of mathematics while arguing a point in labour economics. The badness does not reside in the techniques themselves (political commentary or mathematical argument) but in the person using them, since all techniques can be abused. Aristotle noted that 'if it be objected that one who uses such power of speech unjustly might do great harm, *that* is a charge which may be made in common against all good things except virtue itself'. Cato the Elder demanded that the user of analogy (or in our time the user of regression) be *vir bonus dicendi peritus,* the *good* man skilled at speaking. The protection against bad science is good scientists, not good methodology.

Rhetoric, then, can be good, offering good reasons for believing that the elasticity of substitution between capital and labour in American manufacturing, say, is about 1.0. The good reasons are not confined by syllogism and number. They include good analogy (production is *just like* a mathematical function), good authority (Knut Wicksell and Paul Douglas thought this way, too), good symmetry (if mining can be treated as a production function, so should manufacturing). Furthermore, the reasonings of syllogism and number are themselves rhetorical, that is, persuasive acts of human speech. An econometric test will depend on how apt is an analogy of the error term with drawings from an urn. A mathematical proof will depend on how convincing is an appeal to the authority of the Bourbaki style. 'The facts' and 'the logic' matter, of course; but they are part of the rhetoric, depending themselves on the giving of good reasons.

Consider, for example, the sentence in economics, 'The demand curve slopes down'. The official rhetoric says that economists believe this because of statistical evidence – negative coefficients in demand curves for pig iron or negative diagonal items in matrices of complete systems of demand – accumulating steadily in journal articles. These are the tests 'consistent with the hypothesis'. Yet most belief in the hypothesis comes from other sources: from introspection (what would I do?); from thought experiments (what would they do?); from uncontrolled cases in point (such as the oil crisis); from authority (Alfred Marshall believed it); from symmetry (a law of demand if there is a law of supply); from definition (a higher price leaves less for expenditure, including this one); and above all, from analogy (if the demand curve slopes down for chewing gum, why not for housing and love too?). As may be seen in the classroom and seminar, the range of argument in economics is wider than the official rhetoric allows.

The rhetoric of economics brings the traditions of rhetoric to the study of economic texts, whether mathematical or verbal texts. It is a literary criticism of economics, or a jurisprudence, and from literary critics like Wayne Booth (1974) and lawyers such as Chaim Perelman (Perelman and Olbrechts-Tyteca 1958) much can be learned. Although its precursors in economics are methodological criticisms of the field (such as Frank Knight 1940), censorious joking (such as Stigler 1977), and finger-wagging presidential addresses (such as Leontief 1971, or Mayer 1980), the main focus of the work has been the analysis of how economists seek to persuade, whether good or bad (Klamer 1984; Henderson 1982; Kornai 1983; McCloskey 1986). Econometrics has its own rhetorical prehistory, more self-conscious than the rest (Leamer 1978), reaching back to the founders of decision theory and Bayesian statistics.

The movement has parallels in other fields. Imre Lakatos (1976), Davis and Hersh (1981), and others have uncovered a rhetoric in mathematics; Rorty (1982), Toulmin (1958), and Rosen (1980) in technical philosophy; and numbers of scientists in their own fields (Polanyi 1962; Medawar 1964). Historians and sociologists of science have since the 1960s accumulated much evidence that science is a conversation rather than a mechanical procedure (Kuhn 1977; Collins 1985). The analysis of conversation from scholars in communication and literary studies (Scott 1967) has provided ways of rereading various fields (a sampling of these is contained in Nelson et al. 1987).

A rhetoric of economics questions the division between scientific and humanistic reasoning, not to attack quantification or to introduce irrationality into science, but to make the scientific conversation more aware of itself. It is a programme of greater, not less, rigour and relevance, of higher, not lower, standards in the conversations of mankind.

See Also

▶ Philosophy and Economics

Bibliography

Booth, W. 1974. *Modern dogma and the rhetoric of assent*. Chicago: University of Chicago Press.

Collins, H.M. 1985. *Changing order: Replication and induction in scientific practice*. London: Sage.

Davis, P.J., and R. Hersh. 1981. *The mathematical experience*. Boston: Houghton Mifflin.

Henderson, W. 1982. Metaphors in economics. *Economics* 18 (4): 147–153.

Klamer, A. 1984. *Conversations with economists: New classical economists and opponents speak out on the current controversy in macroeconomics*. Totowa: Rowman and Allanheld.

Knight, F. 1940. 'What is truth' in economics? *Journal of Political Economy* 48: 1–32.

Kornai, J. 1983. The health of nations: Reflections on the analogy between medical science and economics. *Kyklos* 36 (2): 191–212.

Kuhn, T. 1977. *The essential tension: Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Lakatos, I. 1976. *Proofs and refutations: The logic of mathematical discovery*. Cambridge: Cambridge University Press.

Leamer, E. 1978. *Specification searches: Ad hoc inferences with nonexperimental data*. New York: Wiley.

Leontief, W. 1971. Theoretical assumptions and nonobserved facts. *American Economic Review* 61: 1–7.

Mayer, T. 1980. Economics as a hard science: Realistic goal or wishful thinking? *Economic Inquiry* 18: 165–178.

McCloskey, D.N. 1986. *The rhetoric of economics*. Madison: University of Wisconsin Press.

Medawar, P. 1964. Is the scientific paper fraudulent? *Saturday Review*, 1 August.

Nelson, J., A. Megill, and D.N. McCloskey, eds. 1987. *The rhetoric of the human sciences: Papers and proceedings of the Iowa conference*. Madison: University of Wisconsin Press.

Perelman, C., and L. Olbrechts-Tyteca. 1958. *The new rhetoric: A treatise on argumentation*. Notre Dame: University of Notre Dame Press.

Polanyi, M. 1962. *Personal knowledge: Towards a post-critical philosophy*. Chicago: University of Chicago Press.

Rorty, R. 1982. *The consequences of pragmatism: Essays*. Minneapolis: University of Minnesota Press.

Rosen, S. 1980. *The limits of analysis*. New York: Basic Books.

Scott, R. 1967. On viewing rhetoric as epistemic. *Central States Speech Journal* 18 (1): 9–17.

Stigler, G.J. 1977. The conference handbook. *Journal of Political Economy* 85: 441–443.

Toulmin, S. 1958. *The uses of argument*. Cambridge: Cambridge University Press.

Ricardian Equivalence Theorem

Andrew B. Abel

Abstract

The Ricardian equivalence theorem states that government bonds and lump-sum taxes are equivalent means to finance government spending. Thus, a lump-sum tax cut financed

by the issuance of one-year government bonds would not affect consumption. Consumers could hold the newly issued bonds, and use them to pay the higher taxes when the government increases taxes to repay the principal and interest on the bonds. Intergenerational altruism implies that Ricardian equivalence holds even if the recipients of a tax cut die before future taxes are increased to fully repay the bonds. This article explores situations where Ricardian equivalence does or does not hold.

Keywords
Adverse selection; Altruism; Annuities; Bequest motive; Bequests; Bonds; Capital accumulation; Consumption function; Gift motive; Insurance markets; Intergenerational altruism; Life insurance; Liquidity constraints; Lump-sum taxes; Mortality; National income determination; Pensions; Precautionary savings; Public debt; Ricardian equivalence theorem; Taxation; Wealth

JEL Classifications
E6

The Ricardian equivalence theorem is the proposition that the method of financing any particular path of government expenditure is irrelevant. More precisely, whether government purchases are financed by levying lump-sum taxes or by issuing government bonds does not affect the consumption of any household, nor does it affect capital formation. In this sense, financing government purchases by lump-sum taxes is *equivalent* to financing these purchases by issuing bonds. The fundamental logic underlying this proposition was presented by David Ricardo in Chapter XVII ('Taxes on Other Commodities than Raw Produce') of *The Principles of Political Economy and Taxation.* Although Ricardo clearly explained why government borrowing and taxes could be equivalent, he warned against accepting the argument on its face: ' From what I have said, it must not be inferred that I consider the system of borrowing as the best calculated to defray the extraordinary expenses of the state. It is a system which tends to make us less thrifty – to blind us to our real situation' (1821, pp. 162–3).

The question of whether lump-sum taxes and government debt are equivalent arises in the specification of the consumption function. The aggregate consumption function plays an important role in models of national income determination, and aggregate consumption is often specified to depend on contemporaneous aggregate disposable income and on aggregate wealth. The question is whether the public's holding of bonds issued by the government should be treated as part of aggregate net wealth. Indeed, this is the eponymous question of Barro's (1974) classic article on Ricardian equivalence. If consumers recognize that government bonds, in the aggregate, represent future tax liabilities, then these bonds would not be part of aggregate wealth. If, on the other hand, consumers do not recognize, or for some reason do not care about, the implied future tax liabilities associated with these bonds, then they should be counted as part of aggregate wealth in an aggregate consumption function. Patinkin (1965, p. 289), citing Carl Christ and Christ's discussions with Milton Friedman, recognized this question and specified that a fraction k of the stock of outstanding government bonds is to be treated as wealth. Under the Ricardian equivalence view, k would equal zero; under the view that consumers ignore all future tax liabilities, k would equal 1. Bailey (1971) also examined the question of whether future tax liabilities affect aggregate consumption in a model of national income determination, though his formulation of the aggregate consumption function does not explicitly include aggregate wealth.

The question of whether government bonds are net wealth and the question of the effects of alternative means of financing a given amount of government expenditure are, in many contexts, basically the same question. For purposes of exposition, it is clearest to focus on one particular formulation of the question. The discussion here will focus on the question of the choice between current taxation and debt finance.

The underlying logic of the Ricardian equivalence theorem is quite simple and can be displayed by considering a reduction in current lump-sum taxes of 100 dollars per capita. This reduction in government tax revenue is financed by the sale of government bonds on the open market in the amount of 100 dollars per capita. For simplicity, suppose that the bonds are one-year bonds with an interest rate of five per cent per year. In addition, suppose that the population of taxpayers is constant over time. In the year following the tax cut, the bonds are redeemed by the government. In order to pay the principal and interest on the bonds, taxes must be increased by 105 dollars per taxpayer in the second year.

Now consider the response of households to this intertemporal rearrangement of their tax liabilities. Households can afford to maintain their originally planned current and future consumption by increasing their current saving by 100 dollars. In fact, the additional 100 dollars of private saving could be held in the form of newly issued government bonds. In the second year, when the government increases taxes by 105 dollars to redeem the bonds, households pay the extra tax using the principal and interest on the bonds. Thus, the originally planned path of consumption continues to be feasible after the tax change. In addition, since the originally planned path of consumption was chosen by the consumer before the tax change, it would continue to be chosen after the tax change since all relative prices remain unchanged. Therefore, household behaviour is invariant to the switch between tax finance and debt finance for a given amount of government spending.

In the basic example, the tax cut in the current year is financed by the issuance of one-year government bonds. However, the invariance result continues to hold if the current tax cut is financed by the issuance of N-year bonds. The argument is that once again each consumer uses the extra 100 dollars of disposable income in the first year to purchase 100 dollars of newly issued government bonds. If these government bonds pay interest in years before the bond is redeemed, then the government must increase lump-sum taxes in those years to service the bonds. Consumers who are holding the bonds and receive interest use the interest on their bonds to pay the increased taxes. Then, when the bonds mature after N years, each consumer uses the principal and final interest on these bonds to pay the higher taxes that are levied to redeem the debt. Once again, consumers can afford to maintain the originally planned path of current and future consumption and find it optimal to do so.

Having seen that the Ricardian equivalence theorem holds even if long-term bonds are issued to pay for the current tax cut, it is natural to ask whether the invariance result continues to hold even if some or all of the currently living consumers die before the bonds are redeemed. The first answer to this question would appear to be that consumers who are alive during the tax cut, but who die before the newly issued bonds are retired, would have a reduction in the present value of their taxes and thus an increase in the present value of their disposable income. Equivalently, such consumers could afford to increase their current and future consumption. It is not necessary for these consumers to hold the extra bond that is issued in the first year because they will not have to use the bonds to pay for the future tax increase needed to redeem the bonds. Therefore, these consumers would tend to increase their current and future consumption, *ceteris paribus.*

A self-interested consumer who receives a tax cut financed by government bonds will increase his consumption if he knows with certainty that he will die before future taxes are collected to fully repay the newly issued bonds. But if the consumer is uncertain about when he will die, the situation involves some additional considerations. I begin by ignoring survival-contingent assets such as annuities and life insurance, and I will assume that all consumers have positive net financial assets so that I can put aside issues related to borrowing costs for consumers who may die before repaying their loans. To keep the argument simple, suppose that lump-sum taxes are reduced in the current year by 100 dollars per taxpayer, and the government finances the tax cut by issuing 20-year zero-coupon bonds. Twenty years in the future the government will increase lump-sum taxes to pay off these bonds. The present value of the future tax increase is 100 dollars per current taxpayer. If the number of taxpayers 20 years in the future is the same as in the current year, then tax increase in the future will be $100(1 + r)^{20}$ per taxpayer, where r is the annual interest rate on the government bonds. In this case, the current tax cut will not affect the current consumption of any taxpayer. A current taxpayer could use the 100 dollars from the current tax cut to buy 100 dollars of government bonds, and simply plan to hold on to the bonds for 20 years. In the event that the consumer is still alive and consuming 20 years in the future, he can use his bonds, which will have grown in value to $100(1 + r)^{20}$ to pay the additional lump-sum taxes in that year, without changing consumption in that year (or in the current year). In the event that the consumer dies before 20 years elapse, he will, of course, consume the same level, namely zero, as in the absence of the tax cut. Thus, buying and holding 100 dollars of government bonds in the current year just allows the consumer to maintain consumption unchanged at all ages and in all states (that is, the state in which the consumer is alive in 20 years and the state in which he is not alive in 20 years). Therefore, Ricardian equivalence holds in this case.

The example in the preceding paragraph illustrates that Ricardian equivalence can hold even when selfish consumers receive a bond-financed tax cut and, with some unpredictability, die before the taxes are levied to fully repay the bonds. A crucial step in the argument is the assumption that the number of future taxpayers is the same as the number of current taxpayers. But with some taxpayers dying over time, the only way to maintain a constant number of taxpayers is for new taxpayers to arrive – through birth or immigration – at the same rate at which taxpayers are dying. In an economy with a growing population, that is, in an economy in which the sum of the birth rate and net immigration rate exceeds the death rate, the increase in future lump-sum taxes *per taxpayer* will be smaller than $100(1 + r)^{20}$, because the cost of paying off government bonds is spread among a larger number of taxpayers. Therefore, a consumer who receives a tax cut of 100 dollars in the current year will face a future tax increase that has a present value smaller than 100 dollars, and so will increase consumption in the current period (and in the future period, if he is alive). Alternatively, in an economy with a shrinking population, a lump-sum tax increase will have the opposite effect and will reduce current consumption. (An analytic version of this example is in Abel 1989.)

Ricardian equivalence is often illustrated in the context of perfect markets. If consumers face uncertainty about the length of their lives, and if they do not have bequest motives of any sort, they will want to hold annuities, which are assets that pay off if the owner of the annuity is alive, but pay zero if the owner is not alive. If all consumers face the same publicly known probability, p, of dying each year, then the actuarially fair annual gross rate of return on annuities will be $(1 + \mathrm{r})/(1 - P)$. If all consumers invest one dollar in an annuity that pays a lump sum in 20 years, then in 20 years each survivor will receive $[(1 + \mathrm{r})/(1 - p)]^{20}$ dollars. Whether consumers who receive a 100 dollar lump-sum tax cut in the current year will change their current consumption depends on the amount of the tax increase *per taxpayer* 20 years in the future when the bonds used to finance the tax cut are paid off. If the birth rate and the net immigration rate are both zero, then the population of taxpayers in 20 years will be a fraction $(1 - p)^{20}$ of the population in the current year. Thus, to repay the principal and interest on the 100 dollars of bonds issued per current taxpayer, the lump-sum tax will have to increase by $[(1 + r)/(1 - p)]^{20}$ dollars per current taxpayer. Thus, a current taxpayer could use the 100 dollar tax cut in the current period to purchase 100 dollars of annuities in the current period. If the consumer survives for 20 years, the payoff of the annuity, $[(1 + r)/(1 - p)]^{20}$, will be just sufficient to pay the increased lump-sum tax in that year. Thus, Ricardian equivalence holds in this case with perfect annuities and a zero birth rate and zero net immigration rate. Ricardian equivalence will fail to hold, however, if the birth rate is positive, because the tax burden in 20 years will be spread among a group of taxpayers consisting of surviving taxpayers from the current period plus additional taxpayers. In this case, the tax increase per future taxpayer will be smaller than $[(1 + r)/(1 - p)]^{20}$. Thus, recipients of the tax cut in the current period would be able to increase consumption in the current period somewhat and use the remainder of the tax cut to buy enough annuities to pay the increased lump-sum tax and to increase consumption in 20 years.

The examples with uncertain longevity illustrate that, as emphasized by Weil (1989), the departure from Ricardian equivalence does not result solely from the chance of dying before the future tax increase. In the examples in which the tax cut in the current year induces an increase in current consumption, the effect results from the fact that future increase in taxes per future taxpayer is smaller than the current tax cut per taxpayer. In the case of perfect annuities, this effect is made possible by a positive birth rate, and in the case without annuities the effect was made possible by growing population resulting from a death rate lower than the birth rate plus the net immigration rate.

Altruistic Consumers

If consumers are entirely self-interested, then escaping future taxes through death can lead to departures from the Ricardian equivalence theorem, as discussed above. However, Robert Barro (1974) presented an ingenious argument that extends the Ricardian equivalence theorem to situations in which consumers die before future taxes are increased to repay the bonds that are issued to finance the current tax cut. Before discussing the substantive content of Barro's argument, it is interesting to observe that the term 'Ricardian equivalence theorem' apparently was first used by James Buchanan (1976) in a published comment on Barro's paper. Buchanan's comment begins by pointing out Barro's failure to credit Ricardo with the idea that debt and taxes may be equivalent and, indeed, the comment is titled, 'Barro on the Ricardian Equivalence Theorem'. Previously, Buchanan had referred to this result as the 'equivalence hypothesis' (1958, p. 118).

Barro postulated that consumers have bequest motives of a particular form that has been labelled 'altruistic'. An altruistic consumer obtains utility from his own consumption as well as from the utility of his children. Therefore, a consumer who is altruistic toward all of his children cares not only about his own consumption but also indirectly about the consumption of all his children. Furthermore, if all of the altruistic consumer's children are also altruistic and care about the utility of all of their children, then the altruistic consumer cares indirectly about the consumption of all of his grandchildren. Provided that all consumers are altruistic, the argument can be extended ad infinitum with the important implication that an altruistic consumer cares, at least indirectly, about the entire path of current and future consumption of himself and all of his descendants.

Barro's insight that an intergenerationally altruistic consumer cares about the entire path of his family's consumption defuses the argument that consumers who know they will escape future taxes through death will increase consumption in response to a current tax cut. For altruistic consumers, it does not matter whether they themselves or their descendants pay the higher taxes necessary to pay the principal and interest on the newly issued bonds. In response to a 100 dollar tax cut in the current year, an altruistic consumer will not change his consumption but will hold an additional 100 dollars of government bonds. If the bonds are not redeemed until after the consumer dies, he will bequeath them to his children who can then use the bonds to pay the higher taxes in the year in which the bonds are redeemed, or else bequeath the bonds to their children if the bonds are not redeemed during their lifetimes.

The fact that a consumer leaves a bequest is not prima facie evidence that he is altruistic in the sense defined above. Bequests may arise as the accidental outcome of an untimely death or they may arise for motives other than altruism. For instance, if the utility that a consumer obtains from leaving a bequest depends only on the size of the bequest, then he will not care about tax increases that may be levied on his children or his children's children. In this case Ricardian equivalence would not hold.

The argument that each current and future consumer in a family of intergenerationally altruistic consumers cares about his own consumption as well as the consumption of all of his descendants for ever raises the question of whether the government must ever pay off the newly issued government bonds. If the government could roll over the principal and interest on this debt for ever, so that it would never be necessary to increase future taxes, it would seem that a current tax cut financed by issuance of government bonds would reduce the present value of the taxes paid by the current and future members of the family and hence would lead to an increase in the family's consumption. If the government attempted to roll over its debt each year by issuing new bonds, the quantity of these bonds would grow in perpetuity at the rate of interest. If the rate of interest exceeds the economy's growth rate, then these bonds would not willingly be held in private portfolios. Alternatively, if the rate of interest falls short of the economy's growth rate – a condition that signals an inefficient over-accumulation to capital – then, as pointed out by Feldstein (1976), it is possible for the government to roll over the debt permanently. Carmichael (1982) has shown that in this case the altruistic bequest motive will not be operative (that is, the non-negativity constraint on bequests will bind) but that an altruistic gift motive from children to parents (which specifies that a consumer's utility depends on his own consumption and the utility of his parents) may be operative. If the gift motive is operative, then Carmichael argues that Ricardian equivalence will hold, despite the fact that government bonds may be regarded as net wealth.

Departures from Ricardian Equivalence

Now that we have described a fairly general set of conditions under which Ricardian equivalence holds, it is useful to discuss several of the conditions that might lead to a violation of Ricardian equivalence. A clear overview of reasons why the Ricardian equivalence theorem may not provide an accurate description of the actual effects of debt finance vs. tax finance is provided by Tobin (1980).

The basic argument underlying the Ricardian equivalence theorem is that it makes no difference whether the government issues debt in the amount of 100 dollars per capita or whether it collects taxes of 100 dollars per capita since in the latter case consumers can borrow 100 dollars per capita to pay the higher taxes. In the former case, public borrowing is increased by 100 dollars per capita, and in the latter case private borrowing is increased by 100 dollars per capita. Under the appropriate conditions it makes no difference whether the borrowing is by the public sector or by the private sector. In order for the choice between debt finance and tax finance to have an effect, it must be the case that any changes in government borrowing cannot be fully offset by changes in private sector behaviour. Equivalently, there must be something that the government can do in credit markets that the private sector cannot do.

The government can borrow by issuing bonds, but in some situations consumers may not be able to borrow. For instance, a young consumer with a high prospective income might like to borrow to increase his consumption when young with the intention of repaying the loan when his income is higher in the future. However, for a variety of reasons it may simply not be possible for the young consumer to borrow the desired amount; if this is the case, the consumer is described as 'liquidity- constrained'. A liquidity-constrained consumer who receives a tax cut in the current period may choose to consume some, or even all, of the tax cut rather than save the entire tax cut. In effect, the current tax cut allows the consumer to borrow in order to increase current consumption, which is what the liquidity-constrained consumer wanted to do anyway. The current tax cut financed by an issue of government bonds can be viewed as the government borrowing on behalf of the consumer. Although this example makes it seem clear that a liquidity-constrained consumer would increase his current consumption in response to a current tax cut, some caution is required in interpreting this result. Unless the reason for the liquidity constraint is specified, one cannot determine what will be the effect of the tax cut. For example, suppose that a consumer is able to borrow some funds, but is liquidity-constrained in the sense that he would like to borrow even more funds. If his creditors determine how much they are willing to lend by looking at his ability to repay the loan, then, in response to the prospective tax increase accompanying the current tax cut, his lenders may reduce the amount they are willing to lend by the amount of the tax cut. In this case, Ricardian equivalence would hold.

The Ricardian equivalence theorem requires not only that consumers be intergenerationally altruistic, but that their bequest motives be operative in the sense that consumers can bequeath whatever amount they choose subject to their budget constraint. To be more precise, it is possible that an altruistic consumer may like to leave a negative bequest to his children, but he is constrained from leaving a bequest less than zero. The fact that a consumer may want to leave a negative bequest does not necessarily violate the assumption that the consumer is altruistic. It may be that the consumer's children will all be so much wealthier than the consumer that, even though the consumer cares about the utility of his children, he could achieve higher utility by taking some of his children's resources and consuming them himself. Formal conditions that imply that altruistic consumers would like to leave negative bequests have been presented by Drazen (1978) and Weil (1987). Under these conditions, if the consumer is constrained from leaving a negative bequest, he will instead leave a zero bequest. In such cases, a tax cut that is followed by a tax increase after the consumer's death will reduce the present value of the taxes paid by the consumer and he will increase his consumption. In effect, the current tax cut helps the consumer achieve the desired negative bequest by giving him current resources and taking resources away from his descendants.

Another reason for departure from the Ricardian equivalence theorem is that policy may redistribute resources among families that have different marginal propensities to consume. For instance, suppose that one half of the consumers receive a 200 dollar tax cut in the current year and the other half of the consumers have unchanged taxes in the current year. The government finances the tax cut by issuing bonds in the amount of 100 dollars per capita, and in the following period it redeems the bonds and pays the interest. For simplicity, suppose that the population is constant and that the interest rate on government bonds is five per cent per year. Then in the year following the tax cut there is a tax increase of 105 dollars per consumer. Finally, suppose that this tax increase is levied on all consumers equally. In this case, the tax cut in the current year redistributes resources from the consumers whose taxes are unaffected to the consumers whose taxes are reduced in the current year. The recipients of the transfer will increase their consumption and the other consumers will reduce their consumption. The reallocation of consumption across consumers may be viewed as a violation of Ricardian equivalence. Whether aggregate consumption rises or falls depends on the marginal propensities to consume of the recipients of the transfer compared with the marginal propensities to consume of the other consumers. If all consumers have equal marginal propensities to consume, then there will be no effect on aggregate consumption or capital accumulation. However, if, for instance, the recipients of the transfers have a higher marginal propensity to consume than the other consumers, then aggregate consumption would increase. It should be pointed out that, in some sense, this example does not represent a violation of the Ricardian equivalence theorem, because it ignores the possibility that there might exist an insurance market for individual tax liabilities. If there were such a market, then consumers could have insured themselves against the redistribution of taxes. Such markets do not generally exist, but whether the Ricardian equivalence theorem holds may depend on the reason why these markets do not exist.

To see the role of insurance markets in a different context, consider consumers who each contribute 100 dollars to a social security fund during their working life. Suppose that at the end of their working lives some of the consumers die and the others survive and live in retirement. Although the number of consumers who die at retirement may be predictable, the identities of those who will die are not predictable. The surviving retired consumers each receive an equal share of the social security fund (with accrued interest) to which they contributed while they were working. Each survivor's social security income is greater than the 100 dollars (plus interest) which he contributed, because the fund contains the contributions plus interest of his peers who died at the end of the working life.

Does the introduction of this type of social security system affect consumption and capital accumulation or does Ricardian equivalence imply that consumption and capital accumulation will be unaffected? To answer this question, it is useful to observe that this stylized social security system has the characteristics of an actuarially fair annuity. That is, consumers pay a premium when young (the social security tax) and receive a payment if, and only if, they survive to old age. Furthermore, if all consumers face the same probability of dying, the rate of return to the survivors is equal to the actuarially fair rate of return. If there were a competitive annuity market offering the actuarially fair rate of return, the social security system would have no effect on consumption or capital accumulation. Workers who are taxed 100 dollars are essentially forced to hold 100 dollars of the publicly provided actuarially fair annuity called social security; however, these consumers can afford, and will choose, to maintain their originally planned consumption and bequests by reducing their holdings of privately supplied annuities by 100 dollars. This reduction in the holding of private annuities allows consumers to re-establish their initial portfolios of annuities and other assets while maintaining consumption unchanged. Thus, the Ricardian equivalence theorem holds in this example, provided that consumers each originally planned to hold at least 100 dollars of private annuities.

If the probability of surviving until retirement differs across consumers, and if individual consumers are better informed about their own survival probabilities than are insurance companies, then the funded social security system described above will affect consumption. The reason is that, if an insurance company offered annuities at a price that would be actuarially fair to the average consumer, it would suffer from what is known as 'adverse selection'. As a simple example, suppose that insurance companies know the average mortality probability but have no additional information about the mortality probabilities of individual consumers. If an insurance company offered annuities at a price that would be actuarially fair to the average consumer, then consumers who believe they are healthier than average would view these annuities as a bargain; consumers who believe they are less healthy (or engage in more dangerous activities) than average would view these annuities as overpriced because these consumers have a smaller chance of living to reap the rewards. As the healthy consumers would buy a disproportionately large share of annuities, they would, on average, inflict losses on the sellers of these annuities and would induce these sellers to charge a higher price for annuities. However, the social security system can supply its annuities at the actuarially fair price for the average consumer because a compulsory social security system is immune to adverse selection. That is, because the government can determine the amount of the publicly provided annuity held by each individual, it does not have to worry that a disproportionately large share of annuities are held by healthy consumers. Therefore, as shown in Abel (1986), the annuity offered by the social security system would yield a higher rate of return than private annuities, or, equivalently, would be made available at a lower price to consumers. Because of the difference in the prices of the publicly provided and privately supplied annuities, consumers could not exactly offset the effects of social security by transacting in private annuity markets.

The example in which adverse selection leads to violation of the Ricardian equivalence theorem was constructed to obey the strict set of rules demanded by strong adherents to the view that the choice between debt finance and tax finance is irrelevant. In particular, the following assumptions were maintained: (a) consumers are forward-looking and understand that a bond-financed tax cut implies an increase in future taxes; (b) consumers have operative altruistic bequest motives so that they care about taxes after their death; (c) there is a complete set of competitive markets; and (d) only lump-sum taxes are changed. However, actual economies display several important departures from each of these assumptions. Violations of these assumptions are discussed below.

First, despite the widespread appeal of rational expectations in modern economics, it may simply be that consumers do not fully appreciate the link between a current tax cut and an increase in future taxes. If consumers did not understand this link at all, then a current tax cut would tend to increase current consumption.

Second, consumers may not have a bequest motive, either because they have no children or because they do not care about the welfare of anyone else. Even if consumers do have a bequest motive, it may not be operative as discussed above. Even if the bequest motive is operative, it may not be of the appropriate form for the Ricardian equivalence theorem to hold. If a consumer's utility depends directly on the size of the bequest he leaves rather than on the utility of his heirs, then a current tax cut followed by a tax increase on his heirs, would tend to raise the current consumption of the consumer. The reason is that he does not care about his heirs' utility per se. His bequest yields utility directly just as any other consumption good. As a result of the decrease in taxes he must pay over his lifetime, the consumer will have a higher level of lifetime income and can increase his own consumption and the bequest he leaves. If his own consumption and the bequest are both normal goods in his utility function, then he will choose to increase both.

Even if all consumers have operative altruistic bequest motives, a tax cut may increase current consumption. If all consumers have several children, but if each consumer cares about the utility of only one of his children, then there will be consumers in future generations whose utility is ignored by all current consumers. To the extent that future taxes are levied on these consumers, some part of future tax liabilities associated with a current tax cut will be ignored by current consumers. In this case, a tax cut would increase contemporaneous aggregate consumption.

Bernheim and Bagwell (1988) have challenged the plausibility of the assumption of intergenerational altruism by showing that this assumption leads to some untenable conclusions. If consumers A and B are unrelated to each other and both are altruistic toward consumer C, then A and B are effectively linked to each other, if both consumers A and B both plan to give positive transfers (bequests) to consumer C. For example, unrelated grandparents (consumers A and B) who plan to make positive transfers to their common grandchildren (consumers C) are effectively linked to each other. If the government transfers a dollar from consumer A to consumer B, these consumers can, and will choose to, undo this transfer and maintain their originally chosen patterns of consumption. The mechanism for undoing the government transfer is for consumer A to reduce his transfer to consumer C by one dollar and for consumer B to increase his transfer to consumer C by one dollar. Bernheim and Bagwell have argued that, if one takes intergenerational altruism seriously, such linkages are so widespread that all consumers are effectively linked to each other. In this case, all government transfers among consumers, including a transfer from future taxpayers to current taxpayers in the form of a bond- financed tax cut, would have no effect. Bernheim and Bagwell go on to show that even non-lump-sum taxes, and indeed prices themselves, would not affect consumption. Rather than conclude that all taxes and prices are irrelevant, Bernheim and Bagwell conclude that their findings cast doubt on the policy conclusions, including Ricardian equivalence, that are based on the assumption of intergenerational altruism.

Third, various types of insurance markets may be absent or, as described above, may suffer from adverse selection. Chan (1983) and Barsky et al. (1986) have argued that, if there are no markets for insuring against unpredictable fluctuations in after-tax income, then a current tax cut could increase current consumption. The argument, which was outlined by Barro (1974, p. 1115) and Tobin (1980, p. 60), is that to the extent that individual tax liabilities are proportional to income the tax system provides partial insurance against fluctuations in individual disposable income. Therefore, the increase in tax rates that follows a lump-sum tax cut in the current year will reduce the variability of future disposable income. The reduction in the riskiness of future disposable income reduces current precautionary saving that consumers undertake to guard against low future consumption. The counterpart of the reduction in precautionary saving is an increase in current consumption.

Fourth, most taxes are not lump-sum taxes. Generally, taxes are levied on economic activities, and changes in these taxes provide incentives to alter the levels of these activities. Although the existence of distortionary taxes does not in all cases imply that Ricardian equivalence is violated when applied to lump-sum tax changes, it does strain the interpretation of empirical tests of Ricardian equivalence that examine historical data on deficits and consumption.

As discussed above, there are many potential sources of departure from the Ricardian equivalence theorem, and ultimately the importance of these departures is an empirical question. The existing literature that attempts to test empirically whether Ricardian equivalence holds has produced mixed results, some claiming to show that it holds, and others the opposite. In judging the empirical relevance of the Ricardian equivalance theorem, however, the important question from the viewpoint of fiscal policy formulation is not whether the theorem holds exactly but whether there are departures from it that are quantitatively substantial. Existing empirical work has not yet produced a consensus on this question.

See Also

► Government Budget Constraint
► Public Debt
► Public Finance

Bibliography

Abel, A. 1986. Capital accumulation and uncertain lifetimes with adverse selection. *Econometrica* 54: 1079–1097.
Abel, A. 1989. Birth, death and taxes. *Journal of Public Economics* 39: 1–15.
Bailey, M. 1971. *National income and the price level*. 2nd ed. New York: McGraw-Hill.
Barro, R. 1974. Are government bonds net wealth? *Journal of Political Economy* 82: 1095–1117.
Barsky, R., G. Mankiw, and S. Zeldes. 1986. Ricardian consumers with Keynesian propensities. *American Economic Review* 76: 676–691.
Bernheim, B., and K. Bagwell. 1988. Is everything neutral? *Journal of Political Economy* 96: 308–338.
Buchanan, J. 1958. *Public principles of public debt*. Homewood: Richard D. Irwin.
Buchanan, J. 1976. Barro on the Ricardian equivalence theorem. *Journal of Political Economy* 84: 337–342.
Carmichael, J. 1982. On Barro's theorem and debt neutrality: The irrelevance of net wealth. *American Economic Review* 72: 202–213.
Chan, L. 1983. Uncertainty and the neutrality of government financing policy. *Journal of Monetary Economics* 11: 351–372.
Drazen, A. 1978. Government debt, human capital and bequests in a lifecycle model. *Journal of Political Economy* 86: 337–342.
Feldstein, M. 1976. Perceived wealth in bonds and social security: A comment. *Journal of Political Economy* 84: 331–336.
Patinkin, D. 1965. *Money, interest and price*. 2nd ed. New York: Harper and Row.
Ricardo, D. 1821. *The principles of political economy and taxation*, 1911. London: M. Dent and Sons.
Tobin, J. 1980. *Asset accumulation and economic activity*. Chicago: University of Chicago Press.
Weil, P. 1987. 'Love thy children': Reflections on the Barro debt neutrality theorem. *Journal of Monetary Economics* 19: 377–391.
Weil, P. 1989. Overlapping families of infinitely-lived agents. *Journal of Public Economics* 38: 183–198.

Ricardian Socialists

Andrea Ginzburg

The name 'Ricardian Socialists' was given currency by H.S. Foxwell. Introducing the English translation of a work by the Austrian jurist Anton Menger in 1899, he complained that 'the important work' of the Ricardian Socialists had been almost wholly ignored 'until the last few years'. Since that time the name has traditionally been used to refer to certain authors in England, especially between 1820 and 1830, and to a lesser extent in the decade following, who claimed that the workers had a right to the entire product of their labour. Starting from their observation of a contradiction, they developed a critique of existing distribution (with implications for development potentialities). On the one hand, it was admitted that labour was the sole source of value. On the other hand, they observed that much of the product of labour – that part which exceeds the 'necessary consumption' of the labourer – is taken, in the form of rent, profit and taxes, by the owning classes who contribute nothing to production.

Following M. Beer (1919), within this 'movement', whose members were in some respect very different from one another, we may distinguish two groups. On the one hand, there were the 'anti-capitalist economists', like T. Hodgskin, P. Ravenstone (pseudonym of Richard Puller) and some anonymous authors of pamphlets (such as *The Source and Remedy* etc., 1821). On the other hand, there were the 'cooperative socialists' close to Owen, such as W. Thompson, J. Gray and J.F. Bray. The distinction – with the further qualification that Bray's work can be viewed as a moment of synthesis between the two currents – is useful for tracing the broad outlines of the cultural and political background against which these authors stand. And it may also indicate that it is, above all, those belonging to the first group who are of greater interest from the point of view of economic analysis. With their greater interest in pursuing economic themes, these authors show a closer adherence to the mainstream of classical political economy, both as regards the structure of their analysis and its subject matter. This obviously holds true even when the motive behind their writings is frequently – as Foxwell (1899, p. lxxiii) also reminds us – that of opposition to the conclusions reached by Ricardo or other classical economists.

No consensus of opinion exists concerning the importance of these authors' contribution to economic analysis or the Ricardian (or at any rate 'classical') basis of their thought. Marx, among others, devoted several pages of his *Theories of Surplus-Value* to a careful analysis of the writings of the first of the two groups distinguished above (Marx 1910, Part III, ch. XXI). Schumpeter, however, held that the writings of the group called Ricardian Socialists 'which of course is entitled to a great place in the history of socialist thought, offer but little that is relevant to a history of economic analysis' (Schumpeter 1954, p. 479).

Introducing a discussion of these authors, Marx stated that their work derived entirely 'from the Ricardian form'. The link between Ricardo and the English protosocialists was recognized by Foxwell but has, more recently, been denied by Hollander (1980) (but for hints in this direction see Lowenthal 1911, p. 103; Blaug 1958, p. 148; Hutchison 1957, p. 89 and 1978, p. 242; in a very different context from these, Hunt 1977, and 1979, p. 149). In the introduction from which we have quoted, Foxwell took the opportunity to unleash a fierce attack on Ricardo, employing the methodological critique of the German Historical School as well as the judgement of Jevons ('Ricardo's crude generalisations . . . gave modern socialism its fancied scientific basis and provoked, if they did not justify, its revolutionary form'). If instead of being neglected for half a century the work of the Ricardian Socialists had been subjected to 'searching criticism by the best economists of the time' – he wrote – the 'elementary blunder in method' of Ricardo would have been more promptly rectified (Foxwell 1899, pp. xl-xli). Hollander, however, in a study of the writings of Hodgskin, the most important and influential of the Ricardian Socialists, maintains that, on the one hand, the 'vehemence' of his critique of Ricardo, and on the other, his attachment to the so-called 'adding up theory' of value of Smith – which Ricardo rejected – make it impossible to rank Hodgskin in the Ricardian tradition.

1. Hodgskin and the other Ricardian socialists stand at an important crossroads in the development of British economic thought in the 19th century. Thus, aside from their intrinsic interest, their writings can also be studied for the support they lend to one or another of the three main interpretations that have been offered of that development, i.e. those of Marx, Jevons and Marshall. According to Marx (1873, vol. I, pp. 14–15), the end of classical political economy (which comes after Ricardo's death) and, parallel with this, the progressive search for and propagation of alternative theories of

profit, must be traced back – together with the problems of analysis left unsolved by Ricardo – to the employment of the Ricardian theory 'as a weapon against bourgeois economy' used by the Labour writers. They had undermined the possibility of founding on the basis of that theory a 'harmonious' vision of capitalist society which would enable it to go on being considered as a 'definitive form' of social production. The anti-Ricardian reaction of the 1830s and 1840s – set against a background in which class struggles were becoming ever sharper and more extensive – must thus be seen, according to Marx, as a reaction to the '*unpleasant* side of classical political economy' (Marx 1910, p. 502) which these authors had brought out.

Subsequently, Marx's interpretation has found support in Meek's (1950 and 1967) survey of the work of Read, Scrope and Longfield. According to Dobb (1973, p. 166), the 'anti-Ricardian reaction' of the 1830s and 1840s to some extent found its fulfilment in the analogous 'reaction' in the late 19th century associated with Jevons and his followers, among them Foxwell himself, which led to the 'rediscovery' of the Ricardian Socialists. Hollander's recent attempt to deny a Ricardian basis of Hodgskin's writings tends to suggest that after Ricardo's death not only would there have been no grounds for reacting against a 'dangerous' use of the Ricardian theory, but also that there was apparently no such reaction. In this view, Ricardo stands in a continuous line of development carrying the *entire* current of British economic thought from the first half of the 19th century up to Marshall and, more generally, to the marginalist theories of the last quarter of the 19th century.

2. If we except Thompson, who stands very close to Bentham's utilitarianism, even if in a contradictory way, all the other Ricardian Socialists argue from a premise of Natural Law that reflects the influence of Locke, Adam Smith and Godwin (and thus Rousseau). Whether dealing with the nature and origin of capital (Ravenstone and Hodgskin), the theory of value and distribution (Hodgskin) or the obstacles by which profit hinders accumulation (Hodgskin and some anonymous authors of *pamphlets*), the arguments start from very similar logical schemes. The necessary starting point is held to be 'first principles', the very first of which states that 'all wealth is the produce of labour'. On the basis of this principle (to which others are added, as in the case of Bray 1839), a scheme is erected which is supposed to represent the workings of a 'natural society'. This is then contrasted with a representation of actual society. (The statistical foundation of this representation is often the 'Map of Civil Society' showing the distribution of income among 'different Classes' prepared by Colquhoun 1814). The divergences of the second scheme from the first are explained by the presence of 'artificial' components, not intrinsically necessary and thus susceptible of modification, resulting from man-made institutions or contingent historical events. The contrast between 'nature' and 'artifice' is used by these writers to criticise both society and the economists who 'erected the results of their individual experience into general laws. Because a thing was, they thought it could not be otherwise' (Ravenstone 1824, p. 6; see also Anon. 1821, pp. 7–8). Thus we have here not a 'positive' idea of Natural Right but, in the manner of Rousseau, a 'normative' one.

The contrast between 'nature' and 'artifice' finds an application in the idea, shared by many Ricardian Socialists, that in capitalist society the appropriation of the produce of labour takes place through a violation of the 'natural' principle of exchange, according to which each party should give and receive equal quantities of labour. However, there is an important difference between Gray, for instance, and Hodgskin (and Bray). (This difference is much fainter in popular proto-socialist literature of the 1830s.) In Gray's view, the violation emerges exclusively and directly from the comparison between a situation in which profit is absent and a capitalist situation. 'We have endeavoured to show that the real income of the country', Gray writes, '. . . is taken from its producers, chiefly by the rent of land, by the rent of houses, by the interest of money, and by the profit *obtained by persons who buy their labour from them at one price and sell it at another*' (1825, p. 58, italics added). The conception of 'profit upon alienation', like that of the 'adding up theory', is, instead, lacking in Hodgskin. In the latter's conception the violation of the 'natural' principle of exchange does not derive directly from the presence of profit; it emerges from the comparison between exchanges effected *between capitalists,* on the one hand, and *between capitalist and labourer,* on the other. The former exchanges, according to the labour theory of value, are in agreement with the 'natural' principle of exchange; in the case of the latter exchanges, the violation is shown, as we shall see, by the difference between labour commanded and labour embodied.

3. As is well known, the theory of Natural Right has historically performed the function of justifying the application to the realm of history of the same conceptual tools that physicists had used to study the realm of nature. Reference to 'natural laws' was aimed not only at applying moral criteria but also reflected a search for phenomena endowed with 'universality and uniformity'. From the point of view of analysis, the theory of Natural Right, with all its limitations, fulfills the function of a 'counterfactual'; in other words, it enables the observer to stand back from existing society, a position which is in any case necessary in order to analyse the working of that society in depth.

In this connection, Hodgskin represents the most complex personality among the Ricardian Socialists. In his thought, the theory of Natural Right is grafted onto two very important cultural influences: on the one hand, that of the 'Scottish Historical School' (in his writings he quotes freely from Millar, Robertson and Lord Kames); on the other, that of Thomas Brown, an exponent of the Scottish philosophy of Common Sense (see in particular Hodgskin 1827b, 1832). In Hodgskin these influences lead to a peculiar combination of a 'naturalistic optimism' and a philosophy of history based on materialism. Indeed Hodgskin holds that, in the long term, the 'material world' governs the formation of 'beliefs' since experience eventually leads to the correction of mistaken 'beliefs'. At the foundations of progress in knowledge and inventions, which, in turn, is the cause of a 'perpetual' productivity increase, lies a 'natural' phenomenon, the growth of population. 'Necessity is the mother of invention; and the continual existence of necessity can only be explained by the continual increase of people' (Hodgskin 1827a, p. 86; see also Hodgskin in Halévy 1903, p. 77; Ravenstone 1821, p. 177). Since the growth of population, as Smith had argued, also leads to an extension of the division of labour, it ends by governing the development of society, independently of men's intentions and desires. The definition of 'natural' right of property accepted by Hodgskin comes from Locke. At one point in his *Two Treatises* (1690, Bk. II, ch. V, par. 27), Locke had stated that since each man owns his body, he also owns 'the *Labour* of his body' and 'the *Work* of his Hands'. The infringement of this law as of every other 'natural law' by 'artificial' institutions not only violates justice but also, by indirectly slowing down the growth of population, hinders the general progress of society.

As Halévy has noted (1903, p. 59), Hodgskin ascribes the same importance to population as does Malthus, but in a positive way. From the very outset he criticizes the *Malthusian elements* in the writings of Ricardo, expressing the hope that by getting rid of them one will be able to return to Adam Smith. As we shall see, the disagreement with Ricardo was not over his theory of value. As regards the theory of distribution, he argued that the current level of wages and the existence of an 'absolute' component in rent have an historical origin, to be sought in the 'power over labour' which landowners all over Europe have inherited from the previous state of slavery; he concluded that distribution depends 'entirely and exclusively on political regulations' (Hodgskin in Halévy 1903, p. 78). Starting from a critique of the relevance of 'decreasing soil fertility', he proposes a different analysis of the relation between the growth of production (and population) and technical conditions in the production of necessaries. He studies firstly the effects of this relation on distribution, then its effects on development with a given distribution. The basic analytical tool used in this theoretical extension is no less than the Ricardian theory of value.

4. In Hodgskin's most widely read work, *Labour Defended* (1825), he re-expounds the Ricardian theory of wages, profits and differential rents, stating that the theory 'of that ingenious and profound writer' confirms that 'the exactions of the capitalist cause the poverty of the labourer' (Hodgskin 1825, pp. 80–81). Yet in an important letter of May 1820 (first printed by Halévy in 1903), in setting out his opinions on value and distribution, Hodgskin had attacked Ricardo with criticisms of a fiercer and more explicit nature than can be found anywhere else in his work. Since in his subsequent writings his opinions remained substantially unaltered (except to give less attention to the problem of rent), a correct interpretation of that letter is crucial for defining Hodgskin's position. In 1820 he held that Ricardo, unlike Smith, had not made a clear distinction between 'natural' and 'artificial' circumstances. This had led him to mistake the latter for the former and to make incorrect forecasts. Consistent with his general conceptions, the main target of Hodgskin's critique is Ricardo's conclusion that, as population grows, profits tend to fall and development becomes stationary. Thus, on the basis of assumed 'natural laws', Ricardo is accused of having 'set bounds to our hopes for the future progress of mankind in a more definite manner' even than Malthus (Halévy 1903, p. 67).

   Hodgskin's most notable contribution (mentioned below) makes it idle to argue whether he should be labelled as 'Smithian' as opposed to 'Ricardian': he must be viewed as Ricardian reinterpreter of Smith and thus, more simply, as a 'classical political economist'. The contribution referred to is contained in his discussion of the difference of opinion between Smith and Ricardo on the subject of value. He holds that it lies not in the regulator principle of exchange – since both of them see value as determined by labour embodied – but in the different standards adopted for measuring prices. In asserting this Hodgskin clearly dissociates himself from Ricardo's interpretation of Smith, and thus also from the so-called 'adding up' theory.

   According to Ricardo (1821, p. 13), Smith had confined the validity of the labour theory of value strictly to the 'early and rude state of society which precedes both the accumulation of stock and the appropriation of land'. But the entry of profit onto the scene necessitated that the theory be abandoned. In this situation, the natural price would be obtained as the sum of the 'component parts', at their respective natural rates, taken independently of one another. A corollary of this 'adding up' theory was that an increase in profit would determine a rise in prices without any corresponding fall in wages. Ricardo rejected this conclusion and asserted, on the basis of the labour theory of value, that relative prices were independent of variations in distribution: every increase in profits was offset by a fall in wages. Smith, in abandoning (according to Ricardo), the labour theory of value in favour of the 'adding up' theory, had simultaneously taken wages as the unit for measuring prices. This suggested to Ricardo the existence of a correspondence between the 'regulator principle of exchange' and the unit for measuring prices. (He may also have been led into this mistake by the experience with his own theory in the course of his search for an invariable measure of value: see Sraffa 1951, p. xli, note 1.) Thus he charged Smith with contradicting the rule of exchange according to the labour embodied *because* he had adopted the labour commanded as the unit for the measurement of prices.

   Actually Hodgskin rejects this since he reinterprets Smith's text on the basis of the inverse relation of wages to the rate of profit which Ricardo had derived from the labour theory of value. As the rate of profit increases – says Hodgskin (Halévy 1903, p. 74) – price rises *in terms of wage units,* and therefore the labourer must perform a larger quantity of labour in order to purchase the same quantity of goods as before, because wages have fallen. To this fall in wages (rise in profit) he thus traces the divergence between labour embodied and labour commanded which Smith detected when the rate of profit is positive and accounted for, unlike Hodgskin, with the 'adding up' theory. (According to this theory, in any case, reductions in wages should lead to *reductions* in prices.) Hodgskin contrasts exchange between owners of the means of production, on the one hand, and, on the other, between commodities produced under capitalistic condition and labour. He accepts the hypothesis of uniformity in the ratio of profits (and rents) to wages in the price of each commodity (the hypothesis is set out explicitly in Hodgskin 1827, p. 186, and implied in all his writings); this explains why in Hodgskin's view incomes other than wages 'do not enter' into the relative prices of commodities, whereas they do enter into the ratio of prices to wages, and indeed 'constitute the greatest part of it'.

In 1846, reviewing Ricardo's works, Hodgskin (1846, p. 1557), was to admit that Smith had made 'a verbal variation from his own principle' of 'labour paying all price', whereas Ricardo had maintained 'a technical adherence to it'. But he justified Smith as having tried to offer through the exchange of commodity and labour 'a truer representation of what actually occurs in society than Mr. Ricardo's'. (Hodgskin holds that Ricardo has focused his attention on a relatively minor problem, that of 'exchangeable variations in the value of commodities', instead of dealing, as Smith had tried to, with the 'important relations of the labourer to other classes'.) The result Hodgskin achieved by basing himself on the labour theory of value retains its validity – as Sraffa has demonstrated (1960, ch. VI) – within the framework of a rigorous theory of the prices of production. From his 'equations of reduction to dated quantities of labour' it can immediately be seen that when the rate of profit is zero, embodied labour and labour commanded coincide. When the rate of profit increases (i.e. when wages fall in terms of the price of products) the quantity of 'labour commanded' by each commodity increases and is greater than the quantity of 'labour embodied'. Since we are dealing with a 'pricerelation between labour and the given product', Sraffa has remarked (1960, p. 40), this is independent of the 'medium' adopted as a measure of wages and prices.

5. Hodgskin holds that the measurement of prices in terms of labour commanded is important, among other things to show up the mistake in Ricardo's thesis on the 'natural' tendency of profits to fall as the growth of population requires the cultivation of less and less fertile land. He states that the direct and indirect application of 'machinery and ingenuity' to agricultural production has in actual fact *reduced* the quantity of labour embodied in each unit of production (defined as 'natural price'). What, on the contrary, has steadily grown, in the long run, is labour commanded. (Price measured in wage units is defined as 'exchange value' in *Labour Defended,* and as 'social price' in *Popular Political Economy.*) Ricardo has been deceived by the missing distinction between the two 'prices' into underestimating the long-term trend of technical progress, induced by the growth of population, to 'compensate for decreasing fertility'. This technical progress has led to an increase in the ratio between surplus and wages, and thus to an increase in labour commanded. In addition, Ricardo has drawn general conclusions on technology on the basis of the increase in the price of corn occurring in 'a short and single period' (after 1792). But this period, Hodgskin asserts, has been strongly affected by a series of exceptional circumstances and/or 'political regulations' (see Hodgskin 1827a, pp. 226–31 and Hodgskin 1848, p. 1228).

Though Hodgskin's disagreement with Ricardo on this point is important in several respects, it does not concern the theoretical structure of the theory of value and distribution. Rather, it offers an instance of the flexibility injected into the analytic structure of the surplus theories by the separate determination of production, on the one hand, and distribution, on the other (see Garegnani 1984, pp. 296–7). This separation enables various hypotheses about the shape of the relationship between levels of production and returns to be considered. In 1846, Hodgskin was to evince astonishment at Ricardo's disregard of the 'laws of production' especially in a period of rapid advance in output; in this disregard he found one of the reasons for the decline of the Ricardian theories. On the other hand, in a famous letter to Malthus of 9 October 1820, Ricardo had written: 'no law can be laid down respecting quantity, but a tolerably correct one can be laid down respecting proportions' (Ricardo 1887, p. 278) – which may perhaps help to explain the meaning he attached to his acceptance of Say's principle.

6. In *Popular Political Economy,* Hodgskin states that the difference between 'natural price' and 'social price' is important not only 'to understand the natural laws which regulate the progress of nations', but also 'rightly to estimate the causes which retard it' (1827a, p. 220). He denies that redistribution in favour of profits and the very presence of profit itself promote development. First of all, that which enriches the individual capitalist, he writes, does not necessarily add to national wealth: not all the capital which brings profit to its owner 'assist production'. Moreover, the presence of profit requires a part of the product of labour be withdrawn from reproduction and handed over to 'unproductive idlers'. The capitalist neglects those investments that do not promise him sufficient profit, yet these same would provide labourers with a comfortable subsistence. Hodgskin does not deny that profits have a periodic tendency to shrink. In *Labour Defended* (1825, pp. 78–80) his explanation explicitly contradicts those of Smith and Ricardo. The fall in profits is ascribed to the need to balance, periodically, two contrasting forces. On the one hand, the need to obtain a rate of profit not less than the monetary rate of interest prompts capitalists to a continual reinvestment of profits, thus causing a continuous growth, at compound rate, of the bulk of profits (for a given labour force). The appropriation by capitalists of the fruits of technical progress may lead to an increase in the rate of profit which contributes to this growth. On the other hand, reinvestment is not concerned with 'gold or money, but food, clothing and instruments', and labour productivity can-increase continuously so as to satisfy 'the overwhelming demands of compound interest': there is thus a limit to the growth of profits. Hodgskin's thesis may be set alongside Marx's idea that the degree of exploitation of labour has unsurmountable limits bound up with the length of the working day. This idea led Marx to conclude that an increase in the rate of surplus-value could not in the long term determine a counter-tendency to the fall of the rate of profit. Hence Marx's simplifying hypothesis of the constancy of rate of surplus-value, so that his law of the tendency of the rate of profit to fall ultimately depends solely on the increase in the organic composition of capital (see Marx 1910, pp. 298–311 and Marx 1894, pp. 211–66).

7. A recurring theme in the writings of the Ricardian Socialists is the polemic against the thesis, originating with Lauderdale and Say – and partly accepted in revised versions by disciple of Ricardo – which casts capital in the role of an 'active agent' of production. According to these theories, capital is capable of increasing productivity and/or saving labour independently of the application of labour. *Labour Defended* (along with Ravenstone's book, 1821) appears to offer the most coherent and effective arguments, at the time, against such conceptions of 'economic fetishism'. In this work, circulating capital is traced back to 'coexisting labour', while fixed capital is identified with the knowledge and the skilled labour needed to construct and employ the instruments of production. Ultimately, capital for Hodgskin (and here we catch an echo of Smith) is 'a means of obtaining command over labour' (Hodgskin 1825, p. 55). The arguments Hodgskin uses to demonstrate that capital enjoys no independent productivity are, however, mostly confined to the aspect of use-value. There are, moreover, two features

which explain Hodgskin's tendency, noted by Marx, to 'underestimate somewhat the value which the labour of the past has for the labour of the present' (Marx 1910, p. 276). On the one hand, the want of a clear distinction (this in common with other classical economists) between 'concrete labour' and labour as 'value magnitude' which, on the basis of the accepted theory of value, determines the exchange value. This also explains Hodgskin's reluctance to admit the influence of soil fertility on production, to the extent that he considers it 'a contradiction of Mr. Ricardo's own principle' that 'labour pays all cost' (Hodgskin 1848, p. 1228). On the other hand, Hodgskin was attached to Smith's idea (restated by James Mill) that 'what is *annually* produced is *annually* consumed' (Hodgskin 1825, p. 47). However, Marx's main objection to the Ricardian Socialists, reiterated against Lassalle in the *Critique of the Gotha Programme,* turns on the fact that by proposing a society governed by individual exchanges between independent producers, their critique of capitalism stopped short of discussing the market as a coordinating mechanism in the social division of labour. Some hints of criticism of competition do nevertheless occur in the writings of the Owenite current of Ricardian Socialists and, in particular, in the works of Thompson (1824 and 1827) and Gray (1825).

See Also

- ▶ Bray, John (1809–1897)
- ▶ Hodgskin, Thomas (1787–1869)
- ▶ Ravenstone, Piercy
- ▶ Thompson, William (1785–1833)

Bibliography

Anon. 1821. *The source and remedy of the national difficulties. A letter to John Russell*. London: Rodwell and Martin.

Beer, M. 1919. *A history of British socialism*. London: Allen – Unwin.

Blaug, M. 1958. *Ricardian economics*. New Haven: Yale University Press.

Bray, J.F. 1839. *Labour's wrongs and labour's remedy.* Leeds. Reprints of Economics Classics, New York: A.M. Kelley, 1968.

Colquhoun, P. 1814. *A treatise on the wealth, power and resources of the British empire*. London: J. Mawman.

Dobb, M.H. 1973. *Theories of value and distribution since Adam Smith*. Cambridge: Cambridge University Press.

Foxwell, H.S. 1899. Introduction to A. Menger (1886).

Garegnani, P. 1984. Value and distribution in the classical economists and Marx. *Oxford Economic Papers* 36(2): 291–325.

Gray, J. 1825. *A lecture on human happiness*. London. Reprints of Economic Classics. New York: A.M. Kelley, 1971.

Halévy, E. 1903. *Thomas Hodgskin*. Edited in translation with an introduction by A.J. Taylor. London: E. Benn, 1956.

Hodgskin, T. 1825. *Labour defended against the claims of capital,* by a Labourer. London. Reprints of Economic Classics. New York: A.M. Kelley, 1963.

Hodgskin, T. 1827a. *Popular political economy.* London. Reprints of Economic Classics. New York: A.M. Kelley, 1966.

Hodgskin, T. 1827b. *The word belief defined and explained*. London: Charles Tait.

Hodgskin, T. 1832. *The natural and artificial right of property contrasted,* by the author of *Labour defended against the claims of capital*. London.

Hodgskin, T. 1846. Review of *The works of David Ricardo,* ed. J.R. McCulloch. *The Economist* 28.

Hodgskin, T. 1848. Review of H.C. Carey, *The past, the present, and the future. The Economist* 28.

Holander, S. 1980. The post-Ricardian dissension: A case study in economics and ideology. *Oxford Economic Papers* 32(3): 370–410.

Hunt, E.K. 1977. Value theory in the writings of the classical economists, Thomas Hodgskin and Karl Marx. *History of Political Economy* 9(3): 322–345.

Hunt, E.K. 1979. *History of economic thought: A critical perspective*. Belmont: Wadsworth.

Hutchison, T.W. 1957. Review of E. Halévy (1903), ed. A.J. Taylor. *Economica* 24: 88–89.

Hutchison, T.W. 1978. *On revolutions and progress in economic knowledge*. Cambridge: Cambridge University Press.

King, J.E. 1983. Utopian or scientific? A reconsideration of the Ricardian socialists. *History of Political Economy* 15(3): 345–373.

Locke, J. 1690. In *Two treatises of government*, ed. P. Laslett. Cambridge: Cambridge University Press, 1967.

Lowenthal, E. 1911. *The Ricardian socialists*. Reprints of Economic Classics. New York: A.M. Kelley, 1972.

Marx, K. 1873. Afterword to the 2nd German ed, in K. Marx, (1867) *Capital. A critique of political economy,* vol. I. New York: International Publishers, 1967.

Marx, K. 1894. In *Capital. A critique of political economy*, vol. III, ed. F. Engels. New York: International Publishers, 1967.

Marx, K. 1910. *Theories of surplus value, Part III*, , 1972. London: Lawrence & Wishart.

Meek, R.L. 1950. The decline of Ricardian economics in England. *Economica* 17: 43–62.
Meek, R.L. 1967. The decline of Ricardian economics in England. In *Economics and ideology and other essays*, ed. R.L. Meek. London: Chapman – Hall.
Menger, A. 1886. *The right to the whole produce of labour*. With an introduction and bibliography by H.S. Foxwell. London. 1899. Reprints of Economic Classics. New York: A.M. Kelley, 1962.
Ravenstone, P. 1821. *A few doubts . . . on the subjects of population and political economy.* London: Andrews. Reprints of Economic Classics. New York: A.M. Kelley, 1966.
Ravenstone, P. 1824. *Thoughts on the funding system and its effects*. London: Andrews.
Ricardo, D. 1821. *On the principles of political economy and taxation*. 3rd ed in *The works and correspondence of David Ricardo,* ed. P. Sraffa with the collaboration of M.H. Dobb, vol. I. Cambridge: Cambridge University Press, 1951.
Ricardo, D. 1887. *Letters 1819-June 1821*. In *The works and correspondence of David Ricardo,* ed. P. Sraffa, vol. VIII. Cambridge: Cambridge University Press, 1973.
Schumpeter, J.A. 1954. *History of economic analysis*. New York: Oxford University Press.
Sraffa, P. 1951. *Introduction to The works and correspondence of David Ricardo*, vol. I. Cambridge: Cambridge University Press.
Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.
Stark, W. 1943. *The ideal foundations of economic thought*. London. Reprints of Economic Classics. Fairfield : A.M. Kelley, 1976.
Thompson, W. 1824. *An inquiry into the principles of the distribution of wealth most conducive to human happiness*. London. Reprints of Economic Classics. New York: A.M. Kelley, 1963.
Thompson, W. 1827. *Labour rewarded. The claims of labour and capital conciliated, by one of the idle classes*. London. Reprints of Economic Classics. New York: A.M. Kelley, 1966.

Ricardian Trade Theory

Kiminori Matsuyama

Abstract

Ricardian trade theory takes cross-country technology differences as the basis of trade. By abstracting from the roles of factor endowment and factor intensity differences, which are the primary concerns of factor proportions theory, Ricardian trade theory offers a simple and yet powerful framework within which to examine the effects of country sizes, of technology changes and transfers, and of income distributions. Moreover, its simple production structure makes it relatively easy to allow for many goods and many countries, and hence capable of generating valuable insights which are lost in the standard two-country, two-sector model of international trade.

Keywords

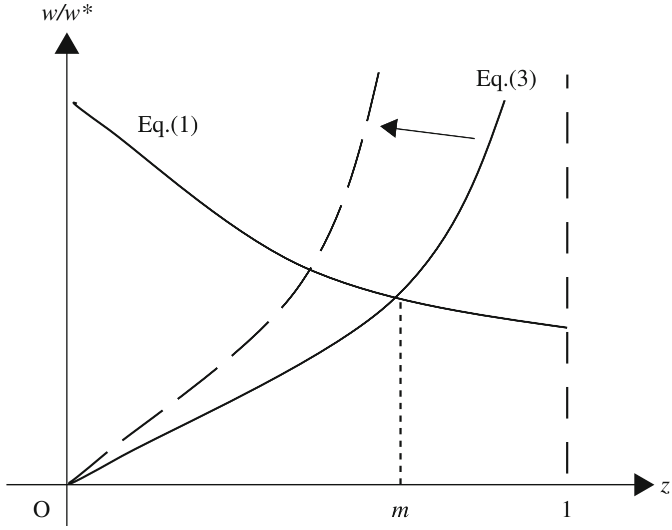
Absolute advantage; Cobb–Douglas functions; Comparative advantage; Constant returns to scale; Diminishing returns to scale; Export-based and import-biased technology; Export subsidies; Factor endowment; Factor intensity; Factor proportions; Gains from trade; Globalization; Iceberg costs; Immiserizing growth; Increasing returns to scale; International trade; North–South trade; Product cycles; Productivity growth; Ricardian trade theory; Size of nations; Specific factor models; Structural change; Technical change; Terms of trade; Trade costs; Trade, technology diffusion and growth; Transfer of technology; Transfer problem

JEL Classifications

F

Ricardian trade theory takes cross-country technology differences as the basis of trade. By abstracting from the roles of cross-country factor endowment differences and cross-industry factor intensity differences, which are the primary concerns of factor proportions theory (such as Heckscher–Ohlin and Specific Factor models), Ricardian trade theory offers a simple and yet powerful framework within which to address many positive and normative issues of international trade. It is particularly well-equipped to examine the effects of country sizes, of technology changes and transfers, and income distributions. Furthermore, its simple production structure makes it relatively easy to allow for many tradable

Ricardian Trade Theory, Fig. 1 The equilibrium factor terms of trade and patterns of specialization



goods and many countries, and hence capable of generating valuable insights, which are lost in the standard two-country, two-goods model of international trade.

Let us start with the Ricardian model with a continuum of tradable goods, adopted from Dornbusch, Fischer and Samuelson (DFS) (1977). The world consists of two countries, Home and Foreign. There is a continuum of competitive industries, indexed by $z \in [0, 1]$, each producing a homogenous tradable good, also indexed by z. There is only one non-tradable factor of production, called labour. (Or, if there are many non-tradable factors, they can be aggregated into a single composite factor.) Let $a(z)$ and $a^*(z)$ be the Home and Foreign unit labour requirements of good z, that is, labour input required to produce one unit of output z at Home and Foreign. Without loss of generality, we can index z so that Home's relative efficiency, $A(z) \equiv a^*(z)/a(z)$, is non-increasing in z. In Fig. 1, it is strictly decreasing. In short, Home (Foreign) has a *comparative advantage* in low-indexed (high-indexed) goods.

Let w and w^* denote the wage rates at Home and Foreign. Then, the prices in autarky are given by $p(z) = wa(z)$ at Home and $p^*(z) = w^*a^*(z)$ at Foreign. Under free trade (and in the absence of any trade costs), the price of each good is equalized across the two countries and is given by $p(z) = p^*(z) = \min\{wa(z), w^*a^*(z)\}$. Then, for a given relative wage rate or a given level of the *factoral terms of trade,* w/w^*, there is a marginal good, m, defined by

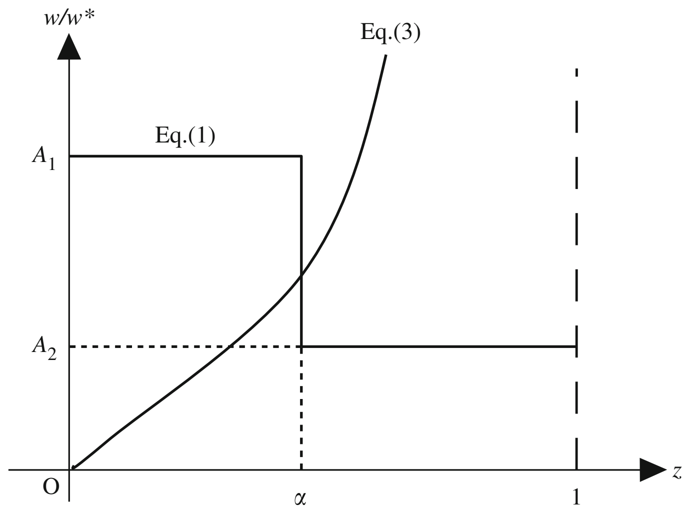
$$\frac{w}{w^*} = A(m), \tag{1}$$

such that Home produces only goods in $[0, m]$, and Foreign produces only goods in $[m, 1]$, and the prices become

$$p(z) = p^*(z) = wa(z), \ \ z \in [0, m]; p(z) = p^*(z) = w^*a^*(z), \ \ z \in [m, 1]. \tag{2}$$

To pin down the relative wage rate, we must specify the demand conditions. To keep it simple, let us assume that there are $L(L^*)$ households at Home (Foreign), each supplying one unit of labour, and that every household shares the symmetric Cobb-Douglas preferences defined over $z \in [0,1]$, as $U = \int_0^1 \log[c(z)]dz$ and $U^* = \int_0^1 \log[c^*(z)]dz$. Then, the world income (and the world total expenditure), $wL + w^*L^*$, is also equal to the world expenditure on each good. Since Home produces the goods in $[0, m]$, the total expenditure on the Home goods is $m \cdot (wL + w^*L^*)$, which must be equal to the Home income, wL, in equilibrium. This condition yields

Ricardian Trade Theory, Fig. 2 The two-goods case



$$\frac{w}{w^*} = \frac{m}{1-m}\left[\frac{L^*}{L}\right] \tag{3}$$

which is depicted by the upward sloping curve in Fig. 1. It is upward-sloping because a higher m means that a larger fraction of the world expenditure goes to the goods produced by the Home labour, hence its relative wage goes up. As shown by the intersection of the two curves in Fig. 1, Eqs. (1 and 3) jointly determine the equilibrium relative wage rate, w/w^*, and the equilibrium patterns of trade and specialization, m, as Home exports and Foreign imports goods in [0,m) and Home imports and Foreign exports goods in $(m,1]$. In short, *the patterns of trade follow the patterns of comparative advantage.*

The standard two-country, two-goods Ricardian model, found in many college textbooks, may be recovered as a special case of this model, where $A(z) = A_1$ for $z \in [0, \alpha]$ and $A(z) = A_2$ for $z \in (\alpha, 1]$, with $A_1 > A_2$, as shown in Fig. 2. By aggregating all the goods in $[0, \alpha]$ as a composite good, called Good 1, and all the goods in $(\alpha, 1]$ as another composite good, called Good 2, the model becomes a two-sector model, where the households have the preferences, $U = \alpha\log(C_1) + (1-\alpha)\log(C_2)$ and $U^* = \alpha\log\left(C_1^*\right) + (1-\alpha)\log\left(C_2^*\right)$. Viewed this way, the model highlights the restrictive feature of the two-good assumption in the textbook Ricardian model.

Gains from Trade and Country Size Effects

The Home and Foreign welfares are measured by $\mathrm{U} = \int_0^1 \log[w/p(z)]dz$ and $\mathrm{U}^* = \int_0^1 \log[w^*/p^*(z)]$ dz, respectively. In autarky, they are equal to

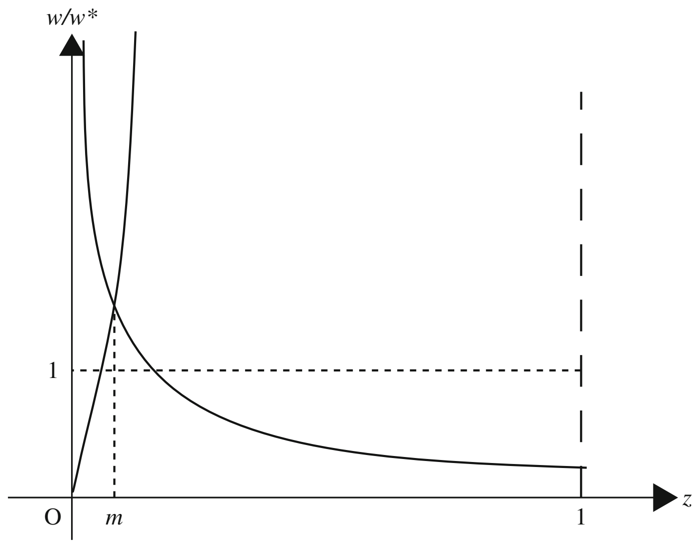
$$U_A = -\int_0^1 \log[a(z)]dz; \quad U_{\mathrm{A}}^* = -\int_0^1 \log[a^*(z)]dz; \tag{4}$$

and, under free trade, they are equal to

$$U_{\mathrm{T}} = -\int_0^m \log[a(z)]dz + \int_m^1 \log\left[\frac{w}{w^*a^*(z)}\right]dz; U_{\mathrm{T}}^* = \int_0^m \log\left[\frac{w^*}{wa(z)}\right]dz - \int_m^1 \log[a^*(z)]dz. \tag{5}$$

Subtracting (4) from (5) and using (1) show that the welfare changes from autarky to free trade by:

$$\Delta U = U_{\mathrm{T}} - U_{\mathrm{A}} = \int_m^1 \log\left[\frac{A(m)}{A(z)}\right]dz; \quad \Delta U^* = U_{\mathrm{T}}^* - U_{\mathrm{A}}^* = \int_0^m \log\left[\frac{A(z)}{A(m)}\right] dz, \tag{6}$$



Ricardian Trade Theory, Fig. 3 Gains from trade and country size effects

both of which are strictly positive in the case depicted in Fig. 1. More generally, both countries gain from trade, as long as $0 < m < 1$ and $A(0) > A(m) > A(1)$. Note that this condition could hold even when one country, say Home, has *absolute advantage* over the other, say Foreign, that is, $A(z) > 1$ for $z \in [0,1]$. Clearly, such absolute advantage allows Home households to enjoy higher wage income and hence a higher standard of living than Foreign households, $w > w^*$, and $U_T - U_T^* = \log(w) - \log(w^*) > 0$. Yet both countries gain from trade as long as trade allows them to specialize in the goods that they are relatively good at producing.

In the two-goods case, shown in Fig. 2, both countries gain from trade only when $m = \alpha$ and $A_1 > w/w^* > A_2$, which requires that $A_2(1 - \alpha)/\alpha < L^*/L < A_1(1 - \alpha)/\alpha$. Home does not gain from trade if $w/w^* = A_2$, which occurs when $L^*/L \geq A_2(1 - \alpha)/\alpha$, and Foreign does not gain from trade if $w/w^* = A_1$, which occurs when $L^*/L \geq A_1(1 - \alpha)/\alpha$. This result of the two-sector model is often interpreted as saying that a large country cannot gain from trade, as it remains *incompletely* specialized, or that a country must *fully* specialize in order to gain from trade, but this is due to an artificial feature of the two-goods model which restricts the cross-country differences in technology.

What is generally true is that, as one country becomes large (small) relative to the rest of the world, its gains from trade become smaller (large). As shown in Fig. 1, an increase in L^*/L, which shifts the upward-sloping curve to the left, leads to a higher w/w^* and a lower m, which implies a higher ΔU and a lower ΔU^*, as seen from (6). For example, a faster population growth in the South (Foreign) allows the North (Home) to specialize further, which improves its factoral terms of trade and its standard of living at the expense of the South. (The same phenomenon might be described by the protectionist in the North as saying, 'because of the cheap labour in the South, the North loses its competitive advantage and industries move from the North to the South'.)

This also suggests that a country with a small population could enjoy higher per capita income even with limited technological superiority. In Fig. 3, Home's technologies are inferior in almost all the goods, yet, thanks to its relatively small population, Home enjoys higher per capita income. This may explain why countries like Norway and Switzerland enjoy a high standard of living even though their geography and climates are not particularly suitable to most economic activities. With smaller populations, they can maintain a high standard of living by specializing in a narrower range of activities that they are particularly good at. This effect is difficult to see within a two-goods model.

These results suggest diminishing returns to scale (DRS) at the country level, even though technologies satisfy the constant returns to scale (CRS) property. This is because the endogeneity of the terms of trade, w/w^*, introduces *de facto* diminishing returns. To see some macroeconomic implications, let us reinterpret the model in the following way. Home and Foreign produce their GDPs, Y and Y^*, by the CRS aggregate production function, $\log(Y) = \int_0^1 \log[c(z)]dz$ and $\log(Y^*) = \int_0^1 \log[c^*(z)]dz$, where $c(z)$ and $c^*(z)$ denote the inputs of the tradeable intermediate goods, $z \in [0, 1]$. The representative household at Home and at Foreign supplies L and L^* units of the composite of the primary factors, which may include not only labour but also physical and human capital. Then, the expressions analogous to (5) become

$$\log\left[\frac{Y}{L}\right] = \int_m^1 \log\left[\frac{w}{w^* a^*(z)}\right]dz - \int_0^m \log[a(z)]dz;$$
$$\log\left[\frac{Y^*}{L^*}\right] = \int_0^m \log\left[\frac{w^*}{wa(z)}\right]dz - \int_m^1 \log[a^*(z)]dz.$$

The effect of a change in L and L^* can be seen by totally differentiating the above expressions, which yields the following growth accounting:

$$\frac{dY}{Y} = \frac{dL}{L} + (1-m)\left[\frac{dw}{w} - \frac{dw^*}{w^*}\right]; \qquad \frac{dY^*}{Y^*} = \frac{dL^*}{L^*} + m\left[\frac{dw^*}{w^*} - \frac{dw}{w}\right].$$

If L and L^* grow at the same rate, w/w^* remains constant, and hence both Y and Y^* also grow at the same rate. However, if L grows faster than L^*, then Y grows slower than L and Y^* grows faster than L^* through the terms of trade effect. This example also suggests that, even when there are increasing returns to scale (IRS) in the aggregate production technologies, naive cross-country growth regression exercises which do not take into account interdependence among countries might fail to uncover economies of scale. See also Acemoglu and Ventura (2002), which studies how such a terms-of-trade mechanism generates stable cross-country distribution of income in the world even when different countries accumulate factors at different rates.

Technology Changes and Transfers

Because it takes cross-country technology differences as the basis of trade, the Ricardian model is well-suited to study the effects of technology changes. Let $g(z) \equiv -\,d\log[a(z)]$ and $g^*(z) \equiv -\,d\log[a^*(z)]$ denote the rate of productivity change in industry z at Home and Foreign. By totally differentiating (5), the Home welfare changes can be expressed as

$$dU = \int_0^m g(z)dz + \int_m^1 g^*(z)dz + \left[\frac{1-m}{1+\xi(m)m(1-m)}\right][g(m) - g^*(m)]$$

where $\xi(z) \equiv -d\log[A(z)]/dz$. For example, it is easy to see that Home always gains from productivity growth in its export sectors, $g(z) > 0$ for $z \in [0, m]$. Does Home gain from productivity growth abroad, as well? If Foreign experiences a *uniform* productivity growth in its export sectors (that is, $g^*(z) = g^* > 0$ for all $z \in [m, 1]$), then the answer is yes because

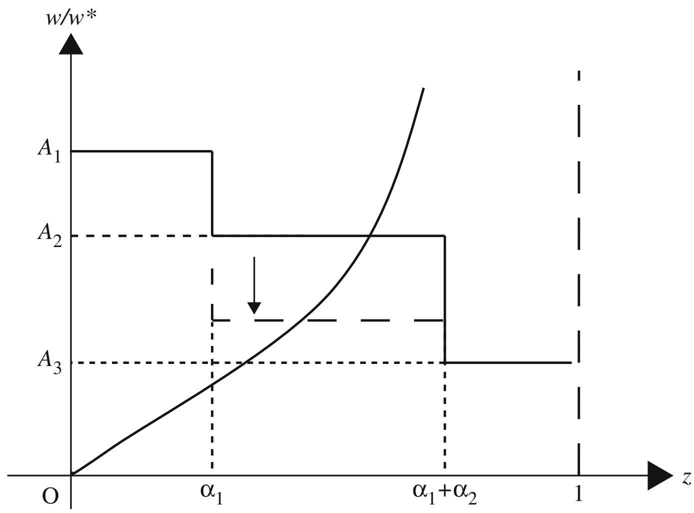
$$dU = \left[\frac{m(1-m)^2\xi(m)}{1+\xi(m)m(1-m)}\right]g^* > 0,$$

unless $\xi(m) = 0$, in which case productivity gains in the Foreign export sectors are entirely offset by the terms of trade change.

On the other hand, Home may lose from Foreign productivity growth if it is concentrated around the marginal sector, m. The following example, taken from Jones (1979), illustrates this possibility. Let $A(z) = A_1 = a_1^*/a_1$ for $z \in [0, \alpha_1]$; $A(z) = A_2 = a_2^*/a_2$ for $z \in (\alpha_1, \alpha_1 + \alpha_2]$, and $A(z) = A_3 = a_3^*/a_3$ for $z \in (\alpha_1 + \alpha_2, 1]$, with $A_1 > A_2 > A_3$. Again, we may view this example as a three-sector model, by aggregating all the goods within each segment, as Goods 1, 2, and 3. When the upward-sloping

R

Ricardian Trade Theory, Fig. 4 The three-goods case



curve intersects with $A(z)$ at its middle segment, $w/w^* = A_2$. Then, the Home and Foreign welfares are given by

$$U_T = (1 - \alpha_1 + \alpha_2)\log\left(a_2^*/a_3^*\right); U_T^* = -(\alpha_1 + \alpha_2)\log\left(a_2^*\right) - (1 - \alpha_1 + \alpha_2)\log\left(a_3^*\right),$$

where we normalize $a_1 = a_2 = a_3 = 1$ without loss of generality to simplify the expression. The Home welfare declines (and the Foreign welfare improves) unambiguously when Foreign productivity growth takes the form of a reduction in a_2^*, depicted by the arrow in Fig. 4. The reason for the Home loss is that the Home purchasing power measured in Good 2 remains unchanged as Foreign productivity growth is completely offset by the increase in the Foreign wage, while the Home purchasing power measured in Good 3 declines, as the increase in the Foreign wage makes it more expensive.

The above example again demonstrates the restrictiveness of the two-goods assumption. In particular, it suggests that the widely used distinction between 'export-biased' and 'import-biased' technology changes in two-sector models, first introduced by Hicks (1953), is of very limited value, as it can be applied only when these changes are uniform across all the export (or import) sectors. Indeed, it can be misleading because the effects of technology changes that take place in *some* export sectors may be very different from those of 'export-biased technological changes' in the two-sector model. (A similar point can be made for the analysis of trade policies. In the standard competitive two-sector model of trade, the government cannot improve its national welfare by providing export subsides to its export sector. This should be interpreted as saying that export subsidies provided *uniformly* to all of its export sectors cannot improve its national welfare. Indeed, using the Ricardian models similar to the one above, Itoh and Kiyono, 1987, showed that *selective* export subsidies that *target* sectors around the marginal sector can improve the national welfare.)

The same mechanism could operate when the technologically lagging country succeeds in catching up with the technologically leading country. Suppose $a(z) = 1$ for all $z \in [0,1]$ and $A(z) = a^*(z) = \Lambda^{1-z}$, where $\Lambda > 1$ is a parameter, representing the extent to which Foreign 'lags behind' Home technologically. For example, each tradable good is produced by performing two tasks by x_1 and x_2, with the Cobb-Douglas production function, $F^z(x_1, x_2) = [x_1/z]^z[x_2/(1-z)]^{1-z}$, and that the unit labour requirement for task 1 is equal to one everywhere, while the unit labour requirement for task 2 is one at Home and $\Lambda > 1$ at Foreign. One may think that Λ

reflects the technology gap, which affects when performing task 2, but not task 1. Then, Home has absolute advantage in all the goods, but Foreign has comparative advantage in the high-indexed goods, which can be produced mostly by performing the simple task 1. (Krugman 1986, offered another story behind a similar parameterization of $A(z)$.) With this parameterization, we have

$$U_{\mathrm{T}} = \frac{(1-m)^2}{2} \log \Lambda, \quad U_{\mathrm{T}}^* = -\frac{(1-m)^2}{2} \log \Lambda,$$

where m is determined by the condition, $mL^*/(1-m)L = \Lambda^{1-m}$. Home could lose if Foreign succeeds in narrowing the gap (that is, a reduction in Λ). The reason is easy to understand. As the gap narrows, Foreign becomes more similar to Home, and Home gains little from trading with a country similar to itself. Indeed, if Foreign catches up completely, $\Lambda = 1$, Home loses all the gains from trading with Foreign because the two countries become identical. Note that the underlying mechanism in this example is the same as in the three-goods example. When Foreign narrows the gap, their productivity growth is not uniform across its export sectors. It is larger in the sectors in which Home has bigger absolute advantages. However, it is false to say that Home suffers because Foreign productivity growth is 'import-biased'. The Home loss is caused by Foreign productivity growth around the marginal sector, not at the lower end of the spectrum.

Nontraded Goods, Trade Costs, and Effects of Globalization

We have been examining the effects of trade by comparing the two extreme cases: autarky, where no goods are traded, and 'free' trade, where all goods are costlessly tradeable. Let us now introduce some trade costs and examine the effects of (partial) trade liberalization by reducing the trade costs. Matsuyama (2007c) conducts such exercises by following DFS (1977), which proposed two alternative ways of introducing trade costs in their model: traded–nontraded dichotomy and uniform iceberg costs.

Traded–Nontraded Dichotomy

Suppose that only the goods in $[0,k]$ are tradable at zero cost and $A(z) \equiv a^*(z)/a(z)$ is continuous, and strictly decreasing in z, within this range. On the other hand, trade costs are so high for the goods in $(k,1]$ that they need to be produced locally. At this point, we do not have to specify the production technologies for these nontradables.

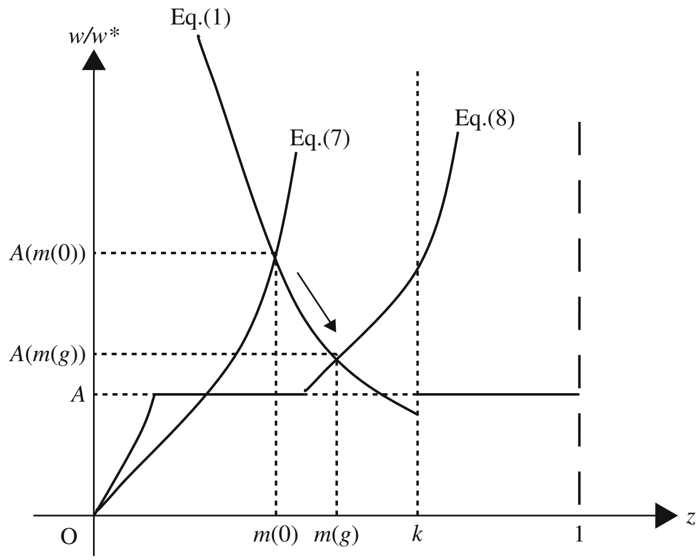
Given the marginal good, $m \in [0, k]$, defined by $A(m) = w/w^*$, Home produces all the goods in $[0, m]$ for both countries and Foreign produces all the goods in $(m, k]$. In addition, each country produces all the goods in $(k,1]$ locally. Therefore, the total expenditure on the goods produced at Home is equal to $m(wL + w^*L^*) + (1-k)\,wL$, which must be equal to the Home income, wL, in equilibrium. This condition yields

$$\frac{w}{w^*} = \frac{m}{k-m}\left[\frac{L^*}{L}\right]. \tag{7}$$

The equilibrium is determined jointly by Eqs. (1 and 7), as shown in Fig. 5. DFS (1977) used this extension to study the classical transfer problem. In the presence of the nontraded goods, the German households spend a larger share of their income on the goods produced in Germany than the households abroad. Because of this 'home bias' in demand, an exogenous income transfer from Germany to the Allies (the war reparations after the Treaty of Versailles of 1919) shifts demand away from the German goods, which leads to a deterioration of the German terms of trade, imposing the additional burden on the German economy.

Let us use this model to study the effects of a globalization. Imagine that some nontradables become tradable. For example, the governments might decide to lift the bans on trading some goods that can be traded costlessly. Or advances in information and communication technologies might open up the possibility of trading some labour services at zero cost. The effects, of course, depend on the relative efficiency of the two

R



Ricardian Trade Theory, Fig. 5 Non-uniform globalization

countries in producing these newly tradables. Consider the case where $A(z) = A$ for all $z \in (k, 1]$, and that a fraction g of these goods becomes newly tradable at zero cost. Then, if $w/w^* > A$, all of the newly tradeables are produced at Foreign. Therefore, given the marginal good, $m \in [0, k]$, Home produces all the goods in $[0,m]$ for both countries and $(1 - g)(1 - k)$ fraction of the goods (those which remain nontradable) locally. Hence, in equilibrium, $wL = m(wL + w^*L^*) + (1 - g)(1 - k)wL$. On the other hand, if $w/w^* < A$, all of the newly tradables are produced at Home. Therefore, Home produces $m + g(1 - k)$ fraction of the goods for both countries and $(1 - g)(1 - k)$ fraction of the goods locally. Hence, in equilibrium, $wL = [m + g(1 - k)](wL + w^*L^*) + (1 - g)(1 - k)wL$. Thus, we have, instead of (7),

$$\frac{w}{w^*} = \frac{m + g(1-k)}{k - m}\left[\frac{L^*}{L}\right] \text{ for } \frac{w}{w^*} < A; \quad \frac{w}{w^*} = \frac{m}{k + g(1-k) - m}\left[\frac{L^*}{L}\right] \text{ for } \frac{w}{w^*} > A; \quad (8)$$

and $w/w^* = A$, otherwise. Note that setting $g = 0$ in Eq. (8) recovers Eq. (7). A higher g shifts the graph to the right above $w/w^* = A$ and to the left below $w/w^* = A$. For each value of g, Eqs. (1 and 8) jointly determines the marginal good and the Home relative wage, which we denote by $m(g)$ and $A(m(g))$.

Suppose that, before globalization, $g = 0$, the equilibrium Home relative wage, $A(m(0))$, is higher than A, as shown in Fig. 5. The arrow indicates the shift caused by an increase in g, which is small enough to keep the Home relative wage higher than A. When some nontraded sectors are opened up, Home abandons the production of these new tradeables, and instead starts producing and exporting the goods in $(m(0), m(g))$, which Home imported previously. The Home relative wage declines as a result, from $A(m(0))$ to $A(m(g))$. The Home and Foreign welfares may be evaluated by

$$U(g) = \int_{m(g)}^{k} \log\left[\frac{A(m(g))}{A(z)}\right] dz + g(1-k)\log\left[\frac{A(m(g))}{A}\right];$$

$$U^*(g) = m(g)\log\left[\frac{A}{A(m(g))}\right] + \int_{m(g)}^{k} \log\frac{A}{A(z)}\, dz - \log A,$$

where we use the normalization, $A(z) = a^*(z)/a(z) = a^*(z)$ for all $z \in [0,1]$, to simplify the expressions. A globalization (an increase in g) affects the Home welfare through two effects that operate in the opposite directions. On one hand, it allows Home to reallocate its labour to the sectors where they have higher relative efficiency, that is, from A to $A(m(g))$ or higher. On the other hand, its relative wage rate, or the factoral terms of trade, $w/w^* = A(m(g))$, deteriorates. The overall effect is generally ambiguous. However, if an increase in g brings down its relative wage rate $A(m(g))$ sufficiently close to A, the positive reallocation effect is dominated by the negative terms of trade effect, so that a further globalization harms the Home welfare. In contrast, a globalization unambiguously improves the Foreign welfare, because both effects operate positively.

The possibility that Home could lose when a globalization takes this form should not be too surprising, because it can be viewed as a form of non-uniform technological changes. Indeed, this mechanism may capture some of the widely held concerns that high-wage countries might lose from 'outsourcing' simple tasks to low-wage countries.

Uniform Iceberg Cost

Suppose now that all the goods, $z \in [0,1]$, are tradeable but subject to the iceberg cost. Each good, when shipped abroad, melts away in transit and only a fraction $g < 1$ arrives at the destination. Thus, in order to supply one unit of each good, the exporter must produce $1/g > 1$ units of the good. Then, Home exports to Foreign only when $a(z)w/g < a^*(z)w^*$, or $w/w^*g < A(z)$, and Foreign exports to Home only when $a(z)w > a^*(z)w^*/g$, or $wg/w^* > A(z)$. Thus, there are two marginal goods, defined by

$$\frac{w}{w^* g} = A(m^-) > A(m^+) = \frac{wg}{w^*}, \tag{9}$$

such that Home produces all the goods in $[0, m^-)$ for both countries; Foreign produces all the goods in $(m^+, 1]$ for both countries, and each produces the goods in $[m^-, m^+]$, which becomes (endogenously) nontraded goods. The demand condition now becomes

$$\frac{w}{w^*} = \frac{m^-}{1 - m^+}\left[\frac{L^*}{L}\right]. \tag{10}$$

Eqs. (9 and 10) jointly determine three endogenous variables, m^-, m^+, and w/w^*.

One could proceed to examine the effects of a reduction in the trade cost, by increasing g. This is left as an exercise for interested readess.

Multiple Countries

The two-country assumption is clearly restrictive for certain purposes, such as analysing the income distribution across countries, studying the patterns of bilateral trade flows, let alone the issues related to the regional integration, such as NAFTA. It is relatively straightforward to extend the two-goods Ricardian model for an arbitrary number of countries; see, for example, Becker (1952) for a finite number of countries and Matsuyama (1996) and Yanagawa (1996) for a continuum of countries. It has been a challenge to allow for an arbitrary number of goods *and* countries in a tractable way. For example, Wilson (1980) extended the DFS model in many dimensions, including a finite number of countries, but it does not permit more than a local perturbation analysis. Acemoglu and Ventura (2002), in their analysis of the cross-country income distribution, assumed the extreme form of technological heterogeneity by adopting the Armington assumption, which prevents the patterns of specialization from changing endogenously.

Eaton and Kortum (2002) developed a parsimonious representation of the Ricardian model with a continuum of goods, which allows for an arbitrary number of countries with the iceberg costs that are uniform across sectors but vary across country pairs. Their key idea is to view the technology heterogeneity across countries as a realization from the Frechet distributions, instead of trying to index the goods in a particular order. This yields simple expressions relating the bilateral trade volumes to technology and geographical barriers, and they use these expressions to estimate the parameters needed to quantify the effects of various policy experiments. For further development, see Alvarez and Lucas (2004).

Multi-Stage Trade and Vertical Specialization

Sanyal (1983) proposed a reinterpretation of the DFS model, according to which the final good is produced through many stages of production, $z \in [0,1]$, in order to analyse trade in intermediate inputs and vertical specialization. If the order in which these inputs need to be produced in the vertical chain of production perfectly coincides with the pattern of comparative advantage, as Sanyal assumed, these inputs are traded only once, as one country specializes in the earlier stages of production and the other specializes in the later stages. Under more general patterns of comparative advantage, however, these inputs may be traded back and forth many times. In such a setting, even a small reduction in trade costs could cause a large and nonlinear increase in the volume of trade, as documented by Hummels et al. (2001).

More General Preferences

With the Cobb–Douglas specification, each good receives a constant share of the expenditure regardless of the prices. Its homotheticity implies that the rich and the poor consume all the goods in the same proportions (when they all face the identical prices), so that the demand compositions are independent of the income distribution within each country. These features, while greatly simplifying the analysis, are too restrictive for addressing many important issues related to growth and development.

Consider, for example, the Fisher–Clark–Kuznets thesis, that is, the changing patterns of sectoral compositions, with the decline of agriculture, the rise and the fall of manufacturing, and the rise of the service sectors. To understand such patterns of structural change in the context of a global economy, Matsuyama (2007b) relaxed the Cobb-Douglas assumption to allow for non-unitary price and non-unitary income elasticities in the three-goods (two tradable and one nontradable) Ricardian model.

Non-homothetic preferences also play the key roles in many models of North–South trade. Flam and Helpman (1987), Stokey (1991), and Matsuyama (2000) all built two-country (North and South) Ricardian models with a continuum of goods, with the open-ended goods space, $z \in [0, \infty)$, and considered non-homothetic preferences with the property that, as the household's income goes up, its demand compositions shift towards higher-indexed goods. When the South, the poorer country, has comparative advantages in lower-indexed goods, the demand has home biases (in spite of the absence of any trade costs). Furthermore, the asymmetry of demand generates many comparative statics results that are absent in the standard Ricardian model. For example, in Matsuyama (2000), immiserizing growth might occur; uniform productivity growth in the South might make the South worse off, as all the benefits go to the North. Or, as the South's population grows, some industries migrate from the North to the South, and new industries are born in the North, generating the patterns of product cycles. Flam and Helpman (1987) and Matsuyama (2000) also looked at the roles of income distributions within each country by endowing different households with different amounts of labour.

Endogenous Technologies and Increasing Returns

So far, we have taken the cross-country differences in technology as exogenous and examined their effects on patterns of specialization and trade. However, the patterns of trade and specialization may also affect technologies. Many Ricardian models with endogenous technologies have been developed to examine such two-way causality between technology and trade. Endogenous technologies have also been used as a natural way of introducing increasing returns in production. Due to the space constraint, we cannot do justice to this vast literature, which contains many alternative approaches to endogenize technologies (static external economies of scale, dynamic increasing returns due to learning-by-doing with or without inter-industry spillovers, agglomeration economies with endogenous product varieties, R&D activities, and so on) with a wide range of results with different policy implications. The interested reader should start with a survey by Grossman and Helpman (1995).

Beyond Technologies: Policy-Induced and Institution-Based Comparative Advantage

The Ricardian set-up has also been used to explain how the differences in national policies and institutions give rise to the patterns of comparative advantage, even in the absence of any inherent technology differences. In Copeland and Taylor (1994), the clean environment is a normal good, so that the rich North chooses a higher pollution tax than the poor South. As a result, the North (South) ends up having comparative advantages in less (more) polluting industries. In Matsuyama (2005), Costinot (2006), and Acemoglu et al. (2007), industries differ according to the severity of agency or contractual problems, and the country with a better (worse) institutional set-up to deal with these problems has comparative advantages in industries that are more (less) subject to these problems. One may view this line of research as an attempt to endogenize technology differences. Unlike the literature surveyed by Grossman and Helpman (1995), however, the main objective here is to look at the deeper or more fundamental causes of technology differences, rather than looking at the two-way causality between technology and trade.

Finally, because the Ricardian trade theory abstracts from the roles of factor endowment differences across countries and factor intensity differences across industries as the basis of trade, it is an ideal set-up in which to isolate the roles of factor endowments and intensity differences that are unrelated to the basis of trade. For example, Matsuyama (2007a) uses a two-country Ricardian model to examine how factor intensity affects the extent of globalization and how globalization affects factor prices when certain factors are used more intensively in international trade than in domestic trade. The model is Ricardian in the sense that the patterns of comparative advantage are determined entirely by the exogenous technological differences. The factor proportions matter, however, because they determine the extent of globalization, as the effective trade costs vary with the relative endowments of the factor used intensively in international trade.

See Also

▶ Comparative Advantage
▶ Globalization
▶ International Trade Theory
▶ Terms of Trade

Bibliography

Acemoglu, D., and J. Ventura. 2002. The world income distribution. *Quarterly Journal of Economics* 117: 659–694.

Acemoglu, D., P. Antras, and E. Helpman. 2007. Contracts and technology adoption. *American Economic Review* 97: 916–943.

Alvarez, F., and R.E. Lucas, Jr. 2004. General equilibrium analysis of the Eaton–Kortum model of international trade. Working Paper, University of Chicago.

Becker, G. 1952. A note on multi-country trade. *American Economic Review* 42: 558–568.

Copeland, B., and S. Taylor. 1994. North–South trade and the environment. *Quarterly Journal of Economics* 109: 755–787.

Costinot, A. 2006. On the origins of comparative advantage. Working paper, UCSD.

Dornbusch, R., S. Fischer, and P.A. Samuelson. 1977. Comparative advantage, trade and payments in a Ricardian model with a continuum of goods. *American Economic Review* 67: 823–839.

Eaton, J., and S. Kortum. 2002. Technology, geography, and trade. *Econometrica* 70: 1741–1779.

Flam, H., and E. Helpman. 1987. Vertical product differentiation and North–South trade. *American Economic Review* 77: 810–822.

Grossman, G., and E. Helpman. 1995. Technology and trade. In *Handbook of international economics*, ed. G. Grossman and K. Rogoff, Vol. 3. Amsterdam: North-Holland.

Hicks, J.R. 1953. An inaugural lecture. *Oxford Economic Papers* 5: 117–135.

Hummels, D., J. Ishii, and K.-M. Yi. 2001. The nature and growth of vertical integration in world trade. *Journal of International Economics* 54: 75–96.

R

Itoh, M., and K. Kiyono. 1987. Welfare enhancing export subsidies. *Journal of Political Economy* 95: 115–137.
Jones, R.W. 1979. Technical progress and real incomes in a Ricardian trade model. In *International trade: Essays in theory*, ed. R.W. Jones. Amsterdam: North-Holland.
Krugman, P. 1986. A 'technology-gap' model of international trade. In *Structural adjustment in developed open economies*, ed. K. Jungenfelt and D. Hague. London: Macmillan.
Matsuyama, K. 1996. Why are there rich and poor countries?: Symmetry-breaking in the world economy. *Journal of the Japanese and International Economies* 10: 419–439.
Matsuyama, K. 2000. A Ricardian model with a continuum of goods under nonhomothetic preferences: Demand complementarities, income distribution and north-south trade. *Journal of Political Economy* 108: 1093–1120.
Matsuyama, K. 2005. Credit market imperfections and patterns of international trade and capital flows. *Journal of the European Economic Association* 3: 714–723.
Matsuyama, K. 2007a. Beyond icebergs: Towards a theory of biased globalization. *Review of Economic Studies* 74: 237–253.
Matsuyama, K. 2007b. Productivity-based theory of manufacturing employment declines: A global perspective. Working paper, Department of Economics, Northwestern University.
Matsuyama, K. 2007c. Uniform versus non-uniform globalization. Working paper, Department of Economics, Northwestern University.
Sanyal, K.K. 1983. Vertical specialization in a Ricardian model with a continuum of stages of production. *Economica* 50: 71–78.
Stokey, N. 1991. The volume and composition of trade between rich and poor countries. *Review of Economic Studies* 58: 63–80.
Wilson, C.A. 1980. On the general structure of Ricardian models with a continuum of goods. *Econometrica* 48: 1675–1702.
Yanagawa, N. 1996. Economic development in a Ricardian world with many countries. *Journal of Development Economics* 49: 271–288.

Ricardo, David (1772–1823)

Terry Peach

Abstract

This article discusses the life and work of David Ricardo. The first section provides a comprehensive overview of his life, his contributions to political economy and his political activities. This is followed by more detailed consideration of his monetary writings (including the 'law of markets' and 'comparative advantage'), his early writings on profits and the 'corn model' interpretation, the labour theory of value, the 'new view' and 'neoclassical' interpretations of his work, and his Sraffa-inspired interpretation as a 'classical' economist.

Keywords

Absolute advantage; Absolute and exchangeable value; Bank of England; Barter; Barton, J.; Bentham, J.; Bullion Committee; Bullion Controversy; Classical distribution theories; Classical economics; Classical theory of money; Comparative advantage; Competition of capitals; Convertibility; Corn Laws; Corn model; Credit; De Quincey, T.; Deflation; Depreciation; Differential rent; Diminishing returns; Direct and indirect labour; Fixed and circulating capital; General equilibrium; Inflation; Ingot plan; Invariable measure of value; King, P., Lord; Labour theory of value; Law of markets; Malthus, T. R.; Malthus's population theory; Market wage; McCullough, J. R.; Marshall, A.; Mill, J.; Mill, J. S.; Money supply; Money wage; National debt; Natural price; Natural wage; Neoclassical economics; Paper money; Political economy; Profit and profit theory; Public debt; Pure labour theory of value; Rate of interest; Rate of profit; Real wage; Rent; Ricardian equivalence theorem; Ricardo, D.; Say, J.-B.; Say's equality; Say's identity; Seigniorage; Shared incidence principle; Smith, A.; Specialization; Specie-flow mechanism; Sraffa, P.; Sraffian economics; Stationary state; Subsistence and subsistence wages; Supply and demand; Surplus; Thornton, H.; Torrens, R.; Utilitarianism; Velocity of circulation

JEL Classifications

B31

David Ricardo was one of the most outstanding political economists of the nineteenth century and one of the most influential of all time. Born on 18 April 1772 at 36 Broad Street Buildings in the City of London, he was the third of 15 surviving children of Abraham Israel Ricardo (1733–1812) and his wife, Abigail Delvalle (1753–1801). Abraham's family were Sephardic Jews who had emigrated from Portugal to Holland at the end of the sixteenth century. His father (David's grandfather) is described as 'a non-official broker in funds and stocks' on the Amsterdam exchange (Heertje 2004, p. 283). Abraham also became a stockbroker, first in Amsterdam and then in London, where he moved in 1760. He married Abigail (from an established London Sephardic family of tobacco and snuff merchants) in 1769 and was granted citizenship in 1771. As related by David's brother, Moses Ricardo, their father

> was a man of good intellect, but uncultivated. His prejudices were exceedingly strong; and they induced him to take the opinions of his forefathers in points of religion, politics, education &c., upon faith, and without investigation. Not only did he adopt this rule for himself, but he insisted on its being followed by his children; his son [David], however, never yielded his assent on any important subject, until after he had thoroughly investigated it. It was perhaps in opposing these strong prejudices that he was first led to that freedom and independence of thought for which he was so remarkable. (Ricardo 1951–1955, *Works*, X, p. 5; hereafter '*Works*')

According to Moses Ricardo, the young David was allotted a 'common-school education', typical of 'those who are destined for a mercantile line of life' (*Works* X, p. 3), the emphasis being on reading, writing and arithmetic. Less typically, at the age of 11 David was sent to Amsterdam for 2 years. Details of the visit are sketchy. It has been suggested that he attended the famous Talmud Tora, attached to the Portuguese Synagogue in Amsterdam, although recent scholarship has favoured a more mundane account in which he was privately tutored in the 'common-school' subjects with the addition of French and Spanish (Heertje 2004). Following his return to London, David's full-time education continued only until he reached the age of 14, when he began working for his father as a clerk and messenger on the London Stock Exchange, although he was allowed 'any masters for private instruction whom he chose to have' during his spare time (Moses Ricardo, *Works* X, p. 3). He was later to complain bitterly of years of neglected education 'at the most essential period of life' (*Works* VII, p. 305), to which he frequently attributed his difficulties in written composition.

In 1792 the Ricardo family moved to Bow, close to the house of Edward Wilkinson, a Quaker and surgeon. Before long, David was romantically involved with Wilkinson's daughter, Priscilla Ann (1768–1829), whom he married on 20 December 1793. The young couple was promptly disowned by both sets of parents. Although the breach with the Wilkinson's was short-lived, it is said that David neither spoke to nor saw his mother again. He was also disinherited and removed from his father's business (he was reconciled with his father after his mother's death). The marriage was the occasion of his breach with the Jewish faith, which, according to Sraffa, was 'the culmination of a gradual estrangement [with Judaism] ... in progress for some time before' (Sraffa, *Works* X, pp. 38–9). He was subsequently to attend the meetings of the non-conformist Unitarians.

The break with his father was not to prove an insurmountable obstacle to David's financial prospects. With the assistance of City friends, he embarked on his spectacularly successful career as a jobber on the stock market and a loan contractor for government stock. Before long he had amassed a considerable fortune, and in 1815, the year in which he made his single most profitable transaction, he began a gradual retirement from business. The total value of his estate at death has been estimated at between £675,000 and £775,000, roughly equivalent to more than £500 million ($950,000,000) at 2006 prices.

As Ricardo's wealth grew, so too did his family and social standing. Eight children were born between 1795 and 1810. From 1812, the family's prestigious London address was 56 Upper Brook Street, Grosvenor Square. To this was added in 1814 Ricardo's country seat of Gatcombe Park near the small village of Minchinhampton, where he is reported to have financed almshouses, endowed a school, provided an infirmary and started a savings bank. His petition for his own coat of arms was also granted in 1814. Having entered the squirearchy (he was High Sheriff of Gloucestershire in 1818), acquired his reputation as a leading intellectual and political economist and become a prominent Member of Parliament (he took his seat in 1818), his company was sought increasingly by luminaries of the aristocracy, the political classes and the intelligentsia. Ricardo was highly gratified by his success, no more so than as a recognized authority on his favourite subject, namely, as he described it, 'political economy'.

It is said that Ricardo's interest in political economy was stimulated by chancing upon a copy of Adam Smith's *Wealth of Nations* in a travelling library while on a visit to Bath. Prior to this time, we are told by his brother that a predilection for subjects of an abstract and general nature had led to a leisurely interest in science, including mathematics, chemistry, geology and mineralogy (in 1807 he was a founding member of the London Institution for the Advancement of Literature and the Diffusion of Useful Knowledge, which was charged with promoting science, as well as literature and the arts, and in 1808 he joined the Geological Society of London). Yet, although his interest was awakened, to the point where he become an avid reader of the early articles on political economy in the *Edinburgh Review*, he was for several years too preoccupied with furthering his financial career to treat political economy as anything more that 'an agreeable subject for half an hour's chat' (as he later reminisced to his old friend Hutches Trower, *Works* VII, p. 246). The turning point came in 1809.

The free convertibility of paper currency into gold had been suspended under the Bank Restriction Act of 1797, following the run on the Bank that had been provoked by fears of a French invasion (in the context of the French wars of 1792–1815). In the aftermath of restriction, the market price of gold (in terms of the now unconvertible paper currency) had risen above the (fixed) mint price and the 'exchanges' had become 'unfavourable', so that premiums were now to be paid on bills of exchange drawn on overseas banks for the purposes of settling international debts. This gave rise to the first phase of the 'Bullion Controversy' (c.1797–1801), with contributions from writers including Henry Thornton and Lord Peter King. The controversy was concerned with the reasons for the depreciation of 'paper' relative to gold and the deterioration in the exchanges, with the 'bullionists' (as represented by King) arguing that the fault lay with the Bank of England for overissuing paper, while the 'anti-bullionists' (including Thornton) emphasized the role of special government payments overseas and poor domestic harvests, independently of the Bank's issues (this was the debate that dominated the early entries on political economy in the *Edinburgh Review*). After 1801, however, the price of gold fell back towards the mint price, the exchanges improved and the controversy duly subsided, to be revived in 1808 by a further marked depreciation of paper relative to gold (of around 20 per cent) and an accompanying fall in the exchanges. It was to this second phase of the controversy that Ricardo's first publication was directed, taking the form of an anonymous letter to the *Morning Chronicle* newspaper, published 29 August 1809.

With uncharacteristic rhetorical excess (*R*icardo intoned gravely about the 'present evil' of the depreciation of paper and the 'disastrous consequences' and 'future ruin' that might follow) the argument in the letter followed the standard bullionist position that the root cause of the 'ills' was the 'over-issues of the Bank [of England]', the remedy being (in a remark that anticipates his later 'Ingot Plan') that 'the Bank be enjoined by Parliament gradually to withdraw ... notes from circulation, without obliging them, in the first instance, to pay in specie' (*Works* III, p. 21). The argument was developed in further letters to the *Morning Chronicle*, and in his (signed) pamphlets, *The High Price of Bullion* (1810–11) and the *Reply to Mr. Bosenquet* (1811). The *Bullion* essay was a straightforward elaboration of Ricardo's position (in which he acknowledged that he 'can add but little to the arguments which have been so ably urged by Lord King', *Works* III, pp. 51–2), with an appendix to the fourth edition in which he developed his plan to resume convertibility by requiring the Bank of England to pay on demand in bullion ingots (not specie) bank notes of the value of at least £20, the alleged advantages being that this would reduce the supply of domestic paper, prevent an excessive demand on the Bank for gold (since the demand for bullion ingots would be less than the demand for specie), and prevent the withdrawal of low face-value bank notes (although Ricardo was later to acknowledge that a secondary market in bullion could facilitate the exchange of small notes pro rata). The second pamphlet was a reply to criticisms of the parliamentary Bullion Committee Report (1810), with which Ricardo broadly agreed, and of Ricardo's own currency writings.

The contributions to the Bullion Controversy brought Ricardo to the attention of political and intellectual figures including Thomas Robert Malthus and James Mill (Ricardo was also in correspondence with Spencer Perceval, the Tory Prime Minister, and the opposition Whig leader, George Tierney). Both Malthus and Mill were to play critical roles in the development of Ricardo's subsequent career, although their influences were profoundly different. At the time of Ricardo's entry on the public stage, Malthus was a seasoned writer, the author of the *Essay on Population* and, arguably, the leading political economist of the day. Although he and Ricardo became, and remained, close friends, their relationship was marked by disagreement over many areas within the new 'science' of political economy. While Ricardo borrowed from some of Malthus's writings (including population theory and the theory of differential rent) other aspects of his work evolved dialectically from epistolary skirmishes with his contemporary.

Malthus introduced himself to Ricardo by letter in June 1811, by which time it seems that Mill was already considered a close friend and an ally on the bullion question. Mill had been an early contributor to the *Edinburgh Review*, but by this time his attention had turned to writing his *A History of British India* and it seems unlikely that he had much influence over the content of Ricardo's political economy with the exception of the 'law of markets' (in short, the doctrine that 'supply creates its own demand'). But that is not to detract from Mill's influence on Ricardo in other respects. Mill advised, encouraged, cajoled and even (only semi-humorously) bullied the ever-reticent Ricardo, who almost certainly would not have completed his major work without Mill's incessant prodding (see J.S. Mill 1873, p. 42). It was also James Mill, as associate and disciple of Jeremy Bentham (with whom Ricardo also became personally acquainted), who was to coach the initially sceptical Ricardo in political utilitarianism and persuade him to enter parliament.

Currency issues dominated the early Ricardo–Malthus correspondence, but by late summer of 1813 their attention began to turn to a new subject, the forces governing movements in the general rate of profit. Up to this point Ricardo had robustly endorsed Adam Smith's 'competition of capitals thesis', according to which the general rate of profit is regulated by the intensity of competition in the labour market (determining movements in wage rates) and in the output market (determining prices). Exactly how Ricardo himself interpreted Smith's doctrine is not clear, nor is it possible to pinpoint the reason for the change of focus in the Ricardo–Malthus correspondence. However, it is certain that Ricardo was newly emphasizing the conditions of producing food as an influence on profits. His position was developed in lost 'papers on the profits of Capital', following which, in response to a clamour by the landed aristocracy for a revision of the old Corn Law, and the report in May 1814 of the Parliamentary Committee on the Corn Trade, his deliberations become more narrowly centred on the effects on profits of restrictions on the free importation of corn. The outcome was his *Essay on The Influence of a Low Price of Corn on the Profits of Stock*; *Shewing the Inexpediency of Restrictions on Importation* (1815; reprinted in *Works* IV).

The central argument of the *Essay* may be given as follows. On the assumption of an economy closed to the importation of foreign corn (the principal subsistence commodity and wage good), the increasing demand for corn from a growing population must be met either by the more intensive cultivation of land or by cultivating land that is less fertile or more disadvantageously situated relative to the final market. Either way, the expansion of output will encounter diminishing returns which in turn lead to a higher corn price, higher money wages and, therefore, a lower agricultural rate of profit. Only the landlords benefit, because they receive more *differential* rent: following Malthus, rent is the difference in return from the 'best' and the 'worst' land on the assumption that the return from the 'worst' land is sufficient only to give farmers the general rate of profit; ergo, landlords benefit as that difference increases. To complete the argument, the reduction in profitability is transmitted to capitalists generally by means of higher money wages. As for labourers, the argument appears to be that they might also suffer in consequence of a ('temporary') fall in labour demand, itself the result of lower profitability. Hence Ricardo's provocative conclusion that 'the interest of the landlord is always opposed to the interest of every other class in the community' (*Works* IV, p. 21).

The *Essay* was a transitional work in which Ricardo repudiated some of the fundamental tenets of the prevailing orthodoxy, as derived from Adam Smith and upheld by Malthus, but failed to supply a fully convincing logical alternative. It was James Mill who persuaded Ricardo to develop his ideas in the form of a major treatise. Two years later Mill's exhortations were rewarded with the publication of *On the Principles of Political Economy and Taxation* (1817; reprinted in *Works* I). However, before Ricardo could begin serious work on the *Principles* he had to fulfil his commitment to Pascoe Grenfell M.P., who had enlisted Ricardo's support for an assault on the Bank of England. Ricardo was more than happy to oblige ('I always enjoy any attack upon the Bank', *Works* VI, pp. 268–9), and the result was his *Proposals for an Economical and Secure Currency; with Observations on the Profits of the Bank of England, as they regard the Public and the Proprietors of Bank Stock* (1816; reprinted in *Works* IV).

In language suggested by James Mill, Ricardo lamented that 'a great and opulent body like the Bank of England' should 'wish to augment their hoards by undue gains wrested from the hands of an overburthened people' (*Works* IV, p. 93). It was intolerable that a mere 'company of merchants' should make vast profits by overcharging on the management of the public debt and other public business, through their 'seignorage' on the issue of paper money and by reducing their unprofitable stock of bullion (as they were enabled to do by the Restriction Act). For the longer term (after the expiry of the Bank Charter in 1833), Ricardo's preferred solution was to strip the Bank of its management of the money supply, which he would entrust to 'commissioners responsible to parliament only, the state' (*Works* IV, p. 114) (this plan was developed in the *Plan for the Establishment of a National Bank* [1824], drafted by Ricardo in 1823, reprinted in *Works* IV). For the shorter term, he suggested that the government should seek more favourable terms for the management of the debt. Above all, however, he again advocated a swift return to a fully convertible paper currency, to be achieved by his Ingot Plan. The result would indeed be an 'economical and secure currency' which, along with its other advantages (of cheapness in comparison with a fully metallic currency and of stability by constraining movements in the market price of gold) would facilitate short-term, compensating changes in the money supply in response to fluctuations in the availability of credit.

With the *Proposals* dispatched to the printers, the way was open for Ricardo to commence work on his *Principles*; or, be more precise, it was *almost* so, for he still had to contend with a hectic social life, recurring bouts of lethargy and defeatism, continuing business interests, the demands of a large family and a 'temptation of being out in the air in fine weather' (*Works* VI, p. 263). Fortunately for posterity, the summer of 1816 offered very few outdoor temptations and Ricardo dedicated himself to his task. The *Principles* was published on 19 April, 1817. It was the result of little more than 6 or 7 months' sustained activity on Ricardo's part.

The 'principal problem in Political Economy' is defined in the *Principles* as the determination of the 'laws' which regulate 'the natural course of rent, profit, and wages' over time. These issues had been addressed in the *Essay* and, indeed, the *Principles* was initially conceived as an *Essay* writ large. In the process of writing the later work, however, its scope was enlarged in previously unforeseen ways as Ricardo developed his ideas. The result was a volume comprising 31 chapters, covering not only the 'laws' governing rent, profit and wages, but also a labour theory of value, a theory of international comparative advantage, monetary theory, several chapters devoted to 'the influence of taxation on different classes of the community', and strictures on the writings of predecessors and contemporaries. The 'core' theoretical analysis as it relates to 'the natural course' may be summarized as follows.

In terms of the newly adopted 'pure' labour theory of value, (changes in) the exchangeable value of competitively produced, freely reproducible commodities are determined exclusively by (changes in) the *quantities* of labour expended on their production, where the relevant quantity of labour is the *greatest* quantity expended per unit of output sold. The theory applies only when commodities exchange at their *natural prices*, defined by uniform wage and profit rates (rent is excluded as a component of price, as explained shortly). In addition, it is assumed that one domestically produced commodity, gold (not to be confused with its real-world namesake), serves as the 'invariable standard' (the *numéraire*) in terms of which all prices are expressed, its 'invariability' defined in terms of a given and unchanging labour input per unit of its output. It follows that any change in a commodity's gold-denominated natural price is an exact reflection of a corresponding change in the labour expended on *its* production. This theory of value was used by Ricardo beyond the first chapter of the *Principles*.

Next, there is the theory of differential rent, derived from Malthus. As Ricardo explained, the relevance of the theory is not confined to agriculture but applies whenever units of the same (homogeneous) class of commodity are produced by different quantities of labour. If all units sell at the same natural price, determined by the *greatest* labour input per unit; and, if the rate of profit from the sale of the unit requiring the greatest labour input is equal to the general (uniform) rate; then, an additional surplus revenue will be earned on units requiring a lower labour input, and it is this additional surplus that constitutes (differential) *rent*. (If we assume with Ricardo that capitalist producers, in agriculture and elsewhere, are profit-maximizers, they will *always* extend production to the point where revenue from the sale of the incremental output, requiring the greatest labour input, is sufficient *only* to yield the general rate of profit; moreover, this greatest labour input *must* determine price, otherwise the general rate of profit could not be received and the output would never be produced.)

On wages, Ricardo introduces the distinction between the market price of labour (or market wage) and the natural price (or natural wage), the latter defined (in money terms) as 'that price which is necessary to enable the labourers, one with another, to subsist and to perpetuate their race, without either increase or diminution' (*Works* I, p. 93). The price that *is* necessary depends on the 'real' natural wage: on 'the quantity of food, necessaries, and conveniences [which] become essential ... from habit' (*Works* I, p. 93). Habits may change over time (*P*erhaps under the influence of education, as Ricardo hoped) but, for analytical purposes, the natural (real) wage is a datum. In the event that the market wage is above or below the natural wage, a Malthusian-style population mechanism is triggered: population expands (or contracts) and the market wage returns to the natural level.

To turn to profits, the eponymous chapter in the *Principles* is chiefly a revision of the central argument from the *Essay*, although it does contain the ingredients for a more general theory of the rate of profit. Now in terms of the labour theory of value, the attempt to expand the output of corn in a closed economy encounters diminishing returns in the form of a greater labour input per unit of output; hence, the (natural) price of corn rises proportionately. This in turn increases money wages, because corn enters the given real wage. The rate of profit (calculated with reference to the output produced by the greatest labour input) must therefore fall (since the rise in the natural price of corn is proportionate only to the increase in the *quantity* of labour expended on its production and does not reflect the increased *cost* of that labour) and, by the reasoning explained above, differential rent increases. Natural prices outside agriculture are either unchanged or, if corn is required as a material input, rise only to reflect the increase in the labour expended on their production; hence, the fall in the agricultural rate is communicated to other sectors by an increase in money wages. Perforce, the 'natural course or rent, profit, and wages' is for rent to increase, the rate of profit to fall and (money) wages to rise, although it must be stressed that this 'prediction' is *entirely* contingent on a host of assumptions and should not be taken as evidence of a gloomy or pessimistic attitude by Ricardo to Britain's economic prospects (such an inference, although commonly made, could not be further from the truth).

Finally, the more general theory from Ricardo's analysis is that (changes in) the rate of profit depend exclusively on (changes in) the labour expended on the production of the given real wage where, to borrow J.S. Mill's distinction, the 'labour expended' covers the 'direct' labour and the 'indirect' labour expended on the production of the non-labour inputs to the production of wage-goods. Provided one grants him his assumptions (of a labour theory of value, a given real wage and known labour conditions of production), Ricardo had thus produced a strong candidate for the first logically coherent theory of the determination of the general rate of profit in the history of economic thought.

There was a great deal riding on the success of the *Principles*, not just Ricardo's growing reputation as a political economist. Mill had suggested in 1815 that Ricardo should enter Parliament, a suggestion from which the latter had recoiled with horror. One year later he was becoming more amenable to Mill's plan, writing to his friend: 'If my book succeeds . . . perhaps my ambition may be awakened, and I may aspire to rank with senators' (*Works* VII, p. 113). Much to Ricardo's relief, the book did succeed to an extent far surpassing his self-deprecatory expectations.

Ricardo entered Parliament on 26 February 1819 as the independent member for the rotten borough of Portarlington in Ireland: a constituency which he never visited, with 12 or so electors in the 'pocket' of Lord Portarlington to whom Ricardo had advanced £25,000 as a loan on the mortgage of his estates.

Ricardo availed himself of every opportunity to educate the House of Commons in the 'true principles of political economy'. These principles dictated the gradual repeal of trade restrictions generally and the of the Corn Law in particular; the gradual repeal of the Poor Laws; the repayment of the National Debt (his heroic proposal to replay the debt over 2 or 3 years by the imposition of a property tax was met with widespread incredulity); minimal taxation and a balanced budget; and a return to a convertible currency. With the signal exception of convertibility (Peel's Bill of 1819 for the Resumption of Cash Payments owed much to his proposals), Ricardo mostly found himself on the losing side, but that did nothing to shake his convictions. His parliamentary contributions are testimony to his belief in political economy as a subject of direct empirical relevance (the view of Ricardo as a pure theorist is a travesty). They also mark him as a zealous advocate of a free-market capitalist system with minimal government interference, who believed that Great Britain 'would be the happiest country in the world, and its progress in prosperity would be beyond the power of imagination to conceive, if we got rid of two great evils – the national debt and the corn laws' (*Works* V, p. 55). Additionally, he spoke out on a range of 'liberal' issues including religious tolerance, slavery, freedom of speech and the right to petition. He also aligned himself with the 'radical' cause for the reform of parliament.

The contention that 'good government' would not be achieved without a reform of parliament had been put to Ricardo by James Mill in 1815 but was at that time rejected on the grounds that Mill exaggerated the 'sinister interest' of politicians in pursuing their own selfish interests and undervalued the corrective influence of enlightened public opinion. Three years later Ricardo's position had changed. Partly as a result of Mill's bombardment of Ricardo with 'radical' messages, partly because of his growing conviction that the Tory government was failing to pursue 'right measures', and after reading Jeremy Bentham's *Plan of Parliamentary Reform*, Ricardo was won over to the 'radical' cause. As he came to argue, 'good government' – government 'administered for the happiness of the *many*, and not for the benefit of the *few'* (*Works* VII, p. 299) – required that politicians should 'legislate for the public benefit only, and not... attend to the interests of any particular class' (*Works* VIII, p. 275); yet, under present arrangements, politicians fell prey to the interests of particular classes, particularly the landed class; hence the necessity for reform. However, Ricardo's proposals fell some way short of those of his 'radical' contemporaries. The introduction of the secret ballot was, for him, an almost sufficient basis for securing good government under existing circumstances, although he did make a case for triennial parliaments and a modest extension of the franchise to include householders. He might therefore be described as a moderate reformer in the utilitarian tradition of Bentham and Mill.

The infamous proposal for the speedy repayment of the national debt was also presented to the public in an invited article on the *Funding System* (1820), written in autumn 1819 for publication in the *Supplement to the Encyclopaedia Britannica*. This article is noteworthy for its exposition of what has come to be known, misleadingly, as the 'Ricardian equivalence theorem'. To take Ricardo's own argument, suppose that a war involves the expenditure of £20 million. This can be financed either by raising £20 m in taxes or by borrowing £20 *m* and repaying by taxes £1 m per annum in perpetuity (at an assumed annual interest rate of 5%), or by borrowing the £20 m and (for example) repaying by taxes £1.2 m per annum, which would clear the interest (at 5%) and the initial £20 m over a period of 45 years. 'In point of economy', as Ricardo stated, 'there is no real difference in either of the modes', because the present value of £1 m per annum in perpetuity or £ 1.2 m over 45 years, both at 5% annually, *is* £20 m, hence the idea of *equivalence*. But, he continued, 'the people who pay the taxes never so estimate them, and therefore do not manage their private affairs accordingly': the different modes are *not* equivalent because, according to him, individuals are prone to undervalue the true cost of repaying a loan over time. That being so, Ricardo's proposal was for the pay-as-you-spend mode of financing which, he believed, would make people 'less disposed wantonly to engage in an expensive contest [namely war], and if engaged in it. . .be sooner disposed to get out of it' (*Works* IV, p. 186).

While Ricardo was writing his *Funding System*, his friend Malthus was putting the finishing touches to his own *Principles*, published in April 1820, which contained an unsparing critique of Ricardo's central doctrines (Malthus's *Principles* together with Ricardo's comments are reprinted in *Works* II). Of all Malthus's criticisms, those levelled at Ricardo's treatment of value were the most acute, thus prompting Ricardo to a major revision of his first chapter for the third edition of his *Principles* (1821) (a lightly revised second edition of the *Principles* had been published in 1819). In addition to the defence against Malthus, the third edition is distinguished by a new chapter 'On Machinery' in which Ricardo, stimulated by the work of John Barton, famously declared that 'the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the correct principles of political economy' (*Works* I, p. 392). To avoid misunderstanding, although there may be a very distant family resemblance between Ricardo's analysis and the standard 'neoclassical' case of factor substitution in response to changes in factor prices within a timeless framework, a principal difference is that Ricardo was describing a process *over time*, in which the substitution of machinery for labour was likely only to apply to new ventures. Nor did his analysis end there, since the capitalists were expected to expand accumulation in consequence of their higher profits, so tending to reverse the fall in the demand for labour (and wages).

Following the third edition of the *Principles*, Ricardo's next and last publication within his own lifetime was *On Protection to Agriculture* (1822; reprinted in *Works* IV): a veritable tour de force, written in little more than 3 weeks.

A new Corn Law had been passed in 1815 which prohibited the importation of foreign corn until the price had been at least 80 shillings per quarter for 6 weeks, by which time the ports could be opened to duty-free importation. Imports had been triggered in 1817–1819, with prices first rising too 111 shillings in June 1817 and then (under the pressure of importation, followed by good domestic harvests) falling steadily to 55 shillings in the second half of 1820. After more bumper harvests, 1822 then witnessed the lowest average corn prices since 1792, with a fall to 34 shillings in November. The 'agricultural distress' was widespread and severe, and the powerful landed interest turned to Parliament for assistance. A parliamentary committee was established to investigate the causes and possible remedies for the distress, with Ricardo as one of its members.

Ricardo was not optimistic that Parliament would, or could, shift its position towards a free trade in corn; as he wrote in correspondence, 'I have no hope of good measures being adopted, the landlords are too powerful in the House of Commons to give us any hope that they will relinquish the tax which they have in fact contrived to impose on the rest of the community' (*Works* IX, p. 158). He was proved right. *On Protection* was his response to the protectionist report of the committee, in which he maintained his position that free trade was the only long-term solution while proposing a revised version of the 1815 Act (with measures for dampening price fluctuations). The pamphlet is also distinguished by sharp restatements of Ricardo's central doctrines, a wealth of detailed empirical analysis and a pungent defence of his own position with regard to Peel's Bill of 1819 for the resumption of cash payments. *On Protection* shows him at the peak of his career, a true master of his subject and a political economist in the most rounded sense.

In the summer of 1822 Ricardo embarked with his family on a four-month grand tour of Continental Europe. Upon his return he resumed his hectic life as an active parliamentarian, attended meetings of the Political Economy Club (which he had co-founded in 1821), drafted his plan for an independent national bank, and continued with his deliberations on 'value'. He was also looking forward to hosting a visit to Gatcombe from his old friend Hutches Trower, to whom he wrote: 'we shall walk and ride, we will converse on politics, on Political Economy, and on Moral Philosophy, and neither of us will be the worse for the exercise of our colloquial powers (*Works* IX, p. 377). But it was not to be. On 11 September 1823 Ricardo died from the effects of an abscess in the middle ear. He was buried at Hardenhuish Park, Wiltshire, on the estate of his daughter, Henrietta, and her husband, Thomas Clutterbuck.

The newspaper obituaries of the time were lavish in their praise of Ricardo's achievements, both as a political economist and as a 'Senator' (see Peach 2003). By his friends, he was applauded for having virtually revolutionized economic theory, not merely for its own sake but as means of guiding government policy and thus promoting the 'general happiness' of society. Of course, his critics were to asses his contributions less kindly, but in producing what was arguably the first coherent supply-side analysis of value, distribution and growth – never mind anything else – his place in doctrinal history was assured.

The following sections consider in more details various aspects of Ricardo's work and a selection of the main interpretative disputes that continue to surround it.

Monetary Contributions, the Law of Markets and Comparative Advantage

As Peake (1978, p. 31) rightly observed, 'Ricardo's total productive output was dominated by monetary questions', and it was in this area that he had the greatest practical influence in his own lifetime.

A simple approximation to Ricardo's 'model' includes the following assumptions: the domestic currency consists entirely of paper money ('paper') issued by a central bank; money prices are a function of the supply of paper (*ceteris paribus*); the bank allows the free convertibility of paper into gold on demand at a permanently fixed mint price, initially equal to the globally determined market price in terms of paper; and all profit-seeking economic agents have virtually perfect market information. Now suppose that the bank increases its supply of paper. Domestic commodity prices rise, as does the market price of gold in terms of paper. (*R*icardo treated gold as just another commodity, a view that later ensnared him in the position that the exchangeable value of gold is determined by the labour expended on its production even though its value is *incessantly* fluctuating in response to changes in the volume and pattern of world trade.) Gold has become relatively cheaper (or 'redundant') because, by assumption, it may be purchased at the lower mint price. Profit-seeking agents therefore exchange paper for gold which, because of its new relative cheapness compared with other domestic commodities, is now exported in preference to those commodities in exchange for foreign produce. Hence, the supply of domestic paper contracts, domestic prices fall, the market price of gold (in paper) returns to the mint price and the *status quo ante* is restored.

Now suppose that paper is no longer freely convertible into gold at the mint price. As before, the supply of paper is increased, but the 'stimulus which a redundant currency gives to the exportation of the coin [namely gold] ... cannot, as formerly, relieve itself' (*Works* III, p. 78), so the market price of gold *remains* above the mint price. In addition, the paper cost of bills of exchange drawn on foreign banks will increase to reflect the fall in paper relative to gold (the foreign exchange becomes 'unfavourable'). Hence (following Lord King) Ricardo's 'two unerring tests' of a depreciation in Bank-notes, 'the rate of exchange and the price of bullion' (*Works* III, p. 75). (Ricardo flatly rejected the measurement of depreciation either in terms of changes in the exchangeable value of gold for domestic commodities – because commodities 'are continually varying in value' among themselves – or in terms of subjectively perceived 'enjoyment', 'because two persons may derive very different degrees of enjoyment from the possession of the same commodity'; *Works* IV, pp. 59, 61.)

As to the consequences of 'depreciation', the picture is mixed. Ricardo stressed that the rate of interest is determined by the rate of profit in the 'real' economy (by the 'competition of capitals' or by the conditions of producing wage goods, in the earlier and later writings respectively); and that the 'trifling' effect on the rate of profit (hence on output) of an increase in paper is confined to an interval 'of momentary duration' before money wages adjust to restore the (assumed) given real wage (*Works* III, pp. 91–2, 318–19, 329; *Works* V, p. 446; *Works* VI, pp. 16–17). This dominant position supports a (mostly) neutral money interpretation, but it also raises the question of why Ricardo became so exercised by depreciation if its real effects were insignificant. The answer is possibly to be found in his concern with the effects of rising prices on recipients of fixed money incomes, especially 'monied men' (see *Works* III, pp. 21, 95–6; *Works* VI, p. 68), regarded by Sayers (1953, p. 65) as a 'shattering' inconsistency with the view 'that long-run effects come quickly and easily'. In addition, the later Ricardo was also to remark on the danger of an 'easy' inconvertible paper-money regime in facilitating speculation and over-trading (*Works* V, pp. 397, 446).

Whatever the economic costs of depreciation, Ricardo campaigned tirelessly for a resumption of convertibility at the pre-restriction mint price of gold. In his evidence before the Parliamentary Committees on Cash Payments (1819) he argued, with heroic simplicity, that, in the prevailing circumstances of a 4% premium in the market over the mint price of gold, a reduction in paper currency of about 4% would be sufficient to restore parity, with a consequent fall of domestic prices generally also of around 4% (*Works* V, pp. 416–17). This objective could be achieved 'in a few months' (*Works* V, p. 396). However, 'by a consideration of the fears which *I* think many people very unreasonably entertained', he was 'reconciled' to a plan for the phased return to the old mint price over one or 2 years (*Works* V, p. 451). The logic, if not the detail, of Ricardo's argument was accepted by the committees, leading to Peel's Bill (1819) with its provision for a staged return to convertibility at the old par over a period from February 1820 to May 1823. Payments were to be made only in bullion ingots in the first two stages, in line with Ricardo's recommendation. Contrary to his proposals, however, the third stage gave the Bank the *option* of making payments in specie, while the fourth and final stage saw a return to *full* convertibility at par.

Ricardo regretted that Parliament had not adopted his plan in its totality, but was otherwise supportive of the bill. Certainly, he did not foresee the events that were to follow which led, on his 1822 estimation, to a 10% depreciation of paper, thus making 'the reverting to a fixed currency as difficult a task to the country as possible' (*Works* IX, pp. 140, 152). The fault lay not with his analysis, however, but with the Bank of England, who had (in his opinion) needlessly purchased large quantities of gold in anticipation of resumption, thus raising its market price *independently* of note issues.

Ricardo may have been justified in blaming the Bank, but he must also stand accused of reasoning *as if* his simple model had captured all relevant aspects of reality. He was aware, on one level, that nominal inflation or deflation was *not* determined exclusively by changes in the Bank's supply of paper, and that the market price of gold could differ from the mint price *independently* of changes in the domestic money supply. He had noted in his early monetary writings that the 'regulator of prices' must include not only the quantity of money, but also 'the rapidity of its circulation' and 'the mass of commodities' (*Works* III, p. 311); later (1815–16), in response to post-war economic conditions, he allowed that the quantity of money was also determined by the independent behaviour (in context, the bankruptcy) of the country banks, and he conceded that 'bullionists', himself included, had underrated the effects on the market price of gold from changes in world demand for the metal (*Works* VI, p. 344; *Works* IV, p. 62); finally, under hostile questioning from some members of the Parliamentary Committees on Resumption, he admitted the further qualification that changes in the general state of confidence could affect domestic prices by influencing the availability of credit, itself a substitute for currency (*Works* V, p. 419). Yet, for the most part (as with his 'four per cent' calculation, noted above), he argued as if these counteracting influences were nugatory to the point that they could be ignored completely. This was a treacherous foundation on which to build economy policy.

It was also Ricardo's habitual presumption that real-world economic actors behaved 'rationally' and it is for this reason that he was highly critical of the Bank for purchasing gold when (on his analysis) it was not in their interest to do so. The same presumption was at the heart of his version of the law of markets, described by Keynes as the (flawed) doctrine that 'supply creates its own demand in the sense that the aggregate demand price is equal to the aggregate supply price for all levels of output and employment' (Keynes 1936, 21–2).

The 'law' is commonly attributed to Jean-Baptiste Say although its roots extend back to Adam Smith's *Wealth of Nations*. It seems likely, however, that it was derived by Ricardo from James Mill, who had sketched out the argument in his 1808 review of William Spence's *Britain Independent of Commerce* in the *Edinburgh Review*. It was first used by Ricardo in his early monetary writings to argue that foreign markets will never be so 'glutted' by British produce as to constrain further British exports when money becomes comparatively dearer (that is, the opposite of 'redundancy'). Later, it was used to bolster the argument that the *only* cause of a permanent reduction in general profitability is an increase in the labour expended on the production of wage goods. It was also invoked by Ricardo in the distressed aftermath of the French wars to support his unshakably optimistic view that recovery was *always* imminent.

Ricardo's version of the 'law' may be reduced to the following propositions: first, commodities will continue to be produced only if they return at least the going general rate of profit; second, capitalist producers (and *only* capitalist producers) are not 'for any length of time...ill-informed of the commodities which [they] can most advantageously produce' (*Works* I, p. 290); third, the desire to consume *something* is 'implanted in every man's breast; nothing is required but the means' (*Works* I, p. 292); fourth, all money income is spent, either by the direct recipients or by those to whom the recipients lend (all) their unspent money income (there is no hoarding). If, to take the extreme case, commodities *always* exchange at natural prices (which implies that the producers earn precisely the going general rate of profit) and all income is *always* spent (on the same output), we have 'Say's identity' version of the law, defined by an excess demand for money of zero at all times. This is the version that Keynes attributed to Ricardo. Yet, although Ricardo's emphasis was always on equilibrium or long-period tendencies, he did (as he was forced to by external events) allow for strictly 'temporary' periods of capital misallocation in which capitalists produce the 'wrong' commodities and demand money (to satisfy creditors) in excess of revenue. The 'Say's equality' version of the 'law', allowing for 'temporary' disequilibria of excess demand for money, would therefore better describe his position. Above all, however, what is most striking is his belief that the 'law' encapsulated real-world tendencies, to the extent that he condemned out of hand all proposals for relief works (because only the capitalists knew best how to allocate capital) and, ultimately, was left totally bemused by the scale and duration of the post-war distress (see *Works* VIII, p. 277).

Economists have been considerably more impressed by his statement of comparative advantage in the chapter 'On Foreign Trade' in the *Principles*. Following Ricardo's own example, assume two commodity bundles of cloth (x_1) and wine (x_2), both of which could in principle be produced in England or Portugal ([x_1, x_2] and $\left[x_1^*, x_2^*\right]$ respectively). To produce the bundles in England would require 100 labourers for cloth (a_1) and 120 labourers for wine (a_2); and to produce them in Portugal would require 90 labourers for cloth $\left(a_1^*\right)$ and 80 labourers for wine $\left(a_2^*\right)$. Portugal therefore has an absolute advantage in the production of *both* bundles. Labour (alias 'capital') is immobile internationally, trade initially takes place by way of real barter and, implicitly, bundles are produced under constant returns to scale in both countries.

As Ricardo states, it will be advantageous for England to specialize in making cloth and Portugal to specialize in wine, because both countries thereby obtain more of the other commodity per unit of their domestic labour than if they attempted to make it themselves. For example, if Portugal used 80 labourers to make cloth she could obtain only $0.89x_1$, but if she can exchange $1x_2^*$ (also the produce of 80 labourers) for $1x_1$, as Ricardo supposes, then it would be 'advantageous for her to export wine in exchange for cloth' (*Works* I, p. 135). Similarly, if England used 100 labourers to produce wine she could obtain only $0.833x_2$, so she also benefits by exchanging $1x_1$ (the produce of 100 labourers) for $1x_2^*$.

Ricardo's example implies that the pattern of specialization is dictated by the 'four magic numbers' (Samuelson 1972, p. 378), namely, $a_1/a_2 < a_1^*/a_2^*$. However, the principal purpose of the analysis was not so much to illustrate comparative advantage per se, but to show that the 'same [labour theory] rule which regulates the relative value of commodities in one country, does not regulate the relative value of the commodities exchanged between two or more countries'(*Works* I, p. 133). If, as Ricardo supposes, there is a rate of exchange of $1x_1$ for $1x_2$, 'England would give the produce of the labour of 100 men, for the produce of the labour of 80': something that 'could not take place between the individuals of the same country' (*Works* I, p. 135).

It was also Ricardo's purpose to show that the introduction of money (gold) would leave the analysis unaffected: gold will be 'distributed in such proportions amongst the different countries ... as to accommodate [itself] to the natural traffic which would take place if no such metal existed, and the trade between countries were purely a trade of barter' (*Works* I, p. 137). To give the flavour of the argument, suppose England and Portugal each produce both commodities and that the initial gold prices are $px_1 = px_1^*$ which, given the 'magic numbers', implies $p_2 > p_2^*$. Wine is therefore exported from Portugal to England and is paid for by gold. But, on Ricardo's quantity theory reasoning, the influx of gold to Portugal, and its efflux from England, will raise prices in the former country and reduce them in the latter. Hence, the specie-flow mechanism ensures that $px_1 = px_1^*$ is unsustainable (the same would apply to $px_2 = px_2^*$, by similar reasoning) and that the price of Portuguese cloth must exceed the price of English cloth (just as the price of wine must be higher in England than in Portugal), thus leading to complete specialization. Contrary to Ricardo, however, ' the natural traffic which would take place if no such metal existed', hence relative world prices, are not unique (as he implied), with the range of possible outcomes defined by the condition of $a_1/a_2 \leq p_1/p_2 \leq a_1^*/a_2^*$.

Debate continues as to whether Ricardo was the true originator of the comparative advantage doctrine, or whether that accolade should be awarded to his contemporary, Robert Torrens (see Ruffin 2002, 2005, for a recent case in Ricardo's favour). But, regardless of who may have crossed the line first, it is with Ricardo's name that comparative advantage has become indelibly linked.

Early Writings on Profit (1813–1815) and the 'Corn Model' Interpretation

In the introduction to his masterful edition of Ricardo's *Collected Works* it was suggested by Piero Sraffa that the early Ricardo had devised a model in which corn is the sole agricultural input and output, thus supplying a 'rational foundation' for the 'principle of the determining role of the profits of agriculture', putatively articulated by Ricardo in 1814 with the words 'it is the profits of the farmer which regulate the profits of all other trades' (*Works* VI, p. 104). By implication, when Ricardo wrote that agricultural profits 'regulate' other profits, he had intended a statement of *unique determination* in full awareness of the logically required assumptions. Sraffa revealed, however, that the corn model (or 'corn ratio theory of profits') was 'never stated by Ricardo in any of his extant letters and papers' although, on the basis of indirect textual evidence, he claimed that Ricardo 'must have formulated it' either in lost papers or conversation (*Works* I, p. xxxi). Later, with the publication of Sraffa (1960), it transpired that the corn model had additional significance as a simple precursor of Sraffa's own 'Standard system' in which corn is the sole 'basic commodity'; and Sraffa also disclosed that the interpretation was the outcome of his *own* theoretical work: 'it was only when the Standard system and the distinction between basics and non-basics had emerged in the course of the present investigation that the ['corn model'] interpretation of Ricardo's theory suggested itself as a natural consequence' (Sraffa 1960, p. 93).

The corn model interpretation was widely embraced. With a beguiling pedagogical simplicity, it could 'explain' Ricardo's regulatory statements and his later development of the pure labour theory, with Malthus entering the story to remind Ricardo that agricultural capital does *not* consist entirely of corn, thus necessitating a new (labour) theory of value. However, beginning in the early 1970s with the work Samuel Hollander, doubts have increasingly been aired about the textual basis for the interpretation. What follows is the view of one such critic. (For a sample of critical interpretations, see Faccarello 1982; Hollander 1973, 1975, 1979; Peach 1984, 1993, 2001. The case for the defence has been made by Eatwell 1975, and Garegnani 1982, among others.)

If Ricardo's writings are sifted for confirmation for the interpretation – in other words, if the corn model is *presumed* – then it is easy enough to find 'evidence' in its favour. Without that presumption, the picture is rather different. Thus, Ricardo's assertion in correspondence that the 'rate of profits and of interest must depend on the proportion of production to the consumption necessary to such production' (*Works* VI, p. 108) is said by Sraffa to be the 'nearest that Ricardo comes to an explicit statement on these [corn model] lines' (*Works* I, p. xxxii). Yet, although it is *possible* to conceive of such a 'proportion' in material terms, the expression itself provides no evidence of the way it was conceived by Ricardo; moreover, very similar expressions had been used by him in his earlier monetary writings in which there is no question of him having adopted corn model assumptions. As for the regulatory statements, while it is *possible* to impose a corn model rationalization, the problem is in establishing that the same rationalization was applied by Ricardo. Here, too, the evidence is disobliging. The *Essay* (1815) is replete with such statements (for example, 'The general profits of stock depend wholly on the profits of the last portion of capital employed on the land'; *Works* IV, p. 21), but we can be sure they were not thought *by Ricardo* to depend on the corn model because he *explicitly* assumed heterogeneous inputs to agriculture (including ‘buildings, implements, &c.’; *Works* IV, p. 10). Indeed, an arresting feature of the *Essay* is that its specious corn model appearance derives from the use of corn (alias wheat) to *value* the physically heterogeneous agricultural capital. It was Ricardo’s ‘failure’ to *revalue* this capital as corn became more difficult to produce (in the initial agricultural phase of the argument) that drew Malthus’s criticism and led, ultimately, to Ricardo’s adoption of the labour theory, not an assumption that agricultural capital *comprises* of corn alone.

A question for those who reject the corn model is how the pre-*Essay* Ricardo could arrive at his ‘regulatory’ position if, as he believed at the time, a rise in the price of corn would be followed by a rise in prices generally. Samuel Hollander has conjectured that Ricardo may have invoked a monetary constraint, so that an agriculturally induced rise in money wages would not be passed on in higher prices. Alternatively, it may have been that his view of pricing was *integral* to the analysis: with price rises common to output *and* the heterogeneous inputs to agriculture, Ricardo might have reasoned that an increase in the capital–output ratio *must* reduce profitability. Admittedly, these alternative interpretations do not have the simple elegance and logical consistency of the corn model but, then again, the period 1813–1815 was one in which Ricardo was struggling to establish new ideas within an inherited theoretical framework, much of which was later to be discarded. The existence of contradictions and unresolved theoretical issues during this period is unsurprising.

The Labour Theory of Value

The unmodified or ‘pure’ labour theory of value (*P*LTV) was adopted by Ricardo in early 1816, on the basis of which he drafted material that would form the first seven chapters of the *Principles* (up to and including the chapter ‘On Foreign Trade’). But then he discovered a source of modification to the PLTV resulting from differences in capital structure between production processes. At first, the discovery impeded his progress, but then it seems he had the inspiration to turn it to his advantage (so he thought) in the form of the ‘ curious effect': the iconoclastic demonstration that prices *fall* following a general rise in wages and consequent reduction in the rate of profit. What he did not do, however, was provide any justification for using the PLTV in the light of the ‘curious effect’ analysis.

A simple ‘dated labour’ framework may serve to illustrate Ricardo’s position. Assume three commodities (x_1, x_2, x_3), each produced by ten units of homogeneous labour (L) applied over two discrete production periods ($t-1$, t), with the following conditions of production: $10L_t \rightarrow x_1$; $5L_t + 5L_{t-1} \rightarrow x_2$; $10L_{t-1} \rightarrow x_3$. If we denote the uniform wage and profit rates as w and R, each taking period t values, the natural price equations for the commodities are:

$$px_1 = 10L_t.w(1+r) \tag{1}$$

$$px_2 = 5L_t.w(1+r) + 5L_{t-1}.w(1+r)^2 \tag{2}$$

$$px_3 = 10L_{t-1}(1+r)^2. \tag{3}$$

A PLTV requires $px_1 = px_2 = px_3$, because each commodity is produced with the same quantity of labour. However, it is evident (with $r > 0$) that $px_3 > px_2 > px_1$.

Moreover, if distribution (between w and r) changes, there will be price fluctuations *even though* labour inputs are unchanged. Thus, assume that x_1 is the *numéraire* commodity (so that $px_1 \equiv 1$); in principle, on the basis of (1) a new (lower) r can be calculated for a given rise in w, and these numbers may be entered into (2) and (3) to obtain the new natural prices of x_2 and x_3. With the ‘compounding’ (or magnification) of the effect of the lower r on px_2 and, even more so, on px_3, the result will be a fall of both prices (expressed in terms of x_1) with px_3 falling to the greater extent.

To relate the above to the chapter ‘On Value’ in the first two editions of the *Principles*, the differences in production conditions (or capital structure) are, in Ricardo’s terms, a reflection of

R

differences in (*a*) the durability of fixed capital; (*b*) the ratios of fixed to circulating capital; (*c*) the durability of circulating capital (added in the second edition), where the fixed–circulating capital distinction depends, essentially, on the time required to repay a capital expenditure (the longer the time, the more 'fixed' the expenditure). In the case given by Ricardo, the two extremes (corresponding to x_1 and x_3) are a commodity produced by unassisted labour in 1 year and a commodity produced by unassisted machinery that lasts 100 years. Then, taking the former commodity as his 'invariable standard' (alias *numéraire*), he calculates that a fall of 7% in the rate of profit would reduce the price of the latter by 68%: a vivid illustration of the 'curious effect' (*Works* I, p. 60).

The effect implies, by Ricardo's own testimony, that the PLTV is subject to a (truly) 'considerable modification' from differences in capital structure, but the really curious feature of the first two editions of the *Principles* is that the PLTV (used beyond the first chapter) had been undermined by its own author. This does, indeed, deserve the obloquy of a shattering inconsistency, and one not lost on Malthus, who, in his *Principles* (1820, *Works* II), employed the ingenious tactic of using Ricardo's own analysis to demonstrate the untenability of the PLTV. His criticisms hit home.

Ricardo comprehensively rewrote the chapter 'On Value' for the third edition of the *Principles*, newly adopting two strategies for the defence of the PLTV. First, he ruthlessly extirpated all the numerical examples that *had* suggested a 'considerable modification' to the PLTV and replaced them by others, according to which the 'greatest effects which could be produced on the relative prices of... goods from a rise of wages, could not exceed 6 or 7 per cent' (*Works* I, p. 36). Second, he introduced a new section 'On an invariable measure of value', where he indicated his desire to find a 'perfect measure of value' in terms of which prices would change *only* to reflect changes in the quantities of labour expended on production. This was tantamount to claiming that the discovery of the 'perfect' standard would itself sanction a PLTV, his problem being, however, that any commodity standard must be produced with *some* capital structure and, as he had demonstrated, the 'unwanted' price fluctuations are inescapable if capital structures differ. Hence his second-best solution of assuming that the standard is produced using an (unweighted) 'average' capital structure (cf. x_2, above), allegedly characteristic of 'most commodities': at least for *them* a PLTV would apply. There would still be a 'curious effect' with a fall in profitability, just as some commodities (such as our x_1) would now *rise* in price, but this was announced *sotto voce* (*Works* I, p. 46) and the effect was nowhere near as 'curious', at least in its magnitude, as it had been before.

Through this process of 'double indemnification' Ricardo had, *for the first time*, justified his use of the PLTV in explicit acknowledgement of the problems caused by differences in capital structures: either the differences are small and can be ignored, or (really a variation on the same theme) all the relevant commodities, including the standard, are part of a 'general mass' with the same capital structure. The 'exceptions' to the PLTV may therefore be ignored.

If we leave aside the dubious merit of the defence, its very inclusion is evidence that Ricardo was not retreating in his advocacy of the PLTV, contrary to claims by earlier commentators including J. Hollander (1904) and Cannan (1929, p. 177). That view was laid to rest by Sraffa, who opined that 'the theory of edition 3 appears to be the same, in essence and in emphasis, as that of edition 1' (*Works* I, p. xxxviii): a view that may itself be criticized for undervaluing the scale and significance of the changes (cf. Hollander 1979, p. 217). But why was the theory so important to him?

One attraction is that it provided him with (in its own terms) a logically coherent framework for establishing his central theoretical propositions, particularly of the dependency of the general rate of profit on the conditions of producing wage goods.

A second possibility is that the theory appealed because of its (supposed) empirical relevance; hence Stigler's attribution to Ricardo of 'an *empirical* labour theory of value, that is, a theory that the relative quantities of labour ... are the dominant determinants of relative value' (Stigler 1958, p. 60; emphasis in original). However, although Ricardo did make empirical claims on behalf of theory (for which, it must be said, no evidence was adduced), those claims were arguably more a reflection of his commitment than its basis.

There was also an increasing tendency on Ricardo's part to identify the very *essence* of value with expended labour time. This 'value', referred to him at different times as 'natural value', 'positive value' and 'real value', was conceived as an attribute of *individual* commodities; hence the criticism, *fully accepted by Ricardo*, that he had moved beyond a purely relative usage of 'value' (as in Stigler's interpretation) and had turned it into something absolute (*Works* IX, p. 38). From this perspective, the role of the 'perfect' standard was to harmonize the 'labour values' with (relative) cost-of-production 'values' (or natural prices), since (changes in) the latter would become an exclusive reflection of (changes in) the former. As Ricardo forlornly conceded, however, 'perfection' is ruled out by unequal capital structures: the point he develops in *Absolute Value and Exchangeable Value* (*Works* IV, pp. 361–412), poignantly truncated by his final illness.

The analytical and 'philosophical' attractions are therefore central to understanding Ricardo's PLTV commitment. Of course, with the benefit of nearly two centuries of hindsight, it could be (and has been) argued that the labour theory can be jettisoned, to be replaced (say) with Sraffa's physically specified input–output equations. That argument may be formally correct, but we would no longer have *Ricardo's* theoretical and conceptual system. For him, the labour theory of value was both fundamental and indispensable.

The 'New View'

The inappropriately styled 'new view' (anticipated by Cannan 1893, pp. 247–53, 350, with modern restatements by Casarosa 1978; Hicks and Hollander 1977; Hollander 1990, 2001, 2002, among others) can be treated either as a stand-alone interpretation of Ricardo's treatment of wages or as part of a more far-reaching attempt to assimilate Ricardo's work to 'neoclassical' economics.

In the second great 'rehabilitation' of Ricardo (the first was J.S. Mill's attempt to have him reinstated as 'the greatest political economist': Mill 1848, p. 397), Alfred Marshall applied his principle of 'generous interpretation' to distance Ricardo from the labour theory of value (by that time with its Marxian connotations) and absorb him within the mainstream intellectual tradition. Thus he averred that Ricardo had been 'feeling his way' towards a subjective utility analysis and that, despite appearances, he had attributed coordinate importance to supply *and demand* in the determination of natural prices (Marshall 1920, Appendix I). Interestingly, however, Marshall's generosity deserted him when it came to Ricardo's treatment of wages, which he regarded as indefensible (Pigou 1925, p. 413).

Most subsequent commentators have considered Marshall's' interpretation as far too generous, the prominent exception being Samuel Hollander, who goes even further in claiming to find a 'fundamentally important core of general-equilibrium economics' in Ricardo's work, implying a 'strong continuity of doctrine' between Ricardo's and later 'neoclassical' analysis (Hollander 1987, pp. 6–7; cf. Morishima 1989). As part of this 'general equilibrium' analysis, Ricardo had (allegedly) treated the wage rate as an *endogenous variable*, and it is this feature that is emphasized by the 'new view' interpretation.

The 'non-wage' aspects of the 'neoclassical' Ricardo may be dealt with briefly. First, with regard to utility (in the sense of subjective satisfaction), there is no question that it was treated as a *precondition* for exchangeable value and, in circumstances of fixed supplies, it was also conjectured by Ricardo that prices would be *proportionate* to 'utilities' (*Works* II, pp. 24–5; *Works* VIII, 276–7). However, there was no attempt by him to develop an analysis of *diminishing* marginal utility and, for the purpose of explaining exchangeable values (at natural prices), his emphasis was on *objective* determination by quantities of labour time. As to the 'coordinate' influence of supply and demand, this confuses the *process* by which market prices tend to their natural levels (which does involve output variations and, therefore, a pre-'neoclassical' species of supply-and-demand reasoning) with the *determination* of the natural price levels: for Ricardo, the latter is *independent* of supply and demand, thus effectively denying any *theoretical* relationship between output and (labour) conditions of production (see, for example, *Works* VIII, p. 207). Finally, on 'general equilibrium', there is not a single developed instance of such an analysis in the entire corpus of Ricardo's writings. It is an interpretation obtained only by reconstructing his work and, in the process, obliterating his own hallmark emphasis on unidirectional relationships.

With the 'new view', however, there is at least a textual basis. The most compelling evidence is from three paragraphs in the chapter 'On Wages'. The first paragraph opens thus: 'In the natural advance of society, the wages of labour will have a tendency to fall, as far as they are regulated by supply and demand', the assumption being that 'the supply of labourers will continue to increase as the same rate, while the demand for them will increase at a slower rate' (*Works* I, p. 101). Wages are therefore falling continuously in 'the natural advance' and only reach their 'natural' level (defined for a stationary population) in the terminal stationary state. However, 'we must not forget, that wages are also regulated by the prices of the commodities on which they are expended' (*Works* I, p. 101) and, particularly, by the rising price of corn (from diminishing returns on the land). Money wages therefore *rise* in the 'natural advance' but not by so much as to fully compensate the labourers for the rising corn price, so that real wages secularly decline as before. The effect of diminishing agricultural returns is in this way 'shared' between capitalists and labourers and has come to be known as the 'shared incidence principle'.

There is no doubt that the new view passages exist, and there are also muted refrains of the analysis elsewhere in the *Principles* (*Works* I, pp. 215, 220). At the same time, the natural wage analysis – with the natural wage, defined for a stationary population, as the active centre of gravity for market wages in *all* stages of society – is by far the dominant analysis in the *Principles*; and, unlike the new view, it is the only one consistent with the repeated claim that real-wage variations are of only 'temporary' significance, particularly with regard to movements in the general rate of profit. Based solely on the *Principles*, the proposition that the new view represents Ricardo's *true* position is difficult to sustain.

Malthus, for one, did not recognize Ricardo as a (kindred) new view theorist; hence the trenchant criticisms of Ricardo's natural wage analysis in his own *Principles* (Malthus 1820 *Works* II, pp. 256–64). As Samuel Hollander (2007) has emphasized, however, Ricardo protested that he maintained 'no other doctrine than that which has been well explained by Mr. Malthus' (*Works* II, p. 288). Yet he also reaffirmed his *own* definition of the natural wage (*Works* II, pp. 227–8), which is inexplicable if he truly agreed with Malthus (for whom Ricardo's natural wage would be irrelevant outside the stationary state).

While it cannot be denied, then, that Ricardo was on some level sympathetic to the new view analysis, he was at no time an unequivocal exponent of that doctrine. Even in his later writings (such as *On Protection to Agriculture*, *Works* IV), in which the natural wage is not mentioned explicitly, the real wage is treated as a given and fixed entity, without a trace of the new view. It is also significant that Ricardo was never to criticize the writings of contemporaries, including their avowed representations of his own position, for the (universal) failure to include the new view (Peach 2007). His own credentials as a new view theorist must therefore remain in considerable doubt.

Conclusion: Ricardo as a 'Classical' Economist?

Ricardo was to achieve great fame as a political economist during the tragically short period in which he wrote on the subject, although his ideas, especially his policy proposals, were often bitterly contested by critics of differing political and theoretical persuasions (see Peach 2003). Following his death, his name, if not always his own doctrines, lived on through the writings of his 'New School' disciples, notably James Mill, Thomas De Quincey and the indefatigable J.R. McCulloch, and through the efforts of J.S. Mill. With the advent of 'neoclassical' economics, however, Ricardo's stock began to plummet, with Marshall's attempted rehabilitation to no avail. By the time Sraffa's edition of the *Collected Works* appeared in the 1950s, Ricardo's positive contribution was not uncommonly reduced to anaemic generalities such as the development of a 'professional frame of mind' or an 'abstract deductive approach', the onward progress of economic science having established the 'inadequacy' of much of his substantive work.

The *Collected Works* prompted a flurry of new scholarly interest in Ricardo, but it was only after the publication of Sraffa (1960) that he was subjected to his third major 'rehabilitation', this time not as a 'mainstream' economist (as with J.S. Mill and Marshall) but as a precursor of *Sraffa's* economics or, as related by Sraffa's followers, as a founder of the 'classical' (or 'surplus') tradition that Sraffa (1960) had revived.

The defining characteristics of 'classical' economics are alluded to in the Preface of Sraffa (1960) and amount to the assumption of *given* outputs and methods of production. The distribution between wages and profits may then be 'solved' by taking one distributive variable as given and calculating the other as a residual (or 'surplus'). But how well does this apply to Ricardo's approach? At one level – the calculation of profit as a 'surplus' – it does so well enough. Where the problems arise is with the other attribution of given outputs.

According to Sraffa, 'The "principal problem in Political Economy" was in [Ricardo's] view the division of the national product between classes and in the course of that investigation he was troubled by the fact that the size of this product appears to change when the division changes' (*Works* I, p. xlviii). Hence, 'the problem of value which interested Ricardo was how to find a measure of value which would be invariant to changes in the division of the product' (*Works* I, p. xlviii); and, as Sraffa remarks parenthetically, Ricardo may have come close to solving his 'problem' with the 'average' standard adopted in the third edition of the *Principles*: 'If measured in such a standard, the average price of all commodities, and their aggregate value, would remain unaffected by a rise or fall of wages' (*Works* I, pp. xliv–xlv).

Several objections can be made against Sraffa's interpretation. First and foremost, it implies that Ricardo's 'principal problem' was with purely 'notional' redistributions of a *given national product* (that is, with given and unchanging outputs). However, Ricardo's own 'principal problem', as he defined it himself, was with the 'natural course of rent, profit, and wages' *over time*, and for the purpose of *his* investigation there will be at least one output, that of corn, that *cannot* be treated as given and unchanging in terms of its conditions of production. Second, as Ricardo clarified, *his* analysis of distribution was to be framed at the level of the individual firm, or farm, not in terms of social or 'national' aggregates. Third, there is no evidence that *he* envisaged a 'price-balancing' function for his 'average' standard (that is, to ensure constancy in the total value of national output); indeed, Ricardo's opinion was that *all* distributions-induced price changes are evidence of a 'defect' in the standard.

Ricardo was not a full-fledged 'classical' economist in the Sraffa mould. To describe him more loosely as a 'surplus theorist' is unexceptionable, although by focusing on only one (albeit important) area of Ricardo's writings it is also a 'thin' characterization. Ricardo was a towering intellectual force whose work ranged over all the main areas of political economy. Forcing him into classificatory boxes of a later construction is a disservice to him and a hindrance to those who would seek to understand the full richness and extent of his historical significance.

See Also

- Sraffian Economics

R

Bibliography

Cannan, E. 1893. *A history of the theories of production and distribution in English political economy from 1776 to 1848*. London: Percival.

Cannan, E. 1929. *A review of economic theory*. London: P.S. King & Son.

Casarosa, C. 1978. A new formulation of the Ricardian system. *Oxford Economic Papers* 1: 38–63.

Eatwell, J.L. 1975. The interpretation of Ricardo's essay on profits. *Economica* 42: 182–187.

Faccarello, G. 1982. Sraffa versus Ricardo: The historical irrelevance of the 'corn profit' model. *Economy and Society* 11: 122–137.

Garegnani, P. 1982. On Hollander's interpretation of Ricardo's early theory of profits. *Cambridge Journal of Economics* 6: 65–77.

Heertje, A. 2004. The Dutch and Portuguese background of David Ricardo. *European Journal of the History of Economic Thought* 11: 281–294.

Hicks, J., and S. Hollander. 1977. Mr. Ricardo and the moderns. *Quarterly Journal of Economics* 91: 351–369.

Hollander, J.H. 1904. The development of Ricardo's theory of value. *Quarterly Journal of Economics* 19: 455–491.

Hollander, S. 1973. Ricardo's analysis of the profit rate, 1813–15. *Economica* 40: 260–282.

Hollander, S. 1975. Ricardo and the corn profit model: Reply to Eatwell. *Economica* 52: 188–202.

Hollander, S. 1979. *The economics of David Ricardo*. London: Heinemann.

Hollander, S. 1987. *Classical economics*. London: Blackwell.

Hollander, S. 1990. Ricardian growth theory: A resolution of some problems in textual interpretation. *Oxford Economic Papers* 42: 730–750.

Hollander, S. 2001. Classical economics: A reification wrapped in an anachronism? In *Reflections on the classical canon in economics*, ed. E.L. Forget and S. Peart. London: Routledge.

Stigler, G.J. 1958. Ricardo and the 93% labour theory of value. *American Economic Review* 48: 357–367. In *David Ricardo: Critical Assessments*, vol. 2, ed. J. Cunningham Wood. Beckenham: Croom Helm, 1985.

Hollander, S. 2002. The canonical classical growth model: Content, adherence and priority. In *Competing economic theories*, ed. S. Nisticò and D. Tosato. London: Routledge.

Hollander, S. 2007. Ricardo as a 'classical economist': The 'new view re-examined: A reply to Dr Peach. *History of Political Economy* 39: 307–312.

Keynes, J.M. *1936*. *The general theory of employment, interest and money*. London: Macmillan.

Marshall, A. 1920. *Principles of economics*. 8th ed, 1979. London: Macmillan.

Mill, J. 1808. Review of Britain independent of commerce, by W. Spence. *Edinburgh Review* 11: 429–449.

Mill, J.S. 1848. *Principles of political economy*, 1987. New York: Kelley.

Mill, J.S. 1873. *Autobiography*, 1989. London: Penguin.

Morishima, M. *1989*. *Ricardo's economics: A general equilibrium theory of distribution and growth*. *Cambridge*: *Cambridge University Press*.

Peach, T. 1984. David Ricardo's early treatment of profitability: A new interpretation. *Economic Journal* 94: 733–751.

Peach, T. 1993. *Interpreting Ricardo*. Cambridge: Cambridge University Press.

Peach, T. 2001. Hollander de Vivo and the 'further evidence' for the corn model interpretation of Ricardo: A conspiracy of silence? *Cambridge Journal of Economics* 25: 685–692.

Peach, T. 2003. *David Ricardo: critical responses, 4 vols*. London: Routledge.

Peach, T. 2007. Ricardo as a 'classical economist': The 'new view' re-examined. *History of Political Economy* 39: 293–306.

Peake, C.F. 1978. Henry Thornton and the development of Ricardo's economic thought. *History of Political Economy* 10: 193–212. In *David Ricardo: Critical Assessments* (second series), vol. 5, ed. J. Cunningham Wood. London: Routledge, 1994.

Pigou, A.C. 1925. *Memorials of Alfred Marshall*. London: Macmillan.

Ricardo, D. *1951–1955*. The works and correspondence of David Ricardo, *ed. P. Sraffa with the collaboration of M.H. Dobb, vols. I–X, Cambridge: Cambridge University Press.*

Ruffin, R.J. 2002. David Ricardo's discovery of comparative advantage. *History of Political Economy* 34: 725–748.

Ruffin, R.J. 2005. Debunking a myth: Torrens on comparative advantage. *History of Political Economy* 37: 711–722.

Samuelson, P.A. 1972. The way of an economist. In *The collected papers of Paul A. Samuelson*, ed. R.C. Merton. Cambridge, MA: MIT Press.

Sayers, R.S. 1953. Ricardo's views on monetary questions. *Quarterly Journal of Economics* 67: 30–49. In *David Ricardo: Critical Assessments*, vol. 4, ed. J. Cunningham Wood. Beckenham: Croom Helm, 1985.

Sraffa, P. *1960*. *Production of commodities by means of commodities: prelude to a critique of economic theory*. *Cambridge*: *Cambridge University Press*.

Ricardo–Hayek Effect

Stefano Zamagni

Dwelling on the familiar Ricardian proposition (ch. I, section V, of the *Principles*) according to which a rise in wages will encourage capitalists to substitute machinery for labour and vice versa, F. von Hayek (1939) coined the expression 'Ricardo effect' for the assertion that a general change in wages relative to the prices of final goods will alter the relative profitability of the different methods of production employing labour and capital in different proportions. To be sure, Schumpeter (1939, pp. 345, 812) refers to the influence of factor prices on the introduction of a new method of production as the 'Hayek effect'. On the other hand, Hayek's Ricardo Effect is not to be confused with the celebrated Ricardo machinery effect: the latter concerns the employment *effects* of the introduction of a new method of production; the former deals with the *causes* of its introduction. To avoid misunderstandings, it is therefore proper to use the expression 'Ricardo–Hayek effect'.

In his 1939 paper, Hayek made use of the effect in order to show that a rise in the demand for consumer goods, with money wages and interest rates remaining unchanged, by causing an increase in prices of consumer goods and a decrement of real wages, will lead to a fall in the demand for capital goods thereby causing unemployment. So, contrary to Keynes, a rising level of consumption must, after a certain point, reduce rather than increase the rate of investment. This peculiar utilization of the effect within the realm of business cycle theory triggered more than one kind of criticism, the fiercest of which was Kaldor's (1939, 1942) who arrived to suggest that the real author of Professor Hayek's proposition is Wicksell and not Ricardo at all.

Responding to his critics, Hayek (1942) stressed that his Ricardo effect is a proposition of general character, whose validity and importance are quite independent from its special application to the problems of industrial fluctuations. Indeed, under the above mentioned assumptions, a fall of real wages, with a uniform rate of interest, will certainly change in the same proportion the costs of producing final goods by different methods, but this does not prevent the attractiveness of investing in different methods of production from being affected differently. Why this is so is due to the fact that the current distribution of the funds at the command of the firms between expenditures in wages (investment in circulating capital) and expenditures on machinery (investment in fixed capital) is determined by the circumstance that as long as the prices of final goods remain high relative to costs, the difference is a source of profit every time the capital is turned over, so that higher profits *per unit of time* will be made as the firm can turn over its capital more frequently.

The appropriate context in which to assess the validity and relevance of the Ricardo–Hayek effect is not – as Kaldor took it – the familiar comparative static exercise whereby an old and a new method of production are compared for their profitability, taking for granted that the equipment appropriate to the new method is already in existence or can be procured instantaneously. Rather the effect is to be couched in a dynamic context, since it deals with the transient phase before the new equipment becomes available during which the firms have to decide the relative rates at which they will spend their current outlay on renewing or adding to the two kinds (fixed and circulating) of capital assets. The essence of the Ricardo–Hayek effect is that 'profits will be higher on the method with the higher rate of turnover, *not* because they would accrue at a higher rate *after* the new equilibrium envisaged by Kaldor had been established . . . but because the profits on the less capitalistic method will *begin to accrue* earlier than those on the more capitalistic method' (Hayek 1942, p. 148). In other words, what Kaldor and others disregarded in their assessment of the effect is that – as shown in Zamagni (1984) – the new position which will be eventually achieved, if at all, is time dependent in the precise sense that it depends on the behaviour exhibited by the firms during the transition, a behaviour which, in turn, is affected by the profits accruing to them as the adjustment process goes on.

References

Kaldor, N. 1939. Capital intensity and the trade cycle. *Economica* 6: 40–66.

Kaldor, N. 1942. Professor Hayek and the concertina-effect. *Economica* 9: 359–382.

Schumpeter, J. 1939. *Business cycles*. New York: McGraw-Hill.
Zamagni, S. 1984. Ricardo and Hayek effects in a fixwage model of traverse. *Oxford Economic Papers* 36-(Supplement): 135–151.
von Hayek, F. 1939. *Profits, interest and investment*. London: Routledge & Kegan Paul.
von Hayek, F. 1942. The Ricardo effect. *Economica* 9: 127–152.

Ricci, Umberto (1879–1946)

Giancarlo Gandolfo

Ricci was born in Chieti, Italy, and died in Cairo, Egypt. After an administrative career, his contributions to economic theory (he was an autodidact without academic training) won him a chair in economics (1912), after which he taught in various universities. A critic of the Fascist regime, an article (1928a) written in his humorous, ironic style was the occasion for the government to deprive him of his Rome chair. He then taught in the Universities of Cairo (1929–40) and Istanbul (1942–6). He was a Fellow of the Econometric Society.

Ricci was a major theoretician in various fields: capital theory, demand and supply theory, public finance. He also wrote on economic policy, statistics, and the history of economic analysis. Although a follower of the Walras–Pareto general equilibrium approach, he upheld the usefulness of Marshall's partial equilibrium approach (1906, 1924). He stressed the importance of the elasticity concept and its connection with outlay (1931, 1932), and offered an original treatment of indivisible commodities (1935). Ricci is one of the independent originators of the cobweb theorem (1930) and among the first to realize the potentialities of the econometric approach to give empirical content to 'hypothetical experiments' (i.e. simulations) in economics (1928b, 1939). He coined the word 'polipolio' (polypoly) to denote a set of monopolies and theorized the extreme case of generalized corporativism as a polypolistic system (1926). Also remarkable is his work in applied statistics as Director of the Statistical Service, International Institute of Agriculture, 1910–22 (1914). His posthumous book (1951), containing the lecture notes of his Istanbul courses, is a lucid treatment of microeconomics.

Selected Works

1906. Curve piane di offerta dei prodotti. *Giornale degli economisti,* September.
1910. *Il Capitale-Saggio di economia teoretica.* Turin: Bocca.
1914. *Les bases théoriques de la statistique agricole internationale.* Rome: Imprimerie de l'Institut International d'Agriculture.
1924. Pareto e l'economia pura. *Giornale degli economisti e rivista di statistica,* January-February. Trans. as 'Pareto and Pure Economics', *Review of Economic Studies* 1, October 1933, 3–21.
1926. *Dal protezionismo al sindacalismo.* Bari: Laterza.
1928a. La scienza e la vita. *Nuovi studi di diritto, economia e politica* 1(3), anno VI.
1928b. Il metodo in economia politica. In *Scritti della Facoltá giuridica di Roma in onore di A. Salandra.* Milan: Vallardi. Enlarged German version, Die Methode in der Nationalökonomie, *Zeitschrift für Nationalökonomie* 4(5), October 1933.
1930. Die 'Synthetische Ökonomie' von Henry Ludwell Moore. *Zeitschrift für Nationalökonomie* 1(5): 656.
1931. Courbes de la demande et courbes de la dépense. *L'Egypte contemporaine* 22: 556–558.
1932. The psychological foundation of the law of demand. *Journal of Political Economy* 40: 145–185.
1935. The modification of the utility curve for money in the cases of indivisible goods and goods of increasing utility. *Economica* NS 2: 168–197.

1939. Una nuova via aperta all'econometrica: la misura dei fatti ipotetici. *Rivista internazionale di scienze sociali* 10: 70–87.

1951. *Eléments d'économie politique pure – Théorie de la valeur.* Milan: Malfasi.

References

A full bibliography of Ricci's works is published in his posthumous book (1951), which also contains obituaries by L. Einaudi and C. Bresciani-Turroni (the first of these is also published in English in *American Economic Review,* September 1946). For an evaluation of Ricci's scientific contributions, see also L. Gangemi, In memoria di Umberto Ricci, *Studi Economici ed Aziendali,* April–June 1946 (with bibliography), and C. Grilli, Umberto Ricci e l'economia psichico matematica, *Studi Economici,* November–December 1951.

Rising Supply Price

Peter Newman

'Rising supply price' is a name that partial equilibrium theorists give to their encounters with general equilibrium reasoning. Such encounters must have occurred ever since economics began but for us the story begins in 1912 with Pigou, who asserted that:

> in industries of increasing returns the supply price is greater than the marginal supply price; in industries of diminishing returns the supply price is less than the marginal supply price . . ..It follows that, other things being equal, in industries of increasing returns the marginal net product of investment tends to exceed, and in industries of diminishing returns to fall short of, the marginal net product yielded in industries in general. (1912, pp. 176–7)

These conclusions led him to argue that taxes should be placed on decreasing returns industries and bounties on increasing returns industries. Assuming, what Clapham (1922) seriously doubted, that actual industries can be sorted into such boxes, this is a policy recommendation that appears remarkably specific in content and general in application.

For Pigou, a decreasing returns industry is one in which the expenses of producing $x + \Delta x$ units exceed those of producing x units by more than the expenses attributable directly to the Δx units; this is what excess of 'marginal supply price' over supply price means. However, since the proposed tax-subsidy policy only makes sense for the long-run, replication of a plant of optimal size is always possible and at once rules out decreasing returns to scale in the physical sense. This makes it difficult to see how Pigovian decreasing returns industries can exist, unless the expansion in output from x to $x + \Delta x$ causes a rise in price of one or more of the resources used by the industry. Pigou was willing to admit this possibility.

But then, as Allyn Young gently pointed out in his review of Pigou's book, there is

> A more serious difficulty when we inquire as to the precise content of the 'resources' which are devoted to the work of production . . ..Changes in the prices of product and of resources are the very essence of the situation. Increased prices for the use of land and the other factors in production do not represent an increased *using up* of resources in the work of production. They merely represent *transferences* of purchasing power. (1913, p. 683, his italics)

Thus the rising supply price that accompanies expansion of the industry (as distinct from expansion of any one of its firms) is simply a consequence of increases in the rents of those resources that it uses relatively heavily. Barring net physical external diseconomies, it does not correspond to any increase in the use of real resources.

It took a long time for Young's fundamental point to sink in. In what was essentially the second, much enlarged and retitled, edition of his book, Pigou acknowledged that Young's criticism was 'very important' but defended himself with the feeble argument that 'each [industry] . . . is supposed to make use of only a small part of the aggregate resources of the country' (1920, pp. 934–6). Thus Frank Knight, who had been Young's graduate student at Cornell when the latter's review of Pigou appeared, felt called upon to point out once more the nature of the errors that Pigou was making. So effective was the famous article in which he did this (1924) that it has been reprinted many times, which Young's prior contribution never was, not even by Young himself in his collection (1927).

But that was welfare economics. In positive economics, Clapham's article of 1922 set off a controversy over increasing returns and competition which exploded like a string of firecrackers in the pages of the *Economic Journal* over the next ten years, until the books on imperfect competition by Chamberlin and Joan Robinson in 1933 brought it sputtering to a close. The controversy inevitably touched upon problems of 'rising supply price', but nowhere did it do so effectively save in an article by Roy Harrod, written in 1928 but not accepted and published until 1930 ('An egoistic footnote' in 1951, p. 159, fn2, attributed the delay to an unfavourable referee's report by Frank Ramsey.) His argument went like this:

> Let us call the proportion in which the factors of production A,B,C . . . are mixed in use at the margin in national industry as a whole a:b:c: . . . if an industry using the factors in the proportion of a + x:b:c: . . . expands, it can only get increasing quantities of A at an enhanced price in terms of B, C . . . No doubt by the law of substitution x will be reduced in consequence of the expansion of this industry; but not to zero . . ..
>
> . . . it follows that every industry which uses an appreciable fraction of the factors of production, unless it be an industry using them at the margin in the proportions of a:b:c: . . ., obeys the law of increasing supply price. . . .
>
> This analysis seems to clear up the problem of the old classical distinction between agriculture and the manufacturing industries. If A is land, and a + x: b:c: . . . the proportion in which the factors are mixed at the margin in agriculture as a whole, x/a is clearly large. Agriculture as a whole is thus markedly subject to increasing supply price. (1930, pp. 240–41)

This is an explicit account of what Young merely sketched, though there is no evidence that Harrod had read Young's review. In the following year Viner (1931) published his much reprinted codification of neoclassical partial equilibrium theory, in which without reference to either Harrod or Young he introduced the idea of 'pecuniary' economies and diseconomies, both internal and external. According to this classification what have been discussed here so far are 'net pecuniary external diseconomies', which at that time Viner did not emphasize. However, almost twenty years later Viner added a Supplementary Note to the 1950 reprint of (1931), in which he felt

> . . . it incumbent upon me, . . . to avoid propagating serious error, to carry the analysis . . . further . . . by departing here from the traditional Marshallian pattern of assumptions to which the article adheres. The partial-equilibrium nature of the Marshallian assumptions leaves a wider range of possibilities to the long-run tendencies of costs for an expanding industry than is consistent with general-equilibrium analysis. I first saw this in 1938, and thereafter pointed it out to my students at the University of Chicago. But the first and, to my knowledge, still the only, analysis in print similar to what I have in mind is in Joan Robinson's excellent article, 'Rising Supply Price,' . . . [1941] . . . which has not attracted the attention which in my opinion it eminently deserves. (Viner 1951, p. 227)

In a further footnote, added to the 1951 reprint of the 1950 version, Viner also acknowledged Harrod's prior contribution in (1930). Joan Robinson's fine article is indeed the culmination of this whole line of reasoning, developing in much greater detail and in crystal-clear prose the mode of analysis that began with Harrod; but it is puzzling that Harrod (unlike Hicks, Marshall, Pigou, Robbins and Sraffa) is never mentioned, in spite of the striking similarities between the two analyses. An interesting sidelight is that, in a letter written soon after the appearance of her article and published in Robinson (1951, pp. 42–3), Keynes took a markedly general equilibrium approach to the problem.

Apart from relevant surveys of external economies by Ellis and Fellner (1943) and Chipman (1965, Section 2.8, pp. 736–49) there has been little further discussion of 'rising supply price', evidence perhaps that its nature is by now well understood. However, even as late as 1954, Scitovsky's well-received article with its Pigovian policy conclusions and remark that 'Pecuniary external economies clearly have no place in equilibrium theory' (1954, 149, 146), showed that confusion still existed. Maybe each generation of partial equilibrium theorists has to learn the lesson anew.

See Also

▶ External Economies
▶ General Equilibrium
▶ Pecuniary and Non-Pecuniary Economies
▶ Pigou, Arthur Cecil (1877–1959)

Bibliography

Arrow, K.J., and T. Scitovsky, eds. 1969. *Readings in welfare economics.* Homewood: Richard D. Irwin. Reprints Knight (1924) and Scitovsky (1954).

Boulding, K.E., and G.J. Stigler, eds. 1951. *Readings in price theory.* Homewood: Richard D. Irwin. Reprints Clapham (1922), Ellis and Fellner (1943), Knight (1924), Robinson (1941) and Viner (1931).

Chipman, J.S. 1965. A survey of the theory of international trade. Part 2. The neo-classical theory. *Econometrica* 33: 685–760.

Clapham, J.H. 1922. Of empty economic boxes. *Economic Journal* 32: 305–314.

Clemence, R.V. ed. 1950. *Readings in economic analysis. Vol. 2: Prices and production.* Cambridge, MA: Addison-Wesley. Reprints Viner (1931).

Harrod, R.F. 1930. Notes on supply. *Economic Journal* 40: 232–241.

Harrod, R.F. 1951. *The life of John Maynard Keynes*. London: Macmillan.

Harrod, R.F. 1952. *Economic essays.* London: Macmillan. Reprints Harrod (1930).

Knight, F.H. 1924. Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics* 38: 582–606.

Knight, F.H. 1935. *The ethics of competition.* London: George Allen & Unwin. Reprints Knight (1924).

Pigou, A.C. 1912. *Wealth and welfare*. London: Macmillan.

Pigou, A.C. 1920. *The economics of welfare*. London: Macmillan.

Robinson, J.V. 1941. Rising supply price. *Economica* NS 8: 1–8.

Robinson, J.V. 1951. *Collected economic papers*. Oxford: Basil Blackwell. Reprints Robinson (1941).

Scitovsky, T. 1954. Two concepts of external economies. *Journal of Political Economy* 62: 143–151.

Viner, J. 1931. Cost curves and supply curves. *Zeitschrift für Nationalökonomie* 3: 23–46.

Viner, J. 1951. Reprint of (1931) in Boulding and Stigler (1951).

Young, A.A. 1913. Pigou's wealth and welfare. *Quarterly Journal of Economics* 27: 672–686.

Young, A.A. 1927. *Economic problems new and old*. Boston: Houghton Mifflin.

Risk

Mark J. Machina and Michael Rothschild

Abstract

The phenomenon of *risk* plays a pervasive role in economics. Without it, financial and capital markets would consist of the exchange of a single instrument each period, the communications industry would cease to exist, and the profession of investment banking would reduce to simple accounting. One need but consult the contents of any recent economics journal to see how the recognition of risk has influenced current research in the discipline. This article presents an overview of the modern economic theory of the characterization of risk and the modelling of economic agents' responses to it.

Keywords

Capital asset pricing model; Expected utility hypothesis; Mean-standard deviation; Portfolio theory; Probabilistic sophistication hypothesis; Probability distribution; Riemann–Stieltjes integral; Risk; Risk aversion; Risk preference; Stochastic dominance; Subjective probability; Uncertainty; von Neumann and Morgenstern

JEL Classifications

D8

R

The phenomenon of *risk* is one of the key determining factors in the formation of investment decisions, the operation of financial markets, and several other aspects of economic activity.

Risk Versus Uncertainty

The most fundamental distinction in this branch of economic theory, due to Knight (1921), is that of 'risk' versus 'uncertainty'. A situation is said to involve *risk* if the randomness facing an economic agent presents itself in the form of exogenously specified or scientifically calculable *objective probabilities*, as with gambles based on a roulette wheel or a pair of dice. A situation is said to involve *uncertainty* if the randomness presents itself in the form of alternative possible *events*, as with bets on a horse race, or decisions involving whether or not to buy earthquake insurance.

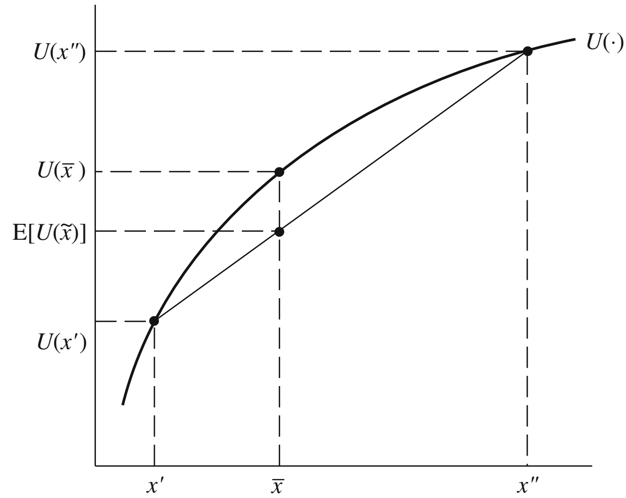
The standard approach to the modelling of preferences under uncertainty (as opposed to risk) has been the *state-preference approach* (for example, Arrow 1964; Debreu 1959, ch. 7; Hirshleifer 1965, 1966; Karni 1985; Yaari 1969). Given the absence of exogenously specified objective probabilities, this approach represents the randomness facing the individual by a set of mutually exclusive and exhaustive *states of nature* or *states of the world* $\mathscr{S} = \{s_1, \ldots, s_n\}$. Depending upon the particular application, this partition of all conceivable futures may either be very coarse, as with the pair of states (it snows here tomorrow, it doesn't snow here tomorrow) or else very fine, so that the description of a single state might read 'it snows more than three inches here tomorrow *and* the temperature in Paris at noon is 73° *and* the price of gold in New York is over $900.00/ounce'. The objects of choice in this framework consist of *state-payoff bundles* of the form $(c_1, ..., c_n)$, which specify the payoff that the individual will receive in each of the respective states. As with regular commodity bundles, individuals are assumed to have preferences over state-payoff bundles which can be represented by indifference curves in the *state-payoff space* $(c_1, ..., c_n)\}$.

Even though the state-preference approach has led to important advances in the analysis of choice under uncertainty (see, for example, the above citations), the advantages of being able to draw on modern probability theory has led economists to hypothesize that an individual's *beliefs* in such settings can nevertheless still be represented by so-called *personal probabilities* or *subjective probabilities*, which take the form of an additive *subjective probability measure* $\mu(\cdot)$ over the state space $\mathscr{S}$. In such a case, a given state-payoff bundle $(c_1, ..., c_n)$ will be viewed as yielding outcome c_i with probability $\mu(s_i)$, so that the individual would evaluate the bundle $(c_1, ..., c_n)$ in the same manner as he or she would evaluate a casino gamble which yielded the payoffs $(c_1, ..., c_n)$ with respective objective probabilities $(\mu(s_1), \ldots, \mu(s_n))$. The hypothesis that individuals have such probabilistic beliefs and evaluate state-payoff bundles in such a manner is termed the *hypothesis of probabilistic sophistication*, and permits a unified application of probability theory to the analysis of decisions under both objective risk and subjective uncertainty. The joint hypothesis of probabilistic sophistication and expected utility risk preferences has been axiomatized by Ramsey (1926), Savage (1954), Anscombe and Aumann (1963), Pratt et al. (1964), and Raiffa (1968, ch.5), and probabilistic sophistication without expected utility has been axiomatized by Machina and Schmeidler (1992).

Choice Under Risk: The Expected Utility Model

For reasons of expositional ease, we consider a world with a single commodity (for example, wealth). An agent making a decision under either risk or probabilistic uncertainty can therefore be thought of as facing a choice set of alternative univariate probability distributions. In order to consider both discrete (for example, finite outcome) distributions as well as distributions with density functions, we represent each such probability distribution by means of its cumulative distribution function $F(\cdot)$, where $F(x) \equiv \text{prob}\,(\tilde{x} \succcurlyeq x)$ for the random variable $\tilde{x}$.

In such a case we can model the agent's preferences over alternative probability distributions in a manner completely analogous to the approach of standard (that is, non-stochastic) consumer theory: he or she is assumed to possess a ranking $\succcurlyeq$ over distributions which is complete, transitive and continuous (in an appropriate sense), and hence representable by a real-valued *preference function* $V(\cdot)$ over cumulative distribution functions, in the sense that $F^*(\cdot) \succcurlyeq F(\cdot)$ (that is, the distribution $F^*(\cdot)$ is weakly preferred to $F(\cdot)$ if and only if $V(F^*) \succcurlyeq V(F)$.



Risk, Fig. 1 Von Neumann–Morgenstern utility function of a risk-averse individual

Of course, as in the non-stochastic case, the above set of assumptions implies nothing about the functional form of the preference functional $V(\cdot)$. For reasons of both normative appeal and analytic convenience, economists typically assume that $V(\cdot)$ is a *linear functional* of the distribution $F(\cdot)$, and hence takes the form

$$V(F) \equiv \int U(x)\, dF(x) \qquad (1)$$

for some function $U(\cdot)$ over wealth levels x, where $U(\cdot)$ is referred to as the individual's *von Neumann–Morgenstern utility function.* (For readers unfamiliar with the *Riemann–Stieltjes integral* $\int U(x) dF(x)$, it represents nothing more than the expected value of $U(\tilde{x})$ when $\tilde{x}$ possesses the cumulative distribution function $F(\cdot)$. Thus if $\tilde{x}$ took the values $x_1,\ldots, x_n$ with probabilities $p_1,\ldots, p_n$ then $\int U(x)dF(x)$ would equal $\sum U(x_i)p_i$, and if $\tilde{x}$ possessed the density function $f(\cdot) = F'(\cdot)$ then $\int U(x)dF(x)$ would equal $\int U(x)f(x)dx$.

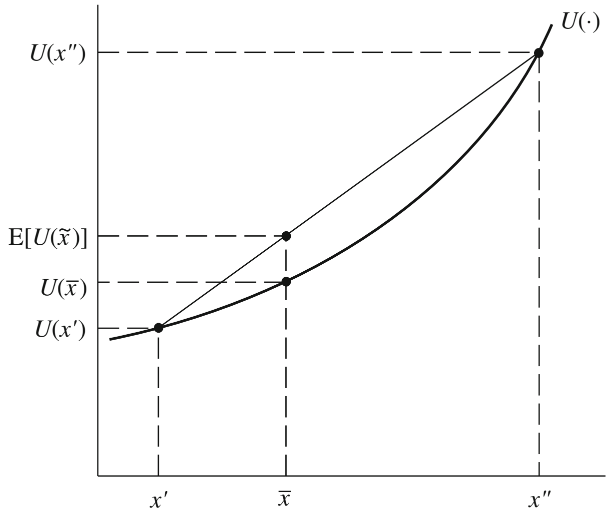
Since the right side of (1) may be thought of as the mathematical expectation of $U(\tilde{x})$, this specification is known as the expected utility model of preferences over random prospects (for a more complete statement of this model, see expected utility hypothesis). Within this framework, an individual's attitudes towards risk are reflected in the shape of his or her utility function $U(\tilde{x})$. Thus, for example, an individual would always prefer shifting probability mass from lower to higher outcome levels if and only if $U(x)$ were an increasing function of x, a condition which we shall henceforth always assume. Such a shift of probability mass is known as a *first order stochastically dominating shift*.

Risk Aversion

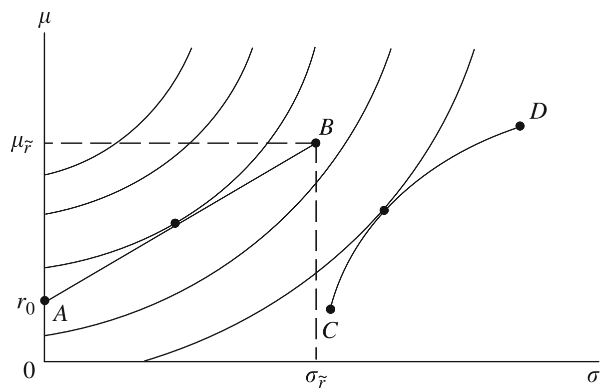
The representation of an individual's preferences over distributions by the shape of his or her von Neumann-Morgenstern utility function provides the first step in the modern economic characterization of risk. After all, whatever the notion of 'riskier' means, it is clear that bearing a random wealth $\tilde{x}$ is riskier than receiving a certain payment of $\bar{x} = E[\tilde{x}]$ (the expected value of the random variable $\tilde{x}$). We therefore have from Jensen's inequality that an individual would be *risk averse,* that is, would always prefer a payment of $E[\tilde{x}]$ (and obtaining utility $U(E[\tilde{x}])$) to bearing the risk $\tilde{x}$ (and obtaining expected utility $E[U(\tilde{x})]$) if and only if his or her utility function were concave. This condition is illustrated in Fig. 1, where the random variable $\tilde{x}$ is assumed to take on the values x' and x'' with respective probabilities 2/3 and 1/3.

Of course, not all individuals need be risk averse in the sense of the previous paragraph. Another type of individual is a *risk lover.* Such an individual would have a *convex* utility

R



Risk, Fig. 2 Von Neumann–Morgenstern utility function of a risk-loving individual



Risk, Fig. 3 Portfolio analysis in the mean-standard deviation model

function, and would accordingly prefer receiving a random wealth $\tilde{x}$ to receiving its mean $E[\tilde{x}]$ with certainty. An example of such a utility function is given in Fig. 2.

Standard Deviation as a Measure of Risk

While the above characterizations of risk aversion and risk preference allow for the derivation of many results in the theory of choice under risk, they say nothing about which of a pair of non-degenerate random variables $\tilde{x}$ and $\tilde{y}$ is the more risky. Since real-world choices are almost never between risky and riskless situations but rather over alternative risky situations, such a means of comparison is necessary.

The earliest and best-known univariate measure of the riskiness of a random variable $\tilde{x}$ is its *variance* $\sigma^2 = E\left[(\tilde{x} - E[\tilde{x}])^2\right]$ or alternatively its *standard deviation* $\sigma = E\left[(\tilde{x} - E[\tilde{x}])^2\right]^{1/2}$. The tractability of these measures, as well as their well-known statistical properties, led to the widespread use of mean-standard deviation analysis in the 1950s and 1960s, and in particular to the development of modern portfolio theory by Markowitz (1952, 1959), Tobin (1958) and others. As an example of this, consider Fig. 3. Points A and B correspond to the distributions of a riskless asset with (per dollar) gross return r_0 and a risky asset with random return $\tilde{r}$ with mean $\mu_{\tilde{r}}$ and standard deviation $\sigma_{\tilde{r}}$. An investor dividing a dollar between the two assets in proportions α:(1 – α)

will possess a portfolio whose return has a mean of $\alpha \cdot r_0 + (1-\alpha) \cdot \mu_{\tilde{r}}$ and standard deviation $(1-\alpha) \cdot \sigma_{\tilde{r}}$, so that the set of attainable (μ, σ) combinations consists of the line segment connecting the points A and B in the figure. It is straightforward to show that, if the individual were also allowed to *borrow* at rate r_0 in order to finance purchase of the risky asset (that is, could sell the riskless asset short), then the set of attainable (μ, σ) combinations would be the ray emanating from A and passing through B and beyond.

If we then represent the individual's risk preferences by means of indifference curves in this diagram, we obtain his or her optimal portfolio (the example in the figure implies an equal division of funds between the two assets). In the more general case of choice between a pair of risky assets, the set of (μ, σ) combinations generated by alternative divisions of wealth between them will trace out a possibly nonlinear locus such as the one between points C and D in the diagram, with the curvature of this locus determined by the degree of statistical dependence (that is, covariance) between the two random returns.

As mentioned, the representation and analysis of risk and risk-taking by means of the variance or standard deviation of a distribution proved tremendously useful in the theory of finance, culminating in the mean-standard deviation-based *capital asset pricing model* of Sharpe (1964), Lintner (1965), Mossin (1966), and Treynor (1999). However, by the late 1960s the mean-standard deviation approach was under attack for two reasons.

The first reason (known since the 1950s) was the fact that an expected utility maximizer would evaluate all distributions solely on the basis of his or her means and standard deviations if and only if their von Neumann–Morgenstern utility function took the quadratic form $U(x) \equiv ax + bx^2$ for $b \Leftarrow 0$. The sufficiency of this condition is established by noting that $E[U(\tilde{x})] = E[a\tilde{x} + b\tilde{x}^2] = a \cdot E[\tilde{x}] + b \cdot \left(E[\tilde{x}]^2 + \sigma^2\right)$. To prove necessity, note that the distributions that yield a 2/3:1/3 chance of the outcomes $(x - \delta) : (x + 2\delta)$ and a 1/3:2/3 chance of the outcomes $(x - 2\delta) : (x + \delta)$ both possess the same mean and variance for each x and δ, so that $(2/3) \cdot U(x - \delta) + (1/3) \cdot U(x + 2\delta) \equiv (1/3) \cdot U(x - 2\delta) + (1/3) \cdot U(x + \delta)$ for all x and δ. Differentiating with respect to δ and simplifying yields $U'(x + 2\delta) + U'(x - 2\delta) \equiv U'(x + \delta) + U'(x - \delta)$ for all x and δ. This implies that $U'(\cdot)$ must be linear and hence that $U(\cdot)$ must be quadratic.

The assumption of quadratic utility is objectionable. If an individual with such a utility function is risk averse (that is, if $b < 0$), then (*a*) utility will decrease as wealth increases beyond $1/(2b)$, and (*b*) the individual will be more averse to constant additive risks about high wealth levels than about low wealth levels – in contrast to the observation that those with greater wealth take greater risks (see for example Hicks 1962, or Pratt 1964).

Borch (1969) struck the second and strongest blow to the mean-standard deviation approach. He showed that, for any two points (μ_1, σ_1) and (μ_2, σ_2) in the (μ, σ) plane which a mean-standard deviation preference ordering would rank as indifferent, it is possible to find random variables $\tilde{x}_1$ and $\tilde{x}_2$ which possess these respective (μ,σ) values and where $\tilde{x}_2$ first order stochastically dominates $\tilde{x}_1$. However, *any* person with an increasing von Neumann-Morgenstern utility function would strictly prefer $\tilde{x}_2$ to $\tilde{x}_1$. In response to these arguments and the additional criticisms of Feldstein (1969) and Samuelson (1967) and others, the use of mean-standard deviation analysis in economic theory waned. See, however, the work of Meyer (1987) for a partial rehabilitation of such two-moment models of preferences.

Besides the variance or standard deviation of a distribution, several other univariate measures of risk have been proposed. Examples include the *mean absolute deviation* $E[|\tilde{x} - E[\tilde{x}]|]$, the *interquartile range* $F^{-1}(.75) - F^{-1}(.25)$, and the classical statistical measures of *entropy* $\sum \ln (p_i) \cdot p_i$ or $\int \ln (f(x)) \cdot f(x) \cdot dx$. Although they provide the convenience of a single numerical index, each of these measures is subject to problems of the sort encountered with the variance or standard deviation. In particular, the entropy measure is based exclusively on the *probability levels* of a random variable, and is particularly unresponsive to its *outcome values* – for

R

example, the 50:50 gambles over the values \$49:\$51 and \$0:\$100 possess identical entropy levels.

Increasing Risk

By the late 1960s, the failure to find a satisfactory univariate measure of risk led to another approach to this problem. Working independently, several researchers (Hadar and Russell 1969; Hanoch and Levy 1969; Rothschild and Stiglitz 1970, 1971) developed an alternative characterization of increasing risk. The appeal of this approach is twofold. First, it formalizes three different intuitive notions of increasing risk. Second, it allows for the straightforward derivation of comparative statics results in a wide variety of economic situations. Unlike the univariate measures described above, however, this approach provides only a partial ordering of random variables. In other words, not all pairs of random variables can be compared with respect to their riskiness.

We now state three alternative formalizations of the notion that a cumulative distribution function $F^*(\cdot)$ is riskier than another distribution $F(\cdot)$ with the same mean. In the following, all distributions are assumed to be over the outcome interval $[0, M]$ unless otherwise indicated.

The first definition of increasing risk captures the notion that 'risk is what all risk averters hate'. Thus an increase in risk must lower the expected utility of all risk averters. Formally:

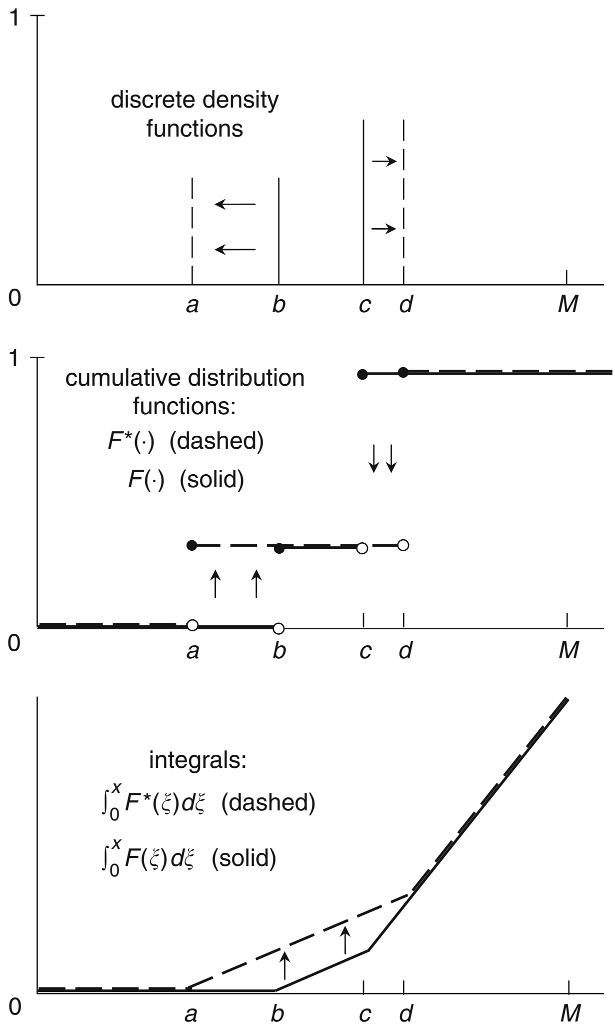
(A) *$F^*(\cdot)$ and $F(\cdot)$ have the same mean and $\int U(x)\,dF^*(x) \leq \int U(x)dF(x)$ for every concave utility function $U(\cdot)$.*

Note that this relationship will *not* be satisfied by every pair of distributions with the same mean. That is to say, there exist pairs $F(\cdot)$ and $F^*(\cdot)$, with the same mean, but such that some risk-averse utility functions prefer $F(\cdot)$ to $F^*(\cdot)$ but other risk-averse utility functions prefer $F^*(\cdot)$ to $F(\cdot)$. This reflects the above-stated fact that comparative risk is a *partial* rather than a *complete* order over the family of probability distributions, even over families of distributions with a common mean. (Although comparative risk is not a complete order, it is a *transitive* order, in the sense that, if the pair $F^*(\cdot)$ and $F(\cdot)$ satisfy condition (a), and the pair $F^{**}(\cdot)$ and $F^*(\cdot)$ satisfy condition (A), then the pair $F^{**}(\cdot)$ and $F(\cdot)$ will also satisfy condition (A).)

The second characterization of the notion that a random variable $\tilde{y}$ with distribution $F^*(\cdot)$ is riskier than a variable $\tilde{x}$ with distribution $F(\cdot)$ is that $\tilde{y}$ consists of the variable $\tilde{x}$ plus an additional zero-mean noise term $\tilde{\varepsilon}$. One possible specification of this is that $\tilde{\varepsilon}$ statistically independent of $\tilde{x}$. However, this condition is too strong in the sense that it does not allow the variance of $\tilde{\varepsilon}$ to depend upon the magnitude of $\tilde{x}$, as in the case of heteroskedastic noise. Instead, Rothschild and Stiglitz (1970) modelled the addition of noise by the condition:

(B) *$F(\cdot)$ and $F^*(\cdot)$ are the respective cumulative distribution functions of the random variables $\tilde{x}$ and $\tilde{x} + \tilde{\varepsilon}$, where $E[\tilde{\varepsilon}|\,x] \equiv 0$ for all values of x.*

The third notion of increasing risk involves the concept, due to Rothschild and Stiglitz (1970), of a *mean preserving spread*. Intuitively, such a spread consists of moving probability mass from some region in the centre of a probability distribution out to its tails in a manner that preserves the expected value of the distribution, as seen in the top panels of Figs. 4 and 5. In the discrete case of Fig. 4, probability mass is moved from the pair of outcome values b and c out to the outcome values a and d. In the continuous density case of Fig. 5, probability mass is moved from the interval (b, c) out to the intervals (a, b) and (c, d). We can unify, generalize and formalize this condition by saying that $F^*(\cdot)$ differs from $F(\cdot)$ by a 'mean preserving spread' if they have the same mean and there exists a single crossing point x_0 such that $F^*(x) \geq F(x)$ for all $x \leq x_0$ and $F^*(x) \leq F(x)$ for all $x \geq x_0$ (see the middle panels of Figs. 4 and 5). Since it is clear that *sequences* of such spreads will also lead to riskier distributions, the third characterization of increasing risk is:



Risk, Fig. 4 Mean preserving spread of a discrete distribution

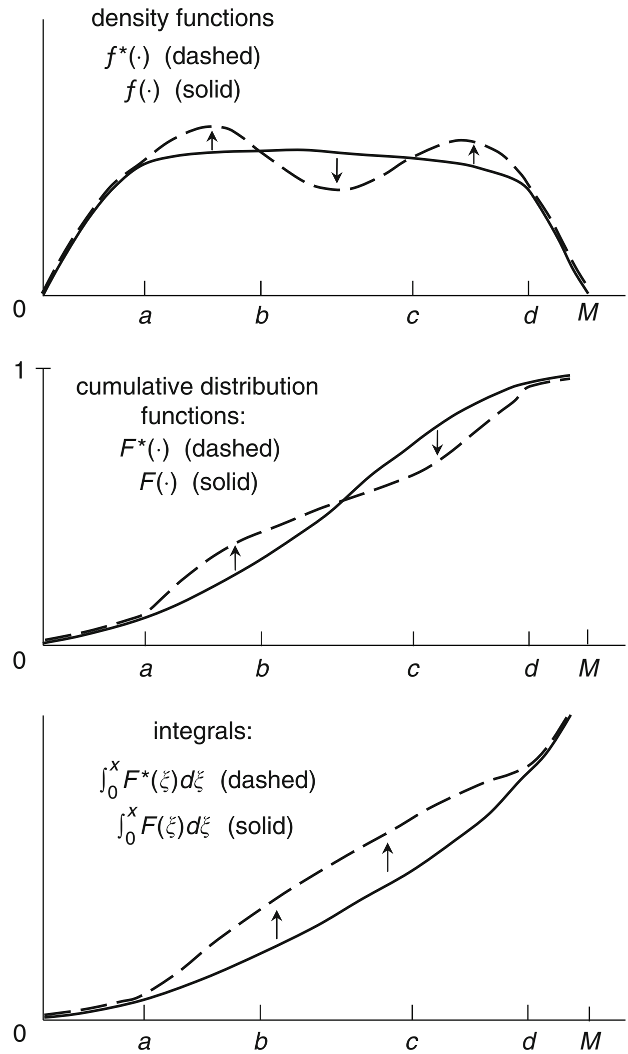
(C) *$F^*(\cdot)$ may be obtained from $F(\cdot)$ by a finite sequence, or as the limit of an infinite sequence, of mean preserving spread.*

Although the single crossing property of the previous paragraph serves to characterize cumulative distribution functions that differ by a single mean preserving spread, distributions that differ by a sequence of such spreads will typically not satisfy the single crossing condition. However, if we consider the integrals of these cumulative distribution functions, we see from the bottom panels of Figs. 4 and 5 that a mean preserving spread will always serve to raise or preserve the value of this integral for each x, and (since $F^*(\cdot)$ and $F(\cdot)$ have the same mean) will exactly preserve it for $x = M$. In contrast to the single crossing property, this so-called 'integral condition' *will* continue to be satisfied by distributions which differ by a sequence of one or more mean preserving spreads. Accordingly, we may rewrite condition (C) above by the analytically more convenient:

(C′) *The integral $\int_0^x [F^*(\xi) - F(\xi)] \cdot d\xi$ is nonnegative for all $x > 0$, and is equal to 0 at $x = M$.*

Rothschild and Stiglitz (1970) concepts of increasing risk are the same by proving that conditions (A), (B) and (C/C′) are equivalent. Thus, a

Risk, Fig. 5 Mean preserving spread of a density function



single partial ordering of distribution functions corresponds simultaneously to the notion that risk is what risk averters hate, to the notion that adding noise to a random variable increases its risk, and to the notion that moving probability mass from the centre of a probability distribution to its tails increases the riskiness of the distribution. The original Rothschild–Stiglitz formulation and proofs have since been further strengthened and extended by Machina and Pratt (1997).

This characterization permits the derivation of general and powerful comparative statics theorems concerning economic agents' responses to increases in risk. The general framework for these results is that of an individual with a von Neumann–Morgenstern utility function $U(x, \alpha)$ which depends upon both the outcome of some random variable $\tilde{x}$ as well as a *control variable* α which the individual chooses so as to maximize expected utility $\int U(x, \alpha)dF(x; r)$, where the distribution function $F(\cdot; r)$ depends upon some exogenous parameter r (x for example might be the return on a risky asset, and α the amount invested in it). For convenience, we assume that $F(0; r) \equiv prob(\tilde{x} \leq 0) \geq 0$ for all r. The first order condition for this problem is then:

$$\int U_{\alpha}(x,\alpha)dF(x;r) = 0 \qquad (2)$$

where $U_{\alpha}(x, \alpha) = \partial U(x, \alpha)/\partial\alpha$, and we assume that the second derivative $U_{\alpha\alpha}(x, \alpha) = \partial^2 U(x, \alpha)/\partial\alpha^2$ is always negative to insure we have a maximum. Implicit differentiation of (2) then yields the comparative statics derivative:

$$d\alpha/dr = \begin{array}{l} -\int U_{\alpha}(x,\alpha)dF_r(x;r)/\int U_{\alpha\alpha}(x,\alpha) \\ \times dF(x;r) \end{array} \qquad (3)$$

where $F_r(x; r) = \partial F(x; r)/\partial r$. Since the denominator of this expression is negative by assumption, the sign of $d\alpha/dr$ is given by the sign of the numerator $\int U_{\alpha}(x, \alpha)dF_r(x; r)$. Integrating by parts twice yields:

$$\begin{aligned} \int U_{\alpha}(x,\alpha)dF_r(x;r) &= \int U_{xx\alpha}(x,\alpha) \\ &\quad \cdot \left[\int_0^x F_r(\xi;r)d\xi\right]dx \\ &= \int U_{xx\alpha}(x,\alpha) \\ &\quad \cdot \left[\frac{d}{dr}\int_0^x F_r(\xi;r)d\xi\right]dx \end{aligned} \qquad (4)$$

Thus, if increases in the parameter r imply increases in the riskiness of the distribution $F(\cdot, r)$, it follows from condition (C′) that the signs of the squarebracketed terms in (4) will be nonnegative, so that the effect of r upon α depends upon the sign of $U_{xx\alpha}(x, \alpha) = \partial^3 U(x, \alpha)/\partial x^2 \partial a$. Thus, if $U_{xx\alpha}(x, \alpha)$ is uniformly negative a mean preserving increase in risk in the distribution of x will lead to a fall in the optimal value of the control variable α, and vice versa. Another way to see this is to note that if $U_{\alpha}(x, \alpha)$ is concave in x then a mean preserving increase in risk will lower the left side of the first order condition (2), which (since $U_{\alpha\alpha}(x, \alpha) \leq 0$) will require a drop in α to re-establish the equality. Economists, mathematicians and scientists routinely use this technique when analysing models involving risk; see for example Rothschild and Stiglitz (1971), Dionne et al. (1993), Eeckhoudt et al. (1996), Jewitt (1987), Ormiston (1992), Tzeng (2001), Nowak (2004), Chateauneuf et al. (2004), Baker (2006), and Beladi et al. (2006).

Related Topics

The characterization of risk outlined in the previous section has been extended along several lines. Diamond and Stiglitz (1974), for example, have replaced the notion of a mean preserving spread with that of a mean *utility* preserving spread to obtain a general characterization of a *compensated increase in risk.* They relate this notion to the well known Arrow-Pratt characterization of comparative risk aversion (see expected utility hypothesis).

In addition, researchers such as Ekern (1980), Fishburn (1982), Fishburn and Vickson (1978), Hansen et al. (1978), Tesfatsion (1976), and Whitmore (1970) have extended the above work to the development of a general theory of *stochastic dominance*, which provides a whole sequence of similarly characterized partial orders on distributions, each presenting a corresponding set of equivalent conditions involving algebraic conditions on the distributions, types of spreads, and classes of utility functions which prefer (or are averse to) such spreads. The comparative statics analysis presented above may be similarly extended to such characterizations. An extensive bibliography of the stochastic dominance literature is given in Bawa (1982). Finally, various extensions of the notions of increasing risk and stochastic dominance to the case of multivariate distributions may be found in Epstein and Tanny (1980), Fishburn and Vickson (1978), Huang et al. (1978), Lehmann (1955), Levhari et al. (1975), Levy and Parousch (1974), Russell and Seo (1978), Sherman (1951), and Strassen (1965); see also the mathematical results in Marshall and Okun (1979).

See Also

▶ Expected Utility Hypothesis
▶ Uncertainty

Bibliography

Anscombe, F., and R. Aumann. 1963. A definition of subjective probability. *Annals of Mathematical Statistics* 34: 199–205.

Arrow, K. 1964. The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies* 31: 91–96.

Baker, E. 2006. Increasing risk and increasing informativeness: Equivalence theorems. *Operations Research* 54: 26–36.

Bawa, V. 1982. Stochastic dominance: A research bibliography. *Management Science* 28: 698–712.

Beladi, H., L. de la Vina, and F. Firoozi. 2006. On information value and mean-preserving transformations. *Applied Mathematics Letters* 19: 843–848.

Borch, K. 1969. A note on uncertainty and indifference curves. *Review of Economic Studies* 36: 1–4.

Chateauneuf, A., M. Cohen, and I. Meilijson. 2004. Four notions of mean-preserving increase in risk, risk attitudes and applications to the rank-dependent expected utility model. *Journal of Mathematical Economics* 40: 547–571.

Debreu, G. 1959. *Theory of value: An axiomatic analysis of general equilibrium*. New Haven: Yale University Press.

Diamond, P., and M. Rothschild, eds. 1989. *Uncertainty in economics: Readings and exercises*. 2nd ed. New York: Academic.

Diamond, P., and J. Stiglitz. 1974. Increases in risk and in risk aversion. *Journal of Economic Theory* 8: 337–360.

Dionne, G., L. Eeckhoudt, and C. Gollier. 1993. Increases in risk and linear payoffs. *International Economic Review* 34: 309–319.

Eeckhoudt, L., C. Gollier, and H. Schlesinger. 1996. Changes in background risk and risk taking behavior. *Econometrica* 64: 683–689.

Ekern, S. 1980. Increasing nth degree risk. *Economics Letters* 6: 329–333.

Epstein, L., and S. Tanny. 1980. Increasing generalized correlation: A definition and some economic consequences. *Canadian Journal of Economics* 13: 16–34.

Feldstein, M. 1969. Mean-variance analysis in the theory of liquidity preference and portfolio selection. *Review of Economic Studies* 36: 5–12.

Fishburn, P. 1982. Simplest cases of n'th degree stochastic dominance. *Operations Research Letters* 1: 89–90.

Fishburn, P. and Vickson, R. 1978. Theoretical foundations of stochastic dominance, in Whitmore and Findlay (1978).

Gärdenfors, P., and N.-E. Sahlin, eds. 1988. *Decision, probability, and utility: Selected readings*. Cambridge: Cambridge University Press.

Hadar, J., and W. Russell. 1969. Rules for ordering uncertain prospects. *American Economic Review* 59: 25–34.

Hanoch, G., and H. Levy. 1969. The efficiency analysis of choices involving risk. *Review of Economic Studies* 36: 335–346.

Hansen, L., C. Holt, and D. Peled. 1978. A note on first degree stochastic dominance. *Economics Letters* 1: 315–319.

Hey, J., ed. 1997. *The economics of uncertainty, vol. II: Uncertainty and dynamics*. Cheltenham: Edward Elgar.

Hicks, J. 1962. Liquidity. *Economic Journal* 72: 787–802.

Hirshleifer, J. 1965. Investment decision under uncertainty: Choice-theoretic approaches. *Quarterly Journal of Economics* 79: 509–536.

Hirshleifer, J. 1966. Investment decision under uncertainty: Applications of the state- preference approach. *Quarterly Journal of Economics* 80: 252–277.

Huang, C., D. Kira, and I. Vertinsky. 1978. Stochastic dominance for multi-attribute utility functions. *Review of Economic Studies* 45: 611–616.

Jewitt, I. 1987. Risk aversion and the choice between risky prospects: The preservation of comparative statics results. *Review of Economic Studies* 54: 73–85.

Karni, E. 1985. *Decision making under uncertainty: The case of state dependent preferences*. Cambridge, MA: Harvard University Press.

Knight, F. 1921. *Risk, uncertainty and profit*. Boston: Houghton Mifflin.

Kyberg, H., and H. Smokler, eds. 1964. *Studies in subjective probability*. New York: Wiley.

Lehmann, E. 1955. Ordered families of distributions. *Annals of Mathematical Statistics* 26: 399–419.

Levhari, D., J. Paroush, and B. Peleg. 1975. Efficiency analysis for multivariate distributions. *Review of Economic Studies* 42: 87–91.

Levy, H., and J. Paroush. 1974. Toward multivariate efficiency criteria. *Journal of Economic Theory* 7: 129–142.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 44: 243–269.

Machina, M., and J. Pratt. 1997. Increasing risk: Some direct constructions. *Journal of Risk and Uncertainty* 14: 103–127.

Machina, M., and D. Schmeidler. 1992. A more robust definition of subjective probability. *Econometrica* 60: 745–780. Reprinted in Hey (1997).

Markowitz, H. 1952. Portfolio selection. *Journal of Finance* 7: 77–91.

Markowitz, H. 1959. *Portfolio selection: Efficient diversification of investment*. New Haven: Yale University Press.

Marshall, A., and I. Okun. 1979. *Inequalities: Theory of majorization and its applications*. New York: Academic.

Meyer, J. 1987. Two-moment decision models and expected utility maximization. *American Economic Review* 77: 421–430.

Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–783.

Nowak, M. 2004. Interactive approach in multicriteria analysis based on stochastic dominance. *Control and Cybernetics* 33: 463–476.

Ormiston, M. 1992. First and second degree transformations and comparative statics under uncertainty. *International Economic Review* 33: 33–44.
Pratt, J. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–136. Reprinted in Diamond and Rothschild (1989).
Pratt, J., H. Raiffa, and R. Schlaifer. 1964. The foundations of decision under uncertainty: An elementary exposition. *Journal of the American Statistical Association* 59: 353–375.
Raiffa, H. 1968. *Decision analysis: Introductory lectures on choice under uncertainty*. Reading: Addison Wesley.
Ramsey, F. 1926. Truth and probability. In *The foundations of mathematics and other logical essays*, ed. R.-B. Braithwaite. London: Routledge & Kegan Paul. 1931. Reprinted in Kyberg and Smokler (1964), Ramsey (1978), and Gärdenfors and Sahlin (1988).
Ramsey, F. 1978. In *Foundations: Essays in philosophy, mathematics and economics*, ed. D.H. Mellor. London: Routledge & Kegan Paul.
Rothschild, M., and J. Stiglitz. 1970. Increasing risk: I. A definition. *Journal of Economic Theory* 2: 225–243. Reprinted in Diamond and Rothschild (1989).
Rothschild, M., and J. Stiglitz. 1971. Increasing risk: II. Its economic consequences. *Journal of Economic Theory* 3: 66–84.
Rothschild, M., and J. Stiglitz. 1972. Addendum to 'increasing risk: I. a definition'. *Journal of Economic Theory* 5: 306.
Russell, W., and T. Seo. 1978. Ordering uncertain prospects: The multivariate utility functions case. *Review of Economic Studies* 45: 605–611.
Samuelson, P. 1967. General proof that diversification pays. *Journal of Financial and Quantitative Analysis* 2: 1–13.
Savage, L. 1954. *The foundations of statistics*. New York: Wiley. Rev. edn, New York: Dover Publications, 1972.
Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.
Sherman, S. 1951. On a theorem of Hardy, Littlewood, Polya, and Blackwell. *Proceedings of the National Academy of Science* 37: 826–831. Errata. *Proc. Nat. Ac. Sci.* 38, 382.
Strassen, V. 1965. The existence of probability measures with given marginals. *Annals of Mathematical Statistics* 36: 423–439.
Tesfatsion, L. 1976. Stochastic dominance and the maximization of expected utility. *Review of Economic Studies* 43: 301–315.
Tobin, J. 1958. Liquidity preference as behavior toward risk. *Review of Economic Studies* 25: 65–86.
Treynor, J. 1999. Toward a theory of market value of risky assets. In *Asset Pricing and Portfolio Performance*, ed. R. Korajczyk. London: Risk Books. In Korajczyk (1999).
Tzeng, L. 2001. Increase in risk and weaker marginal-payoff-weighted risk dominance. *The Journal of Risk and Insurance* 68: 329–337.
Whitmore, G. 1970. Third-degree stochastic dominance. *American Economic Review* 60: 457–459.
Whitmore, G., and M. Findlay. 1978. *Stochastic dominance: An approach to decision making under risk*. Lexington: Heath.
Yaari, M. 1969. Some remarks on measures of risk aversion and on their uses. *Journal of Economic Theory* 1: 315–329. Reprinted in Diamond and Rothschild (1989).

Risk Adjustment

Randall P. Ellis

Abstract

Risk adjustment is used in settings with uncertainty to make payments or allow comparisons of outcomes while controlling for exogenous risk factors that explain variations in the outcome of interest, such as spending, utilisation, quality or death. This article focuses on the conceptual and empirical uses of risk adjustment in health economics, where patient-level risk factors are commonly used to explain spending and other outcomes.

Keywords

Adverse selection; Biased selection; Health economics; Health insurance; Health outcomes; Life insurance; Life tables

JEL Classifications

G22; I11; I13; I182

R

Risk Adjustment

The term 'risk adjustment' is used predominantly in health economics to describe the use of exogenous risk factors to explain variations in health care spending, utilisation, quality or outcomes of interest, such as death or health status (van de Ven and Ellis 2000; Ellis 2008). The term is also used by actuaries and others when predicting other outcomes that vary systematically with covariates. For example, actuaries perform risk adjustment when using life tables to set premiums for life insurance that reflect age and gender (Gründl et al. 2006), when calculating premiums for property insurance related to geography or rate classes, and when adjusting for risk in the finance literature (e.g. Constantinides 1978). However, the most common use of the term is in health care markets, and that is the focus of the remainder of this article.

The American Academy of Actuaries (2010) defines risk adjustment as ‘an actuarial tool used to calibrate payments to health plans or other stakeholders based on the relative health of the at-risk populations’. Risk adjusting payments to health plans by a payer is an important tool for reducing incentives for health plans to adopt strategies that induce favourable selections, and avoid market failures due to adverse selection. Risk adjustment is widely used in publicly funded insurance programs, including US Medicare and Medicaid managed care programs, and the competitive insurance markets in Belgium, Germany, Israel, Netherlands and Switzerland. Risk-adjusted payments play a key role in many proposals to broaden access to insurance and promote incentives for low-cost effective health care, including the health insurance exchanges proposed as part of the Patient Protection and Accountable Care Act (PPACA) of 2010.

Risk adjustment is also commonly used in program evaluation where the interest is in normalising populations with different underlying risks so as to compare outcomes. This use is also often called ‘case-mix’ or ‘severity adjustment’ particularly when used to explain variation in outcomes for a particular procedure or episode of treatment, notably including the literature on Diagnosis Related Groups (DRGs) as introduced by Fetter et al. (1980). It would promote clarity among economists to distinguish risk adjustment from case-mix adjustment, although, the classic book on risk adjustment, edited by physician Lisa Iezzoni (2003) uses risk adjustment to refer to all kinds of risk, case-mix and severity adjustment. There is a rich literature that focuses on the actuarial use of risk adjustment for groups of enrollees (see especially Rice and Smith (2001) and Duncan (2011)), but increasingly, as data improve, risk adjusting based on individual rather than group level data has become the norm. In recent years the terminology ‘predictive modeling’ has come to be used in the USA for models designed to predict individual-level health care utilisation without regard to whether or not the predictive model will be used for payment. Such models can be useful for identifying case management needs or for identifying groups deserving of greater attention independent of payment or incentive concerns.

Theory of Risk Adjustment

The starting point for risk adjustment is the fact that health care spending and utilisation vary in a population for many reasons: health status; health plan coverage (which affects consumer prices and in some cases, utilisation); consumer taste and demand variation (including income); and access measures such as distance to providers, medical technology, provider practice style and provider fees. The classic risk adjustment problem is to control for health status variation while examining how one or more of the other factors also affects spending. Because risk adjustment models can explain much of the variation in a sample, they increase the ability of regression models to detect behavioural changes due to policy interventions. Early risk adjustment efforts focused on using information to capture health status, including self-reported health and chronic conditions, as well as claims-based diagnoses (Ash et al. 1989) Pharmacy information is also used to capture health status, particularly when inpatient diagnoses but not outpatient diagnoses are available (Von Korff et al. 1992) Using prescription drugs information is controversial since usage seems to be more endogenous than diagnostic coding. Nonetheless the Netherlands uses prescription drug data together with inpatient diagnoses as a central part of its risk-adjusted equalisation of funds across competing health plans (van de Ven et al. 2004).

'Conventional risk adjustment' focuses on the statistical problem of maximising the amount of variance in total spending that can be explained with available information. In this work it is frequently noted that although lagged utilisation and spending information is highly predictive, when included in payment models this information not only capture the underlying illness burdens, but also picks up other information, such as consumer taste for treatment and provider practice variation, which one is often trying to avoid rewarding for incentive reasons.

An early contribution to the conventional risk adjustment literature by Newhouse et al. (1989) used fixed effects in panel data to calculate that a 'lower bound on the upper bound' of what is explainable at the individual level using time invariant information is only on the order of 20–25% of total health care spending variation. That is, no model using lagged information should be expected to achieve a higher R^2 when predicting total spending. The potentially achievable R^2 varies with the population, year and data quality, but the upper bound remains on the order of 30% in more recent samples. This limit is very humbling, until one realises that predictability over a one-year period at the individual level in many other insurance settings (e.g. fire, life, property) is even more difficult.

Glazer and McGuire (2000, 2002) were the first to develop careful theoretical models that characterise 'optimal risk adjustment' which optimised a social objective function over a given payment system. Whereas a hallmark of conventional risk adjustment is the goal of unbiasedness – paying each plan or normalising utilisation measures so that predicted levels equal actual levels – the essence of optimal risk adjustment is to allow biased payment models, so that predicted payments do not necessarily track expected actual payments. Glazer and McGuire model the selection problem as one in which competing health plans oversupply services that attract the healthy (e.g. acute care), and undersupply services that disproportionately attract the high-cost, relatively sick (e.g. chronic care services). Since the signals used for risk adjustment are never perfect, by distorting services in this way, even with conventional risk adjustment paying the expected costs, it will be optimal for health plans to distort service offerings so as to attract healthy enrollees within a payment category, and deter the relatively sick. (Ellis and McGuire (2007) document that there is evidence that Medicare health plans have incentives to distort services precisely in this way, even with a rich conventional risk adjustment.) The solution Glazer and McGuire devise is to overpay on signals predicting a greater likelihood of being high cost, and underpay on signals predicting low cost, so as to undo the incentive to undertreat high-cost enrollees. For example, if only half of patients with asthma in a plan have their diagnoses recorded in the base period, and the incremental cost of the observed asthma patients is $500 higher than expected, then the plan should be paid twice this increment, or $1000 to compensate the plan for the under-reported patients with asthma. Conversely, one should pay less than the observed average cost for healthy signals in order to keep overall payments neutral. This twist in payments can in theory undo incentives to undertreat in capitated payment systems. Alternative models of optimal risk adjustment are further developed in Shen and Ellis (2002) and Jack (2006).

There are several challenges to implementing the optimal risk adjustment formulas suggested by Glazer and McGuire. These include the substantial amount of information needed to assess the levels of under- and over-reporting of each signal used for risk adjustment; understanding the reaction function of providers to alternative marginal risk adjustment payment rates; and ensuring that all of the relevant provider behaviour is captured by the model. Glazer and McGuire model the problem as one of service distortion to attract a favourable selection, while policy makers have typically been more concerned about increased incentives to upcode patient severity by recording more diagnoses once risk adjustment is implemented. Optimal risk adjustment is an important concept to keep in mind when using and designing payment formulas, but to date it has proven challenging to implement in practice.

Information Used for Risk Adjustment

Ellis (2008) reviews the different types of information that can be used to predict health care spending. Variation in spending can be decomposed conceptually and empirically into variation due to patient characteristics, characteristics of the providers they see (e.g. specialists, general practitioners, hospitals, nurses), and the prices of the services provided. Depending on the purpose, all of this information may be useful for prediction, but it is not necessarily exogenous, and good instruments to control for endogeneity are scarce. Patient characteristics can be further decomposed into variations due to the underlying health status of the patient, socioeconomic variables such as income and education, geographical location and the benefit design of their insurance all of which influence access and utilisation. The best set of information to use for risk adjustment depends upon the intended use. Health-based payment models and severity adjustment models restrict the information set to use only health status. Needs-adjusted payment, widely used in Europe, broadens the information to reflect further demographic variables such as income, race, geography and access (e.g. distance). Van de Ven et al. (2003) provide a useful overview of how risk adjustment is being done in five European countries with multiple, competing health plans.

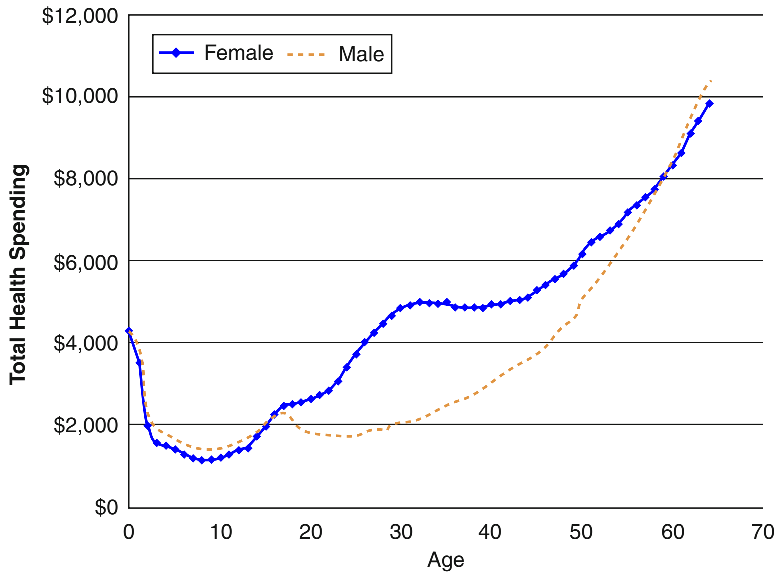
Among the privately insured in the USA, risk adjustment models typically use both diagnoses and pharmacy information for prediction. Because family insurance contracts covering more than one individual are common, risk adjustment models are typically used to generate predictions at the individual level and then sum them up to the family level to make a contract level prediction that is useful for calculating family premiums or assessing family level biased selection. This approach ignores within-family correlations in spending, particularly of the sort predicted by family level deductibles and stop losses (Eichner 1998).

For the US Medicare Advantage (Part C) program, which is offered as a voluntary alternative to conventional Medicare to all enrollees, risk-adjusted payments to health plans from 1985 to 1999 used only age, gender, Medicaid eligibility, institutional status (i.e. whether in a nursing home) and the county of residence of the enrollee. Since 2000 risk adjustment in the US Medicare program has used diagnostic information, initially using only inpatient diagnoses, but since 2004 diagnoses from outpatient clinician claims have also been used (CMS 2006). After considering numerous alternative classification systems for diagnostic information, the Medicare program chose to implement the CMS Hierarchical Condition Category (CMS-HCC) classification system using 70 diagnostic groups for prediction (Ash et al. 2000; Pope et al. 2004). As of 2011, up to 86 HCCs are used, and the system is also used for Medicare Part D which includes prescription drug plans (Robst et al. 2007).

In the United Kingdom (UK), risk adjustment has been used for many years to allocate funds between geographically defined ‘Primary Care Trusts’ (PCTs) using a variety of needs, and done at the group level. Rice and Smith (2001) provide an overview of this approach. More recent efforts in the UK have considered using individual information for risk adjusting payments not only to the geographically defined PCTs, but also to individual general practitioners (Dixon et al. 2011). The main drawback to using individual level diagnostic information has been to obtain this information from office-based physicians, who are not required to record diagnoses as a condition of service payments. Dixon et al. demonstrate that models using only inpatient diagnoses, and counts of office visits and facility visits, have impressively high explanatory power, as discussed further below.

Empirical Risk Adjustment Models

The classic approach to risk adjustment is to use only truly exogenous information, such as age and gender. Figure 1 illustrates the importance of using relatively flexible specifications for capturing age and sex adjustment of total health spending. The figure highlights that babies are disproportionately expensive, childhood males are slightly more expensive than females, and



Risk Adjustment, Fig. 1 Actual health spending by age and gender, US Thomson-Reuters MarketScan Commercially Insured Sample, 2009 ($N = 30{,}333{,}041$). Notes: Sample used is the US 2009 Thomson-Reuters MarketScan Commercially Insured Claims and Encounter Data. All plan types and individual with a valid sex and age < 65 were included, although persons without pharmacy coverage were excluded. Each point plotted is the one year average total covered health spending per capita (medical plus pharmacy spending, including deductibles and copayments, but excluding dental and vision spending) for that one year age and gender group. Means for newborns are adjusted for the fact that babies are on average only present for half a year (Source: Author's original figure)

women cost more than men through their childbearing years.. These patterns are poorly captured by including a linear age term or even by using third or fourth degree polynomials of age, hence most sophisticated risk adjustment models calibrated on large samples use 30 or more age–sex dummy variables to capture this nonlinear pattern (Ash et al. 1989; Pope et al. 2004; Dixon et al. 2011).

Rather than only using exogenous age and gender, the most common approach to risk adjustment is to use the rich information appearing on insurance claims as a proxy for individual health status. The most widespread information used is diagnoses, although pharmacy information is also common. Utilisation measures (e.g. spending, hospitalisations and counts of visits) are also highly predictive of future spending, although they contribute relatively modestly to the predictive power when a rich diagnostic model is used. Although claims-based information is only recorded when a visit to a health care provider is made, and is potentially 'gameable' or amenable to manipulation, their strong predictive power and availability make them highly attractive.

There are a number of careful reviews of alternative risk adjustment models. These include US classics by Ingber (1998) and Pope et al. (1998), the Society of Actuaries (Winkelman and Mehmud 2007), provider profiling models by Thomas et al. (2004), and comparisons done in Canada Berlinguet et al. (2005), Germany (Wasem et al. 2006) and the impressively comprehensive study of multiple classification systems and information datasets from the UK by Dixon et al. (2011). A review of all of the different classification systems available for classifying diagnoses and pharmacy information would take much longer than the space of this article. A glimpse at the dimensions along which five of the major diagnosis based models vary is summarised in Table 1, which was compiled by the State of Florida Medicaid program when comparing alternative diagnosis-based models.

Risk Adjustment, Table 1 Risk adjustment model comparison, State of Florida Medicaid Program, May 2009

Model feature	Adjusted clinical groups (ACGs)	Chronic-illness disability payment system (CDPS)	Clinical risk groups (CRGs)	Diagnostic risk group (DCG)	Episode risk groups (ERGs)
Background					
Model developer	Johns Hopkins	University of California, San Diego (UCSD)	3 M Health Information Systems	Verisk Health (formerly DxCG)	Ingenix (formerly Symmetry)
Marketplace introduction	1992	1996	2000	1996	2001
Disease classification					
Additive/ categorical classification	Categorical	Additive	Categorical	Additive	Additive
Diagnoses (Dx)	Single diagnosis	Single diagnosis	Single diagnosis from inpatient facility or two diagnoses from professionals	Single diagnosis	Single diagnosis from face-to-face encounter or inpatient admissions
Conditions included	Acute and chronic	Chronic only	Acute and chronic	Acute and chronic	Acute and chronic
Model users					
Programs to adjust capitation payments Commercial	175	None	7	300+	60
Estimation capabilities[a]		(Prospective R-squared)			
Without				17.8%	16.4%
Truncation					

Notes: [a]Ross Winkelman, FSA. 2007. *A Comparative Analysis of Claims-based Tools for Health Risk Assessment*, 20 April, Appendix A.

[b]Based on a Medicaid case study, the CRG model's performance was in between the other two models within the study: ACG and CDPS.

For further details of each of the systems listed, see Weiner et al. 1991; Kronick and Dreyfus 1997; Averill et al. 1999; Ash et al. 2000; Symmetry Health Data Systems, Inc., 2001. Source for this table: http://ahca.myflorida.com/Medicaid/quality_management/workgroups/managed_care/5_rar_model_comparison_050709.pdf

Table 2 illustrates a number of important findings from Dixon et al. (2011), using UK data, which have also been shown in other countries. Looking first across the rows, age and gender alone only explain about 3–5% of total variation in spending at the individual level. Also striking is that once diagnostic and utilisation information are included in model (b) surprisingly little further variation is explained by including geographic variation (as captured by 152 primary care trust dummies, which are geographical), 135 need variables (e.g. income, education, and prevalence of selected chronic conditions in the area) and 63 supply side variables (e.g. numbers of providers of various types and distances). Explanatory power at the individual level as measured by R^2 differs only in the third or fourth decimal. The final row reveals that dropping the four utilisation variables has a more significant effect on the model's predictive power, losing about half of the model's explanatory power. Many would argue that the four lagged utilisation variables are not only picking up health status heterogeneity, but also patient and provider taste variation.

Risk Adjustment, Table 2 Selected results from Dixon et al. (2011) predicting FY2008 health spending per capita using prior two years of UK data

ID	Explanatory variables in OLS models:	Number of parameters	Individual level R^2		Practice level R^2
			Estimation Sample	Validation Sample #1	Validation Sample #2
			N = 5,206,651	N = 5,205,747	N = 797
a.	Age and gender only	38	0.0373	0.0366	0.3444
b.	Model (a) plus 152 diagnosis groups and 4 lagged utilisation variables	194	0.2656	0.2610	0.7394
c.	Model (b) plus 151 geographic dummies	345	0.2659	0.2612	0.8046
d.	Model (c) plus 135 attributed need and 63 supply variables	543	0.2662	0.2615	0.8254
e.	Model (c) plus 7 attributed need and 3 supply variables	355	0.2671	0.2622	0.8254
f.	Age/gender, 152 diagnosis groups, 151 geographic dummies, 7 attributed need and 3 supply variables	351	0.1272	0.1229	0.7738

Notes: Diagnosis groups use only inpatient diagnoses from a two year period. Utilisation variables include inpatient episode count, outpatient visit count, dummy = 1 if any priority referral, and dummy = 1 if any outpatient visit; all measures are for prior two years. Estimation sample is a 10% random sample of the UK population. Validation Sample #1 is a different 10% random sample of the UK population drawn without replacement. Validation Sample #2 is a 100% sample of patients at 10% of primary care practices. See further details in Dixon et al. (2011), especially Table 7.4 and Appendix 13, Table 9. http://www.nuffieldtrust.org.uk/sites/files/nuffield/document/Developing_a_personbased_resource_allocation_formula_REPORT.pdf

(Key need and supply side variables are still included in the model.)

Looking across the columns of Table 2 reveals that with 5 million observations in the estimation sample, there is no overfitting problem, even with over 500 righthand side explanatory variables. The final column shows that despite having only modest explanatory power at the individual level, where there is a great deal of individual patient randomness, the models do enormously better at the practice level, where much of this randomness averages out. The third column sums up patient actual and predicted spending to the level of 797 primary care practices (averaging just over 6,500 patients per practice) before using the conventional R^2 to calculate predictive power. The explained variation in spending at the practice level (the R^2) starts at 34% for age and gender, and increases to just over 80% once geographic dummies are added in. Even the final model, which does not use the four utilisation variables capturing patient and provider taste variation, explains 77% of the practicelevel variation in spending.

Econometric Issues

Risk adjustment models have been an active area for testing and developing new estimation techniques. Early models were primarily linear least squares models, in part because the very large sample sizes and large number of explanatory variables made estimation of nonlinear models time-consuming, if not infeasible (Ash et al. 1989). Since the 1990s and 2000s, there has been a surge of interest in building robust nonlinear models that are less sensitive to the large outliers that are common in highly skewed expenditure data. The two-part log linear model pioneered by Duan et al. (1983), and used so widely in the Rand Health Insurance Experiment (Newhouse 2002), was largely laid to rest by Manning and Mullahy (2001) who demonstrated the severe problems caused by uncorrected heterogeneity in such models. Basu et al. (2004) show the superiority of Cox Proportional Hazard models, while Buntin and Zaslavsky (2004) implemented the Generalised Linear Models (GLM), which Manning et al. (2005) further

R

refine, developing an algorithm for choosing among alternative, non-nested model specifications.

In recent years there has been a return of support for least squares models. The preferred approach adopted by the US Medicare program consistently since the 1980s is to use weighted least squares on annualised spending, which is to say that actual spending is divided by the fraction of the year a person is eligible to annualise, and this annualised amount is weighted by the fraction of the year a person is eligible to generate unbiased means. Such an approach replicates the mean exactly in disjoint groups, and is the only demonstrated approach that easily accommodates individuals with partial year eligibility (Ellis et al. 1996). The megasamples of multiple millions of observations, as shown in Fig. 1 and Table 2 in this article, largely alleviate concerns about overfitting: measures of goodness of fit and statistical significance are not overstated with large samples, even with huge outliers and skewness.

Future Directions in Risk Adjustment

As mentioned earlier, risk adjustment figures prominently in the US PPACA of 2010, notably in the proposals for establishing insurance exchanges to serve the individual and small group insurance markets. To keep insurance affordable, premium subsidies will be offered by the government, and premium rate bands will limit premiums variations across age and gender groups to be no more than three to one. It is readily seen from Fig. 1 above that in the absence of regulation, plans would choose to charge 64-year-old males a premium that is nearly ten times that of a 10-year-old male. Such regulated premiums can only be feasible if premium subsidies to plans are risk adjusted so that plans are paid for enrolling the aged and relatively unhealthy.

A second important area for risk adjustment is in bundled payments to Accountable Care Organizations (ACO), which are moderate-size health care provider networks willing to receive a bundled payment in exchange for taking responsibility for providing all care to a panel of patients. Given the modest size of these panels, risk adjustment will be critical for ensuring that both healthy and sick enrollees are welcomed in the ACO.

A third important area for risk adjustment is in bundled payments for primary care, particularly as part of the Patient-Centered Medical Home. In this CMS initiative, the Medicare program is encouraging primary care providers to take responsibility for providing comprehensive primary care for patients from all payers (Medicare, Medicaid and private) and offering increased primary care ‘base payments’ for the extra effort this will take, payments that will be partial capitation amounts, not fee-based. Sizeable bonus payments are also contemplated to reward primary care practices for achieving specified quality, cost and patient satisfaction targets. If either the base payments or bonus payments are not risk adjusted, then primary care practices could potentially act like insurance companies and strive to avoid serving the relatively sick and to attract the healthy, undermining the potential of the PCMH initiative.

Most work on risk adjustment has used a static framework in which signals are taken as given and spending or utilisation in only one period is modelled. Eggleston et al. (forthcoming) show that different payment amounts are needed once prevention and multiple periods are involved, to reward providers for keeping patients healthy. Further work is needed conceptually and empirically on how to refine risk adjustment models in the presence of prevention and treatments that affect outcomes in multiple periods.

To date, risk adjustment models in the USA have relied primarily on demographic and claims-based information to adjust payments, utilisation and outcome measures. Occasionally self-reported information is used, although the relatively high cost of surveys and consumer input limit the widespread adoption of this information. A potentially huge source of information for the future are electronic health records, which capture not only what treatments are done, but also the results of various biometric and laboratory tests and imaging procedures. Health records will be challenging to use, but offer rich possibilities for improved prediction of diverse outcomes of key interest to researchers and policymakers.

See Also

▶ Risk
▶ Risk Sharing

Bibliography

American Academy of Actuaries. 2010. *Issue brief: Risk assessment and risk adjustment*. May 2010. http://www.actuary.org/pdf/health/Risk_Adjustment_Issue_Brief_Final_5-26-10.pdf. Accessed 7 Feb 2012.

Ash, A.S., F. Porell, L. Gruenberg, et al. 1989. Adjusting medicare capitation payments using prior hospitalization data. *Health Care Financing Review* 10(4): 17–29.

Ash, A.S., R.P. Ellis, G.C. Pope, J.Z. Ayanian, D.W. Bates, H. Burstin, L.I. Iezzoni, E. McKay, and W. Yu. 2000. Using diagnoses to describe populations and predict costs. *Health Care Financing Review* 21(3): 7–28.

Averill, R.F., N.I. Goldfield, J. Eisenhandler, J.S. Hughes, B.V. Shafir, and D.E. Gannon, et al. 1999. *Development and evaluation of clinical risk groups (CRGs)*, 3M health information systems. Wallingford.

Basu, A., W.G. Manning, and J. Mullahy. 2004. Comparing alternative models: Log vs. Cox proportional hazard. *Health Economics* 13: 749–765.

Berlinguet, M., C. Preyra, and S. Dean. 2005. *Comparing the value of three main diagnosis based risk adjustment systems (DBRAS)*, Canadian Health Services Research Foundation (CHSRF) report.

Buntin, M.B., and A.M. Zaslavsky. 2004. Too much ado about two-part models and transformation? Comparing methods of modeling medicare expenditures. *Journal of Health Economics* 23: 525–542.

Centers for Medicare and Medicaid Services (CMS). 2006. *Announcement of calendar year (CY) 2007 medicare advantage capitation rates and medicare advantage and part D payment policies*. http://www.cms.hhs.gov/MedicareAdvtgSpecRateStats/Downloads/Announcement2007.pdf. Accessed 23 Aug 2006.

Constantinides, G. 1978. Market risk adjustment in project evaluation. *Journal of Finance* 33: 603–616.

Dixon, P., M. Dushieko, H. Gravelle, S. Martin, and N. Rice, et al. 2011. *Developing a person-based resource allocation formula for allocations to general practices in England*. Nuffield Trust. http://www.nuffieldtrust.org.uk/sites/files/nuffield/document/Developing_a_person-based_resource_allocation_formula_REPORT.pdf. Accessed 15 July 2011.

Duan, N., W. Manning, C.N. Morris, et al. 1983. A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics* 1: 115–126.

Duncan, I. 2011. *Healthcare risk adjustment and predictive modeling*. Winsted: Actex Publications.

Eggleston, K.N., R.P. Ellis, and M. Lu. (forthcoming). Risk adjustment and prevention. *Canadian Journal of Economics*.

Eichner, M.J. 1998. The demand for medical care: What people pay does matter. *The American Economic Review. Papers and Proceedings* 88(2): 117–21.

Ellis, R.P. 2008. Risk adjustment in health care markets: Concepts and applications. In *Paying for health care: New ideas for a changing society*, ed. M. Lu and E. Jonnson. Weinheim: Wiley-VCH.

Ellis, R.P., and T.G. McGuire. 2007. Predictability and predictiveness in health care spending. *Journal of Health Economics* 26: 25–48.

Ellis, R.P., G.C. Pope, L.I. Iezzoni, J.Z. Ayanian, D.W. Bates, H. Burstin, and A.S. Ash. 1996. Diagnosis-based risk adjustment for medicare capitation payments. *Health Care Financing Review* 12: 101–128.

Fetter, R.B., Y. Shin, J.L. Freeman, R.F. Averill, and J.-D. Thompson. 1980. Case mix definition by diagnosis-related groups. *Medical Care* 18(2), Supplement: Case mix definition by diagnosis-related groups. pp. i + iii + v + ix + 1–53. Stable URL: http://www.jstor.org/stable/3764138

Florida Agency for Health Care Administration. 2009. *Risk adjustment model comparison*. http://ahca.myflorida.com/Medicaid/quality_management/workgroups/managed_care/5_rar_model_comparison_050709.pdf. Accessed 17 Oct 2011.

Glazer, J., and T.G. McGuire. 2000. Optimal risk adjustment of health insurance premiums: An application to managed care. *American Economics Review* 90(4): 1055–1071.

Glazer, J., and T.G. McGuire. 2002. Setting health plan premiums to ensure efficient quality in health care: Minimum variance optimal risk adjustment. *Journal of Public Economics* 84(2): 53–173.

Gründl, H., T. Post, and R.N. Schulze. 2006. To hedge or not to hedge: Managing demographic risk in life insurance companies. *The Journal of Risk and Insurance* 73(1): 19–41.

Iezzoni, L.I. (ed.). 2003. *Risk adjustment for measuring healthcare outcomes*, 3rd ed. Ann Arbor: Health Administration Press.

Ingber, M. 1998. The current state of risk adjustment technology for capitation. *Journal of Ambulatory Care Management* 21: 1–28.

Jack, W. 2006. Optimal risk adjustment in a model with adverse selection and spatial competition. *Journal of Health Economics* 25: 908–926.

Kronick, R., and T. Dreyfus. 1997. *The challenge of risk adjustment for people with disabilities: Health-based payment for medicaid programs*. Princeton: Center for Health Care Strategies.

Kronick, R.T., T. Dreyfus, and Z. Zhou. 1996. Diagnostic risk adjustment for Medicaid: The disability payment system. *Health Care Financing Review* 17: 7–33.

Manning, W.G., and J. Mullahy. 2001. Estimating log models: To transform or not to transform? *Journal of Health Economics* 20: 461–494.

Manning, W.G., A. Basu, and J. Mullahy. 2005. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics* 24: 465–488.

R

Newhouse, J.P. 2002. *Pricing the priceless: A health care conundrum*. Cambridge, MA: MIT Press.
Newhouse, J.P., et al. 1989. Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review* 10(3): 41–54.
Pope, G.C., K.W. Adamache, R.K. Khandker, and E.G. Walsh. 1998. Evaluating alternative risk adjusters for medicare. *Health Care Financing Review* 20: 109–129.
Pope, G.C., J. Kautter, R.P. Ellis, A.S. Ash, J.Z. Ayanian, L.I. Iezzoni, M.J. Ingber, J.M. Levy, and J. Robst. 2004. Risk adjustment of medicare capitation payments using the CMS-HCC model. *Health Care Financing Review* 25(4): 119–141.
Rice, N., and P. Smith. 2001. Capitation and risk adjustment in health care financing: An international progress report. *Milbank Quarterly* 79: 81–113.
Robst, J., J.M. Levy, and M.J. Ingber. 2007. Diagnosis-based risk adjustment for medicare prescription drug plan payments. *Health Care Financing Review* 28(4): 15–30.
Shen, Y., and R.P. Ellis. 2002. Cost minimizing risk adjustment. *Journal of Health Economics* 21: 515–530.
Symmetry Health Data Systems, Inc. 2001. *Episode risk groups: ERG user's guide*. Phoenix: Symmetry Health Data Systems, Inc.
Thomas, J.W., K.L. Grazier, and K. Ward. 2004. Comparing accuracy of risk adjustment methodologies used in economic profiling of physicians. *Inquiry* 41: 218–231.
van de Ven, W.P.M.M., and R.P. Ellis. 2000. Risk adjustment in competitive health plan markets. In *Handbook of health Eeonomics*, ed. A.J. Culyer and J.P. Newhouse. North-Holland: Amsterdam. Available at http://www.sciencedirect.com/science/article/pii/S1574006400801730. Accessed 1 Aug 2011.
van de Ven, W.P.M.M., K. Beck, F. Buchner, D. Chernichovsky, and L. Gardiol, et al. 2003. Risk adjustment and risk selection on the sickness fund insurance market in five European countries. *Health Policy* 65: 75–98. Available at http://www.econ.kuleuven.be/public.economics/papers/riadinsu.pdf. Accessed 15 July 2011.
van de Ven, W.P.M.M., R.C.J.A. van Vliet, and L.M. Lamers. 2004. Health-adjusted premium subsidies in the Netherlands. *Health Affairs* 23: 45–54.
Von Korff, M., E.H. Wagner, and K. Saunders. 1992. A chronic disease score from automated pharmacy data. *Journal of Clinical Epidemiology* 45: 197–203.
Wasem, J., L.M. Lauterbach, and W.F. Schrader. 2006. *Klassifikationsmodelle für Versicherte im morbiditätsorientierten Risikostrukturausgleich*. (Classification models for risk adjustment in the morbidity-oriented risk structure reconciliation.) Wissenschaftliches Institut der AOK (WIdO). Available at http://wido.de/fileadmin/wido/downloads/pdf_ggw/wido_ggw_aufs1_0205.pdf. Accessed 17 Oct 2011.
Weiner, J.P., B.H. Starfield, D.M. Steinwachs, and L.M. Mumford. 1991. Development and application of a population-oriented measure of ambulatory care case mix. *Medical Care* 29: 453–472.
Winkelman, R., and S. Mehmud. 2007. *A comparative analysis of claims-based tools for health risk assessment*. Schaumburg: Society of Actuaries.

Risk Aversion

Jan Werner

Abstract

An agent, perhaps an individual or a firm, is said to be *risk averse* if the agent prefers a deterministic outcome equal to the expectation of a risky outcome over that risky outcome. Risk aversion seems to be a common characteristic; introspection suggests as much. More importantly, it gives qualitative explanation to economic behaviour in many instances where risk is present. If individuals and firms were not risk averse, insurance markets would not exist. Needless to say, there are activities which are inconsistent with agents being risk averse. Gambling is perhaps the best example of such an activity.

Keywords

Arrow–Pratt theory of risk aversion; Arrow–Pratt measure of absolute risk aversion; Arrow–Pratt measure of relative risk aversion; Certainty equivalent; Concavity; Dual utility; Equity premium puzzle; Expected utility theory; Insurance markets; Maxmin expected utility; Multivariate risk; Probability measures; Random variables; Rank-dependent expected utility; Relative risk; Representation of preferences; Risk aversion; Risk compensation; Risk–return trade-off; Schur concave functions; State-dependent utility; Stochastic dominance; von Neumann–Morgenstern utility function; Wald's criterion

JEL Classification

D81; G12; D81

Arrow–Pratt Theory of Risk Aversion

The classical theory of risk aversion, due to Pratt (1964) and Arrow (1965), is rooted in the expected utility theory of decision making. An agent's preferences are assumed to have an expected utility representation. The objects of choice are real valued random variables defined either on a finite or infinite set of states of the world with probabilities of states that may be either objective or subjective. The intended interpretation of a random variable is as an agent's risky wealth.

An agent whose expected utility representation of preferences is written $E[u(\tilde{x})]$, where u is the von Neumann–Morgenstern utility function and E denotes the expectation (expected value), is *risk averse* if

$$E[u(\tilde{x})] \leq u(E(\tilde{x})) \quad (1)$$

for every risky wealth $\tilde{x}$. If (1) holds with strict inequality for every non-deterministic $\tilde{x}$, the agent is *strictly risk averse*. Jensen's inequality implies that, if utility function u is concave, the agent is risk averse. The converse is also true. Thus, the concavity of u is a necessary and sufficient condition for risk aversion. Moreover, strict concavity of u is a necessary and sufficient condition for strict risk aversion. Examples of strictly concave von Neumann–Morgenstern utility functions, commonly used in applied work, include the negative exponential utility $u(w) = e^{-\alpha w}$ with $\alpha > 0$, the logarithmic utility $u(w) = \ln(w)$, and the power utility $u(w) = \frac{1}{1-\alpha} w^{1-\alpha}$ with $\alpha > 0, \alpha \neq 1$.

It is useful to have a measure of the intensity of risk aversion. The most natural measure is *risk compensation*. It is by definition the amount $\rho(w\tilde{z})$ of deterministic wealth one could extract from an agent in exchange for relieving her of zero-expectation risk $\tilde{z}$ at an initial deterministic wealth w,

$$E[u(w + \tilde{z})] = u(w - \rho(w, \tilde{z})). \quad (2)$$

A risk-averse agent has non-negative risk compensation for every zero-expectation risk, at every level of initial wealth. Risk compensation makes possible interpersonal comparisons of risk aversion and, for any agent, comparisons of risk aversion at different levels of her initial wealth. If risk compensation $\rho_1(w, \tilde{z})$ of an agent with von Neumann–Morgenstern utility function u_1 is greater than or equal to risk compensation $\rho_2(w, \tilde{z})$ of another agent with utility function u_2, for every deterministic wealth w and risk $\tilde{z}$ with $E(\tilde{z}) = 0$, then the agent with u_1 is said to be *more risk averse* than the one with u_2. An agent has increasing, decreasing or constant risk aversion if, for every zero-expectation risk, her risk compensation is increasing, decreasing or constant in w, for every $\tilde{z}$ with $E(\tilde{z}) = 0$.

Another measure of risk aversion is *certainty equivalent*. It is by definition the amount $c(\tilde{x})$ of deterministic wealth such that an agent is indifferent between this deterministic wealth and risky wealth $\tilde{x}$,

$$E[u(\tilde{x})] = u(c(\tilde{x})). \quad (3)$$

For a risk-averse agent, certainty equivalent is lower than the expectation of risky wealth. Since certainty equivalent and risk compensation are related by $c(w + \tilde{x}) = w - \rho(w, \tilde{z})$, these two measures can be interchangeably used in the Arrow–Pratt theory of risk aversion.

Although with considerable intuitive appeal, risk compensation is not all that practical. It is only implicitly defined in (2). The basic insight of Arrow and Pratt is that there is a simple measure of risk aversion which is in a certain sense equivalent to risk compensation, namely, the *Arrow–Pratt measure of absolute risk aversion*

$$A(w) = -\frac{u''(w)}{u'(w)}. \quad (4)$$

The negative of the second derivative $u''(w)$ is a mathematical measure of the degree of concavity of u. It is rescaled by the first derivative (assumed non-zero) which makes the measure invariant under any affine transformation of u. For 'small' risk $\tilde{z}$ with $E(\tilde{z}) = 0$, risk compensation $\rho(w, \tilde{z})$ equals approximately half the product of the variance $\sigma^2(\tilde{z})$ and the Arrow–Pratt measure at w,

R

$$\rho(w, \tilde{z}) \cong \frac{1}{2} A(w) \sigma^2(\tilde{z}) \tag{5}$$

as can be demonstrated using quadratic approximation of expected utility $E[u(w + \tilde{z})]$.

The important theorem of Pratt establishes an equivalence of the two measures as criteria for comparative risk aversion. For that theorem, let u_1 and u_2 be two strictly increasing von Neumann–Morgenstern utility functions, twice differentiable with continuous second derivatives.

Theorem (Pratt 1964) The following conditions are equivalent:

(a) $A_1(w) \geq A_2(w)$ *for every w.*
(b) $\rho_1(w, \tilde{z}) \geq \rho_2(w, \tilde{z})$ *for every w and every random variable* ~ *z with* $E(\tilde{z}) = 0$,
(c) u_1 *is a concave transformation of* u_2*; that is,* $u_1 = f \circ u_2$ *for f concave and strictly increasing*.

There is a version of the Pratt's theorem which has equalities in (*a*) and (*b*) and an affine transformation in (*c*). This version implies that the Arrow–Pratt measure identifies the von Neumann–Morgenstern utility function up to an affine transformation. For example, the negative exponential utility is (up to an affine transformation) the only strictly concave utility with constant absolute risk aversion. There is also a strict version of the Pratt's theorem with strict inequalities in (*a*) and and a strictly concave transformation in (*c*).

A corollary to the Pratt's Theorem says that an agent has increasing, decreasing or constant risk aversion if and only if her Arrow–Pratt measure of risk aversion is increasing, decreasing or constant. One needs only to consider utility functions $u_1(w) = u(w)$ and $u_2(w) = u(w + \Delta w)$ for arbitrary $\Delta w > 0$.

Arrow (1965) and Pratt (1964) extended their theory of measurement of risk aversion to relative risk, that is, risk per dollar of an agent's wealth. Risk compensation $\rho_r(w, \widetilde{\zeta})$ for relative risk $\widetilde{\zeta}$ with $E(\widetilde{\zeta})$ 0 at initial wealth w is defined by

$$E\left[u\left(w + w\widetilde{\zeta}\right)\right] = u\left(w - w\rho_r\left(w, \widetilde{\zeta}\right)\right). \tag{6}$$

The Arrow–Pratt measure of relative risk aversion is $R(w) = -\frac{u''(w)}{u'(w)} w$. The measures ρ_r and R are related in the same way that their counterparts for absolute risk are related. Power and logarithmic utility functions have constant relative risk aversion.

Risk compensation is defined when the agent's initial position is risk-free. The approximation (5) of risk compensation by the Arrow–Pratt measure holds at a risk-free position, too. Measures of risk aversion that can be used when the initial position is risky have been developed by Ross (1981) and Machina and Neilson (1987).

When random variables are vector-valued (multivariate), the Arrow–Pratt theory can be applied to risk in one coordinate (for example, consumption of one good) when values of other coordinates are deterministic. Alternatively, multivariate risk aversion can be defined by requiring condition (1) to hold for every multivariate random variable $\tilde{x}$. Multivariate random variables arise when objects of choice are consumption plans of multiple goods or consumption plans over multiple time periods. Multivariate risk aversion is equivalent to concavity of the von Neumann–Morgenstern utility function and implies that the induced ordinal preferences over multiple goods under certainty are convex. The theory of comparative risk aversion has been extended to the multivariate case by Kihlstrom and Mirman (1974) under the restriction that utility functions induce the same ordinal preferences.

Rabin and Thaler (2001) have pointed out a peculiar feature of risk aversion under expected utility. If an agent rejects a small actuarially favourable gamble at every level of wealth (or at a big enough range of wealth), then she will reject a gamble with a modest loss and an arbitrarily large gain. They presented a calibration exercise which shows that any risk-averse agent who rejects an even-chance gamble of losing \$10 or winning \$11 will turn down an even-chance gamble of losing \$1000 or winning any sum of money. The significance of Rabin and Thaler's observation is a subject of current debate.

Risk Aversion Without Expected Utility

A representation of preferences which is closely related to but more general than the expected utility is the state-dependent expected utility. For a finite set S of states of the world, it is written $\sum_{s\in S} \pi_s u_s(x_s)$, where π_s is the probability of state s and u_s is the state-dependent utility. Werner (2005a) has shown that an agent with state-dependent expected utility is risk averse in the sense of having a preference for deterministic outcomes over risky outcomes with equal expectations if and only if the utility functions u_s are state independent and concave. When utility functions u_s are state dependent, a risk-free wealth may have risky utility and, more importantly, risky marginal utility. Karni (1985) has developed a theory of aversion to risk in marginal utility defined by an agent being unwilling to take an actuarially fair gamble when starting from a position of risk-free marginal utility of wealth. Measures of risk aversion analogous to the measures introduced by Arrow and Pratt can be defined, and an extension of the Pratt's theorem obtains for utility functions that have the same set of wealth profiles with risk-free marginal utility. State-dependent utilities arise in instances of behaviour under health risk.

The Arrow–Pratt theory of risk aversion is based on the simple notion that every risky outcome is more risky than the deterministic outcome with equal expectation. For preferences that do not have an expected utility representation (state-independent, or not), this concept of 'more risky' is too restrictive to deliver a meaningful notion of risk aversion. A weaker concept of 'more risky' has been introduced by Rothschild and Stiglitz (1970). It is a partial ordering of random variables according to the integrals of their cumulative distribution functions. For two random variables $\tilde{x}$ and $\tilde{y}$ with the same expectation, $\tilde{x}$ is *more risky* than $\tilde{y}$ if

$$\int_{-\infty}^{w} F_y(y)dt \leq \int_{-\infty}^{w} F_x(t)dt, \quad \forall w, \tag{7}$$

where F_x is the cumulative distribution function assigning to each w the probability $Prob\,\{\tilde{x} \leq w\}$.

An agent whose utility function on random variables is monotone decreasing with respect to the ordering of more risky is said to be *strongly risk averse* (see Cohen 1995). It follows that a strongly risk-averse agent, when starting from a risky position $\tilde{x}$, is unwilling to take a gamble $\tilde{z}$ with zero expectation conditional on each possible realization x of $\tilde{x}$, that is, $E(\tilde{z}|\tilde{x}=x) = 0$ for every x. The ordering (7) has been known in mathematics as the second-order stochastic dominance and the strongly risk-averse functions have been known as the Schur concave functions (see Marshall and Olkin 1979). Chew et al. (1987) derived necessary and sufficient conditions for strong risk aversion of two types of utility functions: the rank-dependent expected utility of Quiggin (1982), and the dual utility of Yaari (1987). Characterization results for general utility functions can be found in Machina (1982), Chew and Mao (1995), and Dana (2005). For an expected utility, strong risk aversion and risk aversion in the sense of (1) are equivalent. Concavity of the von Neumann–Morgenstern utility function is necessary and sufficient for strong risk aversion of expected utility (see Rothschild and Stiglitz 1970).

An important representation of preferences under uncertainty, more general than the expected utility, is the maxmin (or multiple-prior) expected utility (see Gilboa and Schmeidler 1989). Under the maxmin expected utility representation an agent has a set || of probability measures (priors) on states instead of a single probability measure, and a von Neumann–Morgenstern utility function u. The set || is assumed to be closed and convex. An agent's utility of risky wealth $\tilde{x}$ is

$$\min_{p\in \mathrm{P}} E_p[u(\tilde{x})], \tag{8}$$

where $E_P[u(\tilde{x})]$ is the expectation of $u(\tilde{x})$ with respect to probability measure P. Multiplicity of probability measures reflects the agent's ambiguous information about states of the world. Taking the minimum reflects the concern with the 'worst case' scenario. If the set || consists of all probability measures, then the maxmin expected utility (8) equals $\min_s u(x_s)$ meaning that the agent follows the Wald's criterion of choice. Maxmin expected utility may exhibit risk aversion with respect to

R

some probability measure in the set of priors. If the von Neumann–Morgenstern utility function u is concave, the agent prefers deterministic wealth in the amount of $E_P(\tilde{x})$ over risky wealth $\tilde{x}$ for every probability measure P in her set of priors. Thus the agent is risk averse in the Arrow–Pratt sense with respect to every P in the set ||. Wald's minimum utility is also strongly risk averse with respect to every probability measure. Many maxmin expected utility functions are not distribution invariant under any probability measure on states, rendering the question of strong risk aversion meaningless for these functions. Werner (2005b) proposes a concept of more risky, stronger than (7), such that adding a gamble with zero conditional expectation makes an initial risky position more risky, but without identifying random variables with their probability distributions. For many (but not all) sets of priors, maxmin expected utility with concave von Neumann–Morgenstern utility function is risk averse in that sense under some probability measure from the set of priors.

Some Implications of Risk Aversion

The choice of insurance coverage provides a good illustration of implications of risk aversion on agents' behaviour. Suppose that an expected-utility maximizing individual with initial wealth w faces a risk of losing L with probability π or not losing it with probability $1-\pi$. She is offered insurance against the loss at price p per dollar of coverage. A strictly risk-averse individual will choose full coverage giving her risk-free wealth $w - pL$, if the insurance is priced actuarially fair, that is $p = \pi$. If it is priced above the fair value, that is $p > \pi$, then the optimal coverage will be partial. Schlesinger (1997) shows that these results continue to hold under risk aversion without expected utility.

A risk-averse investor who invests her initial risk-free wealth among many risky assets and a risk-free asset will choose an optimal portfolio with risky payoff only if the expected return on that portfolio exceeds the risk-free return. For a strictly risk-averse investor, the expected return on the optimal portfolio must strictly exceed the risk-free return if the payoff is risky. This is the fundamental *risk–return trade-off* in asset markets and it is a consequence of risk aversion. It continues to hold when the investor's initial position includes an endowment portfolio of assets. In an equilibrium in competitive asset markets where many strictly risk-averse investors trade their endowment portfolios, the market portfolio (that is, the outstanding supply of assets) must have expected return that exceeds the risk-free return. This is so because the return on the market portfolio is a weighted sum of the returns on investors' optimal portfolios, with weights equal to the respective shares of total wealth. Expected returns on optimal portfolios exceed the risk-free return, with some exceeding it strictly, if the payoff of the market portfolio is risky. Thus, risk aversion provides a qualitative explanation of the expected return in equity markets exceeding the risk-free return. Attempts to give a quantitative explanation of the observed high excess return on equities over risk-free bonds have led to the *equity premium puzzle* (see Mehra and Prescott 1985).

See Also

- ▶ Expected Utility Hypothesis
- ▶ Non-expected Utility Theory
- ▶ Risk
- ▶ Stochastic Dominance

Bibliography

Arrow, K. 1965. *Aspect of the theory of risk bearing*. Helsinki: Yrjö Jahnsson Foundation.

Chew, S., and M. Mao. 1995. A Schur concave characterization of risk aversion for non-expected utility preferences. *Journal of Economic Theory* 67: 402–435.

Chew, S., E. Karni, and Z. Safra. 1987. Risk aversion in the theory of expected utility with rank dependent preferences. *Journal of Economic Theory* 42: 370–381.

Cohen, M. 1995. Risk-aversion concepts in expected-and non-expected-utility models. *Geneva Papers on Risk and Insurance Theory* 20: 73–91.

Dana, R.-A. 2005. A representation result for concave Schur-concave functions. *Mathematical Finance* 15: 613–634.

Gilboa, I., and D. Schmeidler. 1989. Maxmin expected utility with nonunique prior. *Journal of Mathematical Economics* 18: 141–153.
Karni, E. 1985. *Decision making under uncertainty*. Cambridge, MA: Harvard University Press.
Kihlstrom, R., and L. Mirman. 1974. Risk aversion with many commodities. *Journal of Economic Theory* 8: 361–388.
Machina, M. 1982. Expected utility analysis without the independence axiom. *Econometrica* 50: 277–323.
Machina, M., and W. Neilson. 1987. The Ross characterization of risk aversion: Strengthening and extension. *Econometrica* 55: 1139–1149.
Marshall, A., and I. Olkin. 1979. *Inequalities: Theory of majorization and its applications*. New York: Academic.
Mehra, R., and E. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15: 145–161.
Pratt, J. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 132–136.
Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3: 323–343.
Rabin, M., and R. Thaler. 2001. Anomalies: Risk aversion. *Journal of Economic Perspectives* 15(1): 219–232.
Ross, S. 1981. Some stronger measures of risk aversion in the small and in the large with applications. *Econometrica* 49: 621–638.
Rothschild, M., and J. Stiglitz. 1970. Increasing risk. I: A definition. *Journal of Economic Theory* 2: 225–243.
Schlesinger, H. 1997. Insurance demand without the expected-utility paradigm. *Journal of Risk and Insurance* 64(1): 19–39.
Werner, J. 2005a. A simple axiomatization of risk-averse expected utility. *Economics Letters* 88: 73–77.
Werner, J. 2005b. Risk and risk aversion when states of nature matter. Mimeo.
Yaari, M. 1987. The dual theory of choice under risk. *Econometrica* 55: 95–115.

Risk Sharing

Ethan Ligon

Abstract

Agents increase their expected utility by using state-contingent transfers to share risk; many institutions seem to play an important role in permitting such transfers. If agents are suitably risk-averse, then in the absence of any frictions the benchmark Arrow–Debreu model predicts that *all* risk will be shared, so that idiosyncratic shocks will have no effect on individuals; we call this full risk sharing. Real-world tests of full risk sharing tend to reject it; accordingly, researchers have devised models incorporating various frictions to try to explain the partial risk sharing evident in the data.

Keywords

Arrow–Debreu economy; Commitment; Credit; Euler equations; Financial markets; Full risk sharing; Idiosyncratic risk; Insurance; Lagrange multipliers; Pareto efficiency; Partial risk sharing; Permanent-income hypothesis; Risk aversion; Risk sharing; Separability; Sharecropping; State-contingent transfers; von Neumann–Morgenstern preferences; von Neumann–Morgenstern utility functions

JEL Classifications

O1

Any two agents may be said to share risk if they employ state-contingent transfers to increase the expected utility of both by reducing the risk of at least one. A very wide variety of human institutions seem to play an important role in risk sharing, including insurance, credit, financial markets, and sharecropping in developing countries.

To be precise, consider a set of agents indexed by $i = 1,\ldots, n$ each with von Neumann–Morgenstern utility function U_i and a finite set of possible states of the world $s = 1,\ldots, S$, each of which occurs with probability $p(s)$. For simplicity, suppose that each agent i receives a quantity of a single consumption good $x_i(s)$ in state s, thus receiving expected utility

$$EU_i(x_i) = \sum_{s=1}^{S} p(s)U_i(x_i(s)),$$

where x_i denotes the random variable, $\{x_i(s)\}$ denotes its realizations, and E is the expectation operator. We assume that U_i is strictly increasing, weakly concave, and continuously differentiable for all $i = 1,\ldots, n$, so that all agents are at least weakly risk averse. Define the risk faced by agent i to be a quantity

R

$$R_i(x_i) = U_i(Ex_i) - EU_i(x_i).$$

This cardinal measure orders probability distributions in the same manner as Rothschild and Stiglitz (1970). We say that i faces *idiosyncratic risk* if $R_i(x_i) > 0$ and $\mathrm{corr}\left(U_i'(x_i), U_j'(x_j)\right) < 1$ for some j, where U_j' denotes j's marginal utility. If any agent i bears idiosyncratic risk, then there exists a set of state-contingent transfers of the consumption good between i and j, $\left\{\tau_i^j(s)\right\}$ which will strictly increase the expected utility of each, while strictly decreasing the risk of at least one of i and j. Implementing such transfers is risk sharing.

Full Risk Sharing

What might be termed *full risk sharing* (Allen and Gale 1988; Rosenzweig 1988) is a situation in which all idiosyncratic risk is eliminated. While agents may still face risk, this risk is shared, so that marginal utilities of consumption are perfectly correlated across all agents. Full risk sharing is a hallmark of any Pareto-efficient allocation in an Arrow–Debreu economy, provided that agents have von Neumann–Morgenstern preferences, no one is risk-seeking, and at least one agent is strictly risk averse.

Let us establish the necessity of full risk sharing for any interior Pareto-efficient allocation in a simple multi-period endowment economy. The environment is similar to that described above, but agents consume in several periods indexed by $t = 1, \ldots, T$, with agent i discounting future expected utility using a discount factor β_i. Different states of the world are realized in each period, with the probability of state $s_t \in \{1, \ldots, S\}$ being realized in period t allowed to depend on the period, and so given by $p_t(s_t)$. Then consider the problem facing a social planner, who assigns state-contingent consumption allocations to solve

$$\max_{\{(c_{it}(s))\}} \sum_{i=1}^{n} \lambda_i \sum_{t=1}^{T} \beta_i^{t-1} \sum_{S_t=1}^{S} p_t(s_t) U_i(c_{it}(s_t))$$

subject to the resource constraints

$$\sum_{i=1}^{n} c_{it}(s) \leq \sum_{i=1}^{n} x_{it}(s_t),$$

which must be satisfied at every period t and state s_t; the planner takes as given the initial state s_0 and a set of positive weights $\{\lambda_i\}$. By varying these weights one can compute the entire set of interior Pareto-efficient allocations (Townsend 1987).

If we let $\mu_t(s_t)$ denote the Lagrange multiplier associated with the resource constraint for period t in state s_t, then the first order conditions for the social planner's problem are

$$\lambda_i \beta_i^{t-1} p_t(s_t) U_i'(c_{it}(s_t)) = \mu_t(s_t). \qquad (1)$$

Since this condition must be satisfied in all periods and states for every agent, it follows that

$$U_i'(c_{it}(s_t)) = \frac{\lambda_j}{\lambda_i}\left(\frac{\beta_j}{\beta_i}\right)^{t-1} U_j'(c_{it}(s_t))$$

for any period t, any pair of agents (i, j) and any state s_t, so that $\mathrm{corr}\left(U_i'(c_{it}), U_j'(c_{jt})\right) = 1$, and we have full risk sharing.

Thus far, we've considered risk sharing only in the context of an endowment economy. However, the thrust of the claims advanced above holds much more generally. If we were, for example, to add production and some kind of intertemporal technology (for example, storage), the first order conditions of the planner's problem with respect to state-contingent consumptions (1) would remain unchanged – the effect of these changes would be that the Lagrange multipliers $\{\mu_t(s_t)\}$ would change. This is an illustration of what is sometimes called 'separability' between production and consumption, which typically prevails only when there is full risk sharing (see, for example, Benjamin 1992).

Risk sharing can also be thought of as a means to smooth consumption across possible states of the world. This suggests a connection to the permanent-income hypothesis, which at its core involves agents smoothing consumption across periods. And indeed, it's easy to show that full risk sharing in every period implies the kind of smoothing across periods implied by the consumption Euler equation. However, the consumption Euler equation doesn't imply full risk sharing.

Tests of Full Risk Sharing

The insight that Pareto-efficient allocation among risk-averse agents implies full risk sharing has led to tests of versions of (1). The usual strategy involves adopting a convenient parameterization of U_i, and then calculating the logarithm of both sides of (1). For example, if $U_i(c) = \frac{c^{1-\gamma}}{1-\gamma}$, with $\gamma > 0$, then this yields the relationship

$$\gamma \log c_{it}(s_t) = \log \frac{\mu_t(s_t)}{p_t(s_t)} - \log \frac{\lambda_i}{\beta_i} - t \log \beta_i. \quad (2)$$

This is a simple consumption function, which we would expect to be consistent with any efficient allocation. The quantity $\frac{\mu_t(s_t)}{p_t(s_t)}$ is related to the aggregate supply of the consumption good. Note that this is the only determinant of consumption which depends on the random state. This reflects the fact that the only risk borne by agents in an efficient allocation will be aggregate risk. The second term varies with neither the state nor the date, and can be thought of as depending on the levels of consumption that agent i can expect (in a decentralization of this endowment economy, λ_i could be interpreted as a measure of i's time zero wealth). The final term has to do with differences in agents' patience.

Now, suppose one has panel data on realized consumption for a sample of agents over some period of time. If we let $\tilde{c}_{it}$ denote observed consumption for agent i in period t, (2) implies the estimating equation

$$\log \tilde{c}_{it} = \eta_t + \alpha_i + \delta_i t + \varepsilon_{it}, \quad (3)$$

where $\eta_t = \log \frac{\mu_t(s_t)}{\gamma p_t(s_t)}, \alpha_i = -\log \frac{\lambda_i}{\lambda \beta_i}$, $\delta_i = -\log \beta_i$, and ε_{it} is some disturbance term. Because this final disturbance term isn't implied by the model, it's typically motivated by assuming that it's related either to measurement error in consumption or to some preference shock.

The reduced form consumption Eq. (3) can be straightforwardly estimated by using ordinary least squares, but this doesn't constitute a test of full risk sharing. To construct such a test, one typically uses data on other time-varying, idiosyncratic variables which would plausibly influence consumption under some alternative model which predicts less than full risk sharing. Perhaps the most obvious candidate for such a variable is some measure of income, for example the observed endowment realizations $\tilde{x}_{it}$ referred to in the model above. Then one can add (the logarithm of) this variable to reduced form as an additional regressor, yielding an estimating equation of the form

$$\log \tilde{c}_{it} = \eta_t + \alpha_i + \delta_i t + \varphi \log \tilde{x}_{it} + \varepsilon_{it} \quad (4)$$

(Mace 1991; Cochrane 1991; Deaton 1992; Townsend 1994). Then full risk sharing and an auxiliary assumption that ε_{it} is mean independent of the regressors implies the exclusion restriction $\varphi = 0$, which can be easily tested.

Partial Risk Sharing

Restrictions along the lines of (4) have been used to test for full risk sharing in a wide variety of settings, including within-dynasty risk sharing (Hayashi et al. 1996) in the United States, risk sharing across countries (Obstfeld 1994), risk sharing within networks in the Philippines (Fafchamps and Lund 2003), and risk sharing across households in India (Townsend 1994), Africa, or the United States (Mace 1991). A typical finding is that the estimated response of consumption to income shocks is small but significant, leading one to reject the null hypothesis of full risk sharing. In this case it is tempting to interpret the estimated relationship as determining the response of consumption to income. However, this is generally a mistake. By rejecting the hypothesis of full risk sharing one also rejects the model which generated the hypothesis, so that theory no longer supports the interpretation of (4) as a consumption function.

Given this kind of evidence against full risk sharing, scholars have been led to devise and test alternative models in which some kind of friction leads to agents bearing some idiosyncratic income risk. Two promising frictions are private information and limited commitment. In the case of private information, realized or announced incomes may provide a useful signal regarding hidden actions or information, and thus an agent's consumption will optimally depend on this signal, leading to a balance between risk sharing and incentives (Holmström 1979); Ligon (1998) tests this weaker risk-sharing hypothesis in three Indian villages, and is unable to reject it. In the case of limited commitment, an agent who receives an unusually large endowment realization may be tempted to renege on a pre-existing risk-sharing arrangement unless she receives a larger share of resources (Kocherlakota 1996); a test of this model in the same three Indian villages by Ligon et al. (2002) finds that this model predicts a response of consumption to income of just the right magnitude. Still, the construction, estimation, and testing of well-specified models which predict only partial risk sharing remains in its infancy.

See Also

- ► Euler Equations
- ► Permanent-Income Hypothesis

Bibliography

Allen, F., and D. Gale. 1988. Optimal security design. *Review of Financial Studies* 1: 229–263.

Benjamin, D. 1992. Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica* 1: 287–322.

Cochrane, J. 1991. A simple test of consumption insurance. *Journal of Political Economy* 99: 957–976.

Deaton, A. 1992. *Understanding consumption*. Oxford: Clarendon Press.

Fafchamps, M., and S. Lund. 2003. Risk-sharing networks in rural Philippines. *Journal of Development Economics* 71: 261–287.

Hayashi, F., J. Altonji, and L. Kotlikoff. 1996. Risk-sharing between and within families. *Econometrica* 64: 261–294.

Holmström, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.

Kocherlakota, N. 1996. Implications of efficient risk sharing without commitment. *Review of Economic Studies* 63: 595–610.

Ligon, E. 1998. Risk-sharing and information in village economies. *Review of Economic Studies* 65: 847–864.

Ligon, E., J. Thomas, and T. Worrall. 2002. Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *Review of Economic Studies* 69: 209–244.

Mace, B. 1991. Full insurance in the presence of aggregate uncertainty. *Journal of Political Economy* 99: 928–956.

Obstfeld, M. 1994. Risk-taking, global diversification, and growth. *American Economic Review* 84: 1310–1329.

Rosenzweig, M. 1988. Risk, implicit contracts and the family in rural areas of low-income countries. *Economic Journal* 98: 1148–1170.

Rothschild, M., and J. Stiglitz. 1970. Increasing risk: I. A definition. *Journal of Economic Theory* 2: 225–243.

Townsend, R. 1987. Microfoundations of macroeconomics. In *Advances in economic theory: Fifth World Congress*, ed. T. Bewley. Cambridge: Cambridge University Press.

Townsend, R. 1994. Risk and insurance in village India. *Econometrica* 62: 539–591.

Risk-Coping Strategies

Stefan Dercon

Abstract

Risk is part of life in developing countries. Despite generally imperfect credit and lacking insurance markets, households use a variety of strategies to manage and cope with risk, including savings and informal credit transactions, mutual support networks, and income and asset diversification. Most evidence suggests that these strategies achieve only partial consumption smoothing and risk-sharing, while they are not without long-term costs in terms of investment and poverty. This article discusses the nature and evidence on the typical strategies used, and explores its implications. It also highlights some outstanding questions.

Keywords

Asymmetric information; Buffer stocks; Credit markets in developing countries; Health insurance; Household portfolios; Income diversification; Income shocks; Income smoothing; Insurance markets; Intertemporal consumption smoothing; Labour supply; Micro-insurance; Optimal portfolio mix; Precautionary savings; Risk; Risk aversion; Risk coping strategies; Risk diversification; Risk sharing; Risk-sharing contracts; Risk-sharing networks; Self-insurance

JEL Classification

O1

Risk is pervasive, not least in developing countries. A high dependence on agriculture for income and employment means that people's livelihoods are strongly affected by climatic vagaries. Poor health care and immunization means that illness is common, affecting labour supply. Poor infrastructure and market institutions result in limited market integration, leading to a high sensitivity of prices to local shocks. Economic instability, conflict and political instability add further to a high-risk environment in many developing countries.

While risk is widespread, in developing countries insurance markets are typically missing or incomplete. This causes potentially serious welfare losses, especially since government-led alternatives such as social security are largely missing, compounded by imperfections in credit markets that limit the extent to which credit can be used as to substitute for insurance. Economic agents are typically risk averse and, even in the poorest settings, the evidence suggests that they do not just passively undergo the consequences of risk. Instead, given risk aversion, they try to make activity and asset portfolio choices to balance their need to make a living, but without exposing themselves to too much risk. The strategies used to achieve this are often referred to as 'risk management and coping strategies'. The economic analysis of these strategies has been one of the areas in which research based on some of the poorest high-risk rural settings in developing countries has made a substantial contribution to the economic literature in general. Our focus in this overview is on the empirical literature, often drawing on the evidence generated from the sample based on six Indian villages for which the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) collected exceptionally detailed panel data over 10 years in the high-risk environment of the semi-arid tropics of India (Morduch 2004).

Self-Insurance via Savings

It is possible to distinguish a number of commonly observed different risk strategies. First, we can consider strategies that aim to cope with the consequences of shocks. Risk aversion is sufficient to induce households to try to smooth consumption or nutrition, and indeed standard models of consumption smoothing when insurance and credit markets are imperfect can shed light on this type of strategy (Deaton 1992). When shocks occur, households may decide to curb their consumption loss through the sacrifice of existing assets. Households may pre-empt this by trying to self-insure against risk through precautionary savings. A precautionary motive for savings would be sufficient for savings to increase in response to increased risk, so that a buffer stock is built to deplete when shocks occur. Even though formal credit markets in high-risk settings are typically underdeveloped, informal credit transactions may also be used for smoothing purposes.

There is a large literature testing whether households in developing countries smooth consumption, building on standard models of permanent income and often using shocks to identify the relevant effects (for example, Paxson 1992). Furthermore, which assets tend to be used for this purpose in the face of shocks has occasionally been investigated. (Finding positive evidence of using assets to smooth is not sufficient to show that any savings were built up for precautionary motives to start with. This would require evidence that greater risk indeed increased savings, which is harder to test.) Nevertheless, the evidence suggests that in some settings productive assets are sold off for smoothing, for example in work by Rosenzweig and Wolpin (1993) using the ICRISAT data, while in other settings (for example, in Burkina Faso, as in Fafchamps et al. 1998) livestock were not sold off despite serious income shocks. Furthermore, there is much anecdotal evidence that, during famines in mixed farming environments, livestock is not being sold despite serious human nutritional losses and risks.

Squaring these findings with basic theoretical models remains a challenge, and different suggestions can be made. For example, different technologies may underlie the pattern of returns to different assets, so that optimal portfolio mixes would suggest that assets are depleted at different rates when shocks occur. Another possibility is that asset returns and prices are risky as well, so that the reliability of the buffer is limited. For example, if incomes, asset prices and returns have a high positive covariance, then selling assets when incomes are low may not be the optimal strategy, since the future gains from holding on to assets is actually high. Alternatively, behavioural theories based on experiments, such as that risk-loving rather than risk-averse behaviour may be displayed in the face of losses that are potentially very large (as in Kahneman and Tversky 1979), may provide some insight to these puzzles.

A related strategy to self-insurance observed in poor settings involves the key asset available to the poor, namely, labour. In response to shocks, labour supply is adjusted to increase involvement in productive activities, including off-farm or temporary migration (for example, Kochar 1995). Furthermore, children may be taken out of school and into work in response to income shocks (Jacoby and Skoufias 1997).

Risk-Sharing Through Mutual Support

A second common strategy to cope with the consequences of risk involves non-market institutions based on risk sharing, whereby households or other economic agents use transfers to smooth outcomes across a group of people when shocks occur. Conceptually, this is the cross-sectional equivalent of standard permanent income models: it is concerned with smoothing over space rather than over time. Unless risk preferences differ between economic agents, unlike self-insurance strategies this strategy is relevant only for idiosyncratic shocks, not covariate risk. Townsend (1994) is the seminal paper on high-risk developing-country environments. The basic prediction of Townsend's model, under specific assumptions, is that household consumption is dependent on average village resources but not on individual income realizations. This provides a clear basis for empirical testing of the perfect risk-sharing hypothesis: do idiosyncratic shocks to income affect consumption or nutrition outcomes within a well-defined setting? Using the ICRISAT data from India, he finds that perfect risk-sharing is rejected within the village, even though substantial risk-pooling takes place. Other studies (including some on the same data) suggest that some risk-sharing typically occurs in villages, but the evidence is typically not consistent with perfect risk-sharing.

The failure of perfect risk-sharing in village settings has attracted much attention in terms of theoretical work. Work has focused on accounting for information asymmetries, private savings and the role of enforceability problems related to these informal risk-sharing contracts *ex post* in repeated game contexts (for example, Ligon et al. 2002). The work on enforceability has found that constrained efficient contracts will contain updating rules offering higher weights in the risk-sharing contracts in particular states of the world to those facing stronger incentives to leave, and that over time, the weights have memory. The consequence is that these risk-sharing contracts take on features more common in credit contracts, leading to their description as 'quasi-credit' arrangements. There is some empirical evidence consistent with these models.

The lack of perfect risk-sharing within villages has led to further work investigating whether risk-sharing occurs in other settings, for example within households or extended families, or in other social groupings. Partial risk-sharing has been documented across ethnic groups as well as within families. More recently, risk-sharing across networks has been explored as well. It is relatively straightforward to analyse risk-sharing within networks beyond specifically exogenously defined groupings, such as caste, but for most group or network links network membership has to be endogenously modelled. Theoretical work has emerged analysing the formation and stability of insurance networks, including in the face of group deviations (for example, Genicot and Ray 2003). Better data-sets focusing on linkages between households in communities has also allowed further evidence to emerge of the extent and nature of risk-sharing within networks and groups. Integrating the endogeneity of network formation for insurance purposes in empirical analysis nevertheless remains a challenge.

The literatures on intertemporal consumption smoothing and risk-sharing in developing countries appear to converge at least in terms of diagnosis: consumption is relatively smooth in the face of income shocks, but not perfectly. Nevertheless, it is often not clear in the tests whether consumption smoothing in practice occurs through transfers or through self-insurance or credit. Even in the ICRISAT data on India, as used in the Townsend data, this has remained an issue of contention (Townsend 1995; Morduch 2004). One strategy is to specifically study the actual responses to shocks (such as dissavings, credit transactions or transfers) and their contribution to smoothing. Deaton and Paxson (1994) provide an alternative test to distinguish whether insurance or credit is responsible for observed smoothing by looking at the changing distribution over time of consumption for a particular cohort in a number of countries, including Taiwan. If consumption smoothing is present due to formal or informal insurance, then inequality can be expected to remain constant over time. If smoothing consumption occurs through credit market transactions, then inequality can be expected to increase over time due to changes in permanent income. Their results suggest that credit markets are more important than insurance for the observed patterns of smoothing.

Income Smoothing

Households do not just use strategies that aim to cope with the consequences of shocks; they may try to reduce or mitigate the risk they face, not least given the limits to risk-sharing and self-insurance. To put it more directly, they may aim to smooth income (Morduch 1995). In rural settings, this strategy has been a central force in shaping farming systems and institutions. The most straightforward strategy is income diversification, whereby income sources are combined and the resulting portfolio faces reduced risk even if the underlying income processes are equally risky when taken separately (as long as they are not perfectly covariate). In some cases, risk management may imply diversifying into (or even specializing in) specific low-risk technologies or activities, such as growing drought-resistant crops or gathering firewood for sale. Social institutions have also developed to include means of reducing exposure to risk. Geographically dispersed marriage patterns, such as those observed by Rosenzweig and Stark (1989) in villages in southern India, can be interpreted as linked to risk diversification within extended families. Local institutions to manage commons and natural resources or land tenure arrangements appear to include risk management as part of their rationale.

Testing the specific role of risk in observed diversification patterns in activities and assets is nevertheless not straightforward. Given the multiple market imperfections faced by the poor, for example in labour markets, the fact of observed multiple activities is not sufficient to sustain the conclusion that risk is its cause. Furthermore, the opportunity to shape risk faced by households also implies econometric problems in standard tests of consumption smoothing and risk sharing, requiring exogenous sources of variation in risk faced. While rainfall variation may provide a useful instrument for common risk, finding sources of exogenous variation for the identification of idiosyncratic shocks is more difficult. These problems suggest that further exploration of risk management strategies will remain methodologically challenging. Furthermore, most investigations of risk strategies in developing countries have been in rural settings, with a focus on agricultural households in relatively stationary environments. This was mainly due to the lack of suitable panel data-sets in urban settings. With more long-term data-sets becoming available, more attention can be paid to the changes in risk strategies following increased integration of local economies and the rise in migration opportunities.

Risk Strategies and Persistent Poverty

All the above hints at an important consequence. While risk strategies contribute to avoiding serious consumption fluctuations, they are not without consequences for welfare, investment and poverty. More specifically, households tend to trade risk and smooth consumption in the short run for lower mean welfare outcomes in the long run. Precautionary motives for saving and credit constraints may induce asset portfolios to focus more on liquidity than on returns. Sales of productive assets for smoothing or taking children out of school to increase labour supply will reduce permanent income. Income smoothing strategies will involve leaving aside profitable opportunities for activity and asset portfolios with a lower mean return. Evidence from villages in the ICRISAT sample in India suggests that these effects may well be substantial, with those households with limited protection against risk (identified by rainfall variability) opting for portfolios with lower returns (Rosenzweig and Binswanger 1993). More specifically, they find that an increase in rainfall variability by one standard deviation would reduce returns to assets by 35% for the poorest wealth tercile of farmers but have no effect on the richest tercile, which is likely to be better protected against risk through its assets or access to credit. If anything, this type of evidence suggests that risk and the lack of appropriate insurance or protection may well be one of the factors that keep poor people poor.

Risk Strategies and Policy Responses

The perceived failure to keep consumption smooth in the face of risk and the long-run costs attached to existing risk strategies has also stimulated an increasing interest in finding appropriate policy responses, not least since insured risk appears to affect the ability of many poor people to grow out of poverty. Standard transfer schemes, such as food aid or cash transfers, may all provide protection against shocks. However, how existing risk strategies could be strengthened remains less explored (Dercon 2004). For example, it is clear that self-insurance through savings could offer substantial benefits in terms of protecting against the consequences of risk. Even if insurance and credit markets were not functioning well, improving the availability of better savings product could assist poor people to improve their risk management. Similarly, informal insurance schemes could be strengthened, for example by linking micro-insurance initiatives to indigenous group-based systems. A number of insurance-related initiatives have been taken in this respect by governments and NGOs. For example, weather-indexed insurance contracts that trigger payments if rainfall falls below a predetermined level are being piloted in a number of countries. Health insurance schemes, often based at local health facilities, have also become more widespread. Much work is still needed on developing these initiatives, not least in terms of evaluating their effectiveness with appropriate techniques, for example randomized interventions.

See Also

► Famines
► Kahneman, Daniel (Born 1934)
► Microcredit
► Risk
► Risk Aversion
► Risk Sharing

Bibliography

Deaton, A. 1992. *Understanding consumption*. Oxford: Clarendon.

Deaton, A., and C. Paxson. 1994. Intertemporal choice and inequality. *Journal of Political Economy* 102: 437–467.

Dercon, S. (ed.). 2004. *Insurance against poverty*. Oxford: Clarendon.

Fafchamps, M., and S. Lund. 2003. Risk-sharing networks in rural Philippines. *Journal of Development Economics* 71: 261–287.

Fafchamps, M., C. Udry, and K. Czukas. 1998. Drought and saving in West Africa: Are livestock a buffer stock? *Journal of Development Economics* 55: 273–305.

Genicot, G., and D. Ray. 2003. Group formation in risk-sharing arrangements. *Review of Economic Studies* 70: 87–113.

Jacoby, H., and E. Skoufias. 1997. Risk, financial markets and human capital in a developing country. *Review of Economic Studies* 64: 311–335.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–292.

Kochar, A. 1995. Explaining household vulnerability to idiosyncratic income shocks. *American Economic Review* 85: 159–164.

Ligon, E., J. Thomas, and T. Worrall. 2002. Informal insurance with limited commitment: Theory and evidence from village economies. *Review of Economic Studies* 69: 209–244.

Morduch, J. 1995. Income smoothing and consumption smoothing. *Journal of Economic Perspectives* 9(3): 103–114.

Morduch, J. 2004. Consumption smoothing across space. In *Insurance against poverty*, ed. S. Dercon. Oxford: Clarendon.

Paxson, C. 1992. Using weather variability to estimate the responses of savings to transitory income in Thailand. *American Economic Review* 82: 15–33.

Rosenzweig, M., and H. Binswanger. 1993. Wealth, weather risk and the composition and profitability of agricultural investments. *Economic Journal* 103: 56–78.

Rosenzweig, M., and O. Stark. 1989. Consumption smoothing, migration, and marriage: Evidence from rural India. *Journal of Political Economy* 97: 905–926.

Rosenzweig, M., and K. Wolpin. 1993. Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investment in bullocks in India. *Journal of Political Economy* 101: 223–244.

Skees, J., P. Varangis, D. Larson, and P. Siegel. 2004. Can financial markets be tapped to help poor people cope with weather risks. In *Insurance against poverty*, ed. S. Dercon. Oxford: Clarendon.

Townsend, R. 1994. Risk and insurance in village India. *Econometrica* 62: 539–591.

Townsend, R. 1995. Consumption insurance: An evaluation of risk-bearing systems in low-income economies. *Journal of Economic Perspectives* 9(3): 83–102.

Rist, Charles (1874–1955)

Roger Dehem

Born at Prilly, Switzerland, 1874; died at Versailles, 1955. Professor at Montpellier (1899–1912) and Paris (1913–33), Rist was the most notable and influential thinker and actor in the field of money in France in the first half of the 20th century. As a member of the *Comité des experts* (1926) and as a vice-governor of the Bank of France (1926–8), he took an active part in monetary reconstruction in the 1920s. He supported the novel idea of stabilization with devaluation (1926–8). He was also involved as an expert in monetary reforms in Romania (1928), Austria, Turkey and Spain. He was France's delegate at the London Economic Conference (1933).

Although Rist is most widely known for his *History of Economic Doctrines,* written in cooperation with Charles Gide, his lasting claim to fame rests on his profound and consistent interpretation of monetary history and thought as demonstrated in his masterwork, *History of Monetary and Credit Theory.* Based on his first-hand experience in times of great instability, Rist's critical analysis of monetary thought from a long-run viewpoint provides an impressive perspective on the evolution of money. By emphasizing the 'store of value' function of money, and by postulating the inability of the state to safeguard it, Rist is critical of authors who supported some form of non-metallic currency, such as John Law, Smith, Ricardo, Wicksell, Knapp and Keynes. He is in sympathy with Cantillon, Galiani, Turgot, Thornton, Tooke and Walras. What he describes as the confusion between money and credit is to be dispelled by drawing a distinction between money proper (gold), credit instruments (convertible banknotes and deposits) and inconvertible paper money. In strong opposition to Keynesianism, Rist is a sceptic in regard of managed currencies and international agreements of the Bretton Wood type. Rist provides the key to the understanding of the French position in

R

monetary matters as opposed to the typical Anglo-American stance in the past 60 years.

Selected Works

1915. (With Ch. Gide.) *A history of economic doctrines*, 2nd edn. London: G. Harrap, 1948.
1924. *La déflation en pratique.* Paris: Giard.
1933. *Essais sur quelques problèmes économiques et monétaires.* Paris: Sirey.
1940. *History of monetary and credit theory from John law to the present day.* London: Allen and Unwin.
1961. *The triumph of gold.* New York: Philosophical Library.

RMB Internationalisation

Gao Haihong

Abstract

China's growing economic weight in the world economy is the fundamental driver for RMB internationalisation. The rationale behind this includes the drawbacks of the current international monetary system, the benefits for Chinese enterprises and financial institutions, and the intention of inviting foreign pressure for domestic reform.

The use of the RMB began in cross-border trade and investment, and soon expanded to offshore markets and financial transactions. It has also become a foreign reserve currency held by some central banks. The RMB's future depends on many conditions, including the pace of China's capital opening, exchange rate flexibility, domestic financial market development and market acceptance.

Keywords

Bilateral currency swaps; Capital account openness; China, currency convertibility; Currency internationalisation; Domestic financial market; Exchange rate flexibility; Foreign reserves; International monetary system; Offshore RMB markets; RMB

JEL Classifications

E52; O53

RMB internationalisation is China's strategy aimed at boosting its currency's international functions as a store of value, unit of account and medium of exchange. It was triggered by the fear of tremendous capital loss of China's foreign exchange reserves due to the US dollar's dominance in the international reserve currency system. It was also China's response to the absence of substantial reform of the international monetary system, a system where its influence did not reflect its increasing economic share in the world. As a result of China's increasing financial openness and domestic financial reform, international use of the RMB expanded rapidly in trade and investment settlement, reserves asset management and the RMB offshore business. However, RMB internationalisation is constrained by China's limited currency convertibility and its underdeveloped domestic financial market. The debate over RMB internationalisation overlaps with other important debates surrounding China's future policy choices and over the likelihood of a diversified international reserve currency system.

The Meaning of RMB Internationalisation and Literature Review

The definition of currency internationalisation generally falls into three basic roles of a fiat money transacted beyond national borders: store of value, medium of exchange and unit of account. Kenen (1983) put forward some early thoughts on the functions of international currencies. Chinn and Frankel (2005) gave an analytical framework and divided the functions into two categories: private and public purposes. For instance, for private purposes an international currency can be used as currency substitution, trade settlement and investment denomination. For public purposes it is used as an official reserve, for foreign exchange intervention and as an anchor for exchange rate pegging. Ito (2011) revisited the functions of international currency and clarified the different functions of invoicing and settlement in trade transactions. An early framework was provided by Gao and Yu (2012), using the up-to-date data to estimate the extent to which the RMB was being used internationally or regionally.

What are the factors that make a currency an international one? The general conditions highlighted by the literature, such as Frankel (1999) and Michalopoulos (2006), include the issuing country's economic size; the stable intrinsic value of the currency; the credibility of the central bank; financial strength and openness; the degree of financial market development; and political and military power.

The degree of capital account openness is the key condition for international use of a currency. A national currency could be regarded as an international currency if most of the free convertible conditions hold: the government must remove all restrictions on the freedom of any entity, domestic or foreign, to buy or sell its country's currency; and domestic and foreign firms are able to use the currency to invoice their trade, both foreign and domestic (Kenen 2012). Although those conditions are critical for a currency's free accessibility, the argument ignores the fact that the offshore market of a currency normally grows under the existence of national capital controls, because market participants can use the offshore market to avoid domestic financial regulations (Gao 2013a).

From the perspective of history, network externality is another key factor for a currency's international status. For instance, for decades, network externality gave the US dollar an 'exorbitant privilege' as a leading international currency (Eichengreen 2011). The scale of bilateral trade and capital flows, the level of development of the financial market and common languages are also important determinants of the geographical distribution of international currencies (Qing et al. 2015).

The experiences of the Japanese yen and the euro provided tremendous reference points for RMB internationalisation. One of the leading papers in this area (Takagi 2012) reviewed the experience of Japan's attempt to internationalise its currency from 1984 to 2003. The study provided rich documentary evidence with regard to Japan's policies to boost the yen's international status and the actual roles that the yen played in the domestic and international markets. Kawai and Takagi (2011) also drew lessons for the RMB from the experience of the Japanese yen. They concluded that the past experience of the yen suggested that strong economic fundamentals could raise the international role of a currency to some extent, but fundamentals alone could not qualify the currency as a key international currency, and the road to that status would not be easy for the RMB. With regard to the euro, Moss (2012) investigated its role in the first years after its introduction and confirmed that international use of the euro tended to be very slow-moving, being characterised by considerable inertia of the US dollar. The paper's historical evidence also suggested changes in the use of international currencies which tended to be associated with large structural breaks in societal, political and economic forces. The European Central Bank in its annual report 'The International Role of the Euro' traced the roles of the euro in the global market and evaluated its importance in official foreign reserve holdings, although euro internationalisation has never been a policy objective of the ECB.

The study on RMB internationalisation was initially focused on estimating the extent of RMB circulation in China's neighboring countries, beginning in the late 1990s. It soon became a policy concern for China, as the level of RMB circulation grew very rapidly. However, most transactions were beyond the official statistic coverage.

RMB internationalisation has both benefits and costs. From the Chinese government's point of view, it can only occur if the benefits are in excess of the costs for the Chinese economy. It is generally believed that global use of the RMB would help to mitigate exchange rate risk for Chinese firms, strengthen the competitiveness of the Chinese financial sector and avoid China's capital loss of foreign reserves invested in US dollars (Gao and Yu 2012). However, economists are also concerned about the potential risks for the Chinese economy. For instance, Gao (2010) warned of the uncertain impact of the currency's international use on the effectiveness of domestic monetary policy through the channel of money aggregates, currency substitution and arbitrage. For the world, economists concentrate on the potential impact of RMB internationalisation on the international monetary system. Prasad and Ye (2012) believe that RMB internationalisation would have the potential to change the international monetary order. However, as the international monetary system becomes more diversified, whether a multi-polar system is more stable than a 'hegemonic' one, such as one in which the US dollar dominates, is unclear. The short-run impact of such regime shifting – moving away from the dollar's dominance towards a multi-currency international monetary system – would increase volatility in foreign exchange markets (Bénassy-Quéré and Pisani-Ferry 2011).

Whether there exists an optimal policy sequence and to what extent capital opening and domestic financial reform matter are also key issues discussed. Views are divergent on the speed and path of the RMB internationalisation, and the sequence of capital account liberalisation, domestic financial market development and exchange rate flexibility has become the focus (Gao 2013b). The People's Bank of China (PBC) is believed to be supportive of relatively radical capital account openness (Gao and Volz 2012). For instance, the PBC Project Team (2012) argued that the risk of capital account openness was controllable, based on factors such as the limited size of foreign debts in the Chinese enterprises, a generally sound financial system and healthy banking sectors in China. Sheng (2013) also believed that the conditions for RMB internationalisation have already been met. even though further capital account opening could help facilitate RMB cross-border movements and enhance the status of the currency in trade and investment transactions. However, Zhang B. (2011a) and Zhang M. (2011b) were sceptical, worrying about the risks of careless openness without necessary reform in the domestic sector and exchange rate policy. Yu (2011, 2014) repeatedly warned that radical capital account opening before exchange rate flexibility and development of a domestic financial market would entail massive speculative capital flows, jeopardising domestic financial stability, which in turn would have a reverse impact on the process of RMB internationalisation.

The rapid development of the offshore RMB market raised discussions on the driving factors and supervision of cross-border capital flow, and the relations between the two markets. The policy motivations behind the growth of the offshore RMB market drew attention from economists because China decided to develop the offshore RMB market in Hong Kong whilst maintaining most capital controls. Destais (2012) argued that the policy was inspired by a form of pragmatism which is both bold and cautious, reflecting China's approach to economic reform since the latter began in the early 1980s. Additionally, a new asset – the offshore RMB – once established would provide business opportunities for both Chinese companies and foreign entities. Such an approach, however, was insufficient for complete currency internationalisation. From the perspective of market functions, some early discussions investigated the functions of offshore markets for currency internationalisation. He and McCauley (2010) emphasised the intermediate functions of offshore markets based on the evidence of major reserve currencies, arguing that offshore markets could help increase the recognition and acceptance of the currency if a country wanted to increase the share of their international balance sheets denominated in their own currencies. This process could begin while substantial capital controls were still in place. Li and Liu (2008) also believed that offshore RMB markets in Hong Kong could facilitate RMB transactions, whilst the risks were controllable under limited currency convertibility. However, Zhang and Xu (2012) found that the massive cross-border flow was driven by arbitrage, irrelevant to the purpose of the RMB being used in trade settlements. The change in size of offshore RMB deposits and cross-border RMB fund flows also drew some attention in studies by McCauley (2015) and McCauley and Shu (2016). The rapid development of offshore RMB markets also raised the question of future labour division in RMB global centres. Since the development of the Shanghai financial centre relies on the agenda of domestic financial deregulation and capital account opening, it would take time for Shanghai to develop itself into one comparable to New York and other established onshore financial centres. It is also reasonable to believe that Hong Kong is likely to be a wholesale market, as it remains the place for intermediating RMB liquidity from and to the Mainland China (Subacchi and Huang 2012).

The role of the RMB in Asia attracted academic attention because RMB internationalisation began with its regionalisation in Asia (Gao and Yu 2012). Furthermore, empirical evidence indicated that the influence of the RMB's exchange rate on other currencies in Asia has increased, especially since China relaxed its exchange rate restrictions in 1995. The evidence during the period 2000–2014 showed that whilst the US dollar played the dominant role in Asia, the impact of the RMB on other Asian currencies increased significantly (Kawai and Pontines 2014). Specifically, the offshore RMB market (CNH) in Hong Kong and the onshore market (CHY) in Mainland had statistically and economically significant impacts on changes in Asian currency rates against the US dollar, showing that the RMB was beginning to exert its growing influence in the Asia-Pacific region (Shu et al. 2015).

The Background and Rationales

In surpassing Japan in 2010, China became the world's second largest economy. However, its weight in the international monetary and financial world was much lower. For instance, the US dollar remains the dominant currency for China's trade and financial transactions. China's trillions of foreign exchange reserves have been mainly invested in dollar assets and fallen into the 'dollar trap' described by Krugman (2009). Such a position makes China's foreign assets highly sensitive to the dollar's yield curve. In 2009, the PBC's governor, Zhou Xiaochuan, published an article arguing that the 'Triffin Dilemma' – the intrinsic flaw of a single currency's dominance in the international reserve system – still existed and was one of the major sources of global imbalance and financial instability. He proposed that a 'super-sovereign reserve currency', independent of individual monetary authorities, was a desirable alternative to the US dollar as an international reserve currency (Zhou 2009). Although the idea was widely regarded as a theoretical concept, it actually was the first time that China explicitly expressed its intention for a diversified reserve currency system.

RMB internationalisation was initially an *ad hoc* process because the Chinese government wanted to be convinced that it was beneficial for China and clear on what costs it would incur before implementing any policy actions.

The potential benefits and costs of RMB internationalisation are summarised by Gao and Yu (2012) as the following. First, it would reduce the exchange rate risks that Chinese firms are exposed to when the RMB is used as unit of account. Second, it can improve the funding efficiency of Chinese financial institutions and hence increase their international competitiveness, because they will enjoy the advantage of having easier access to the vast pool of RMB funds. Third, increasing use of the RMB could reduce holdings of the US dollar as a medium of exchange and store of value, which in turn reduces seigniorage paid to the USA.

However, the degree of RMB internationalisation is conditional on the degree of capital account liberalisation and the convertibility of the currency. It is fair to say that the reluctance of the Chinese government to give up capital controls is the most important impediment to internationalisation of the RMB. The major concerns for rapid financial opening lie in the following. First, China's financial system is still fragile, reflected in market fragmentation, numerous arbitrage and rent-seeking opportunities, the prevalence of government intervention, poor corporate governance, the myopic and speculative tendency of investors, and so on. Second, China's economy is over-monetised when measured by M2 to GDP ratio, which in turn means that outflows of capital could be huge if capital controls were to be dismantled. Third, China's capital markets are still too shallow. Any significant changes in cross-border capital flows may easily lead to large fluctuations in China's asset prices. Fourth, China's economic structure is still rigid. Enterprises are slow to adjust to exchange rate and interest rate changes. Hence enterprises need capital controls to provide breathing space to allow them to make the adjustment. Fifth, China's financial institutions lack competitiveness and some protection is still needed (the infant industry argument).

There is also an argument of 'committed device': the reformers are looking for a lock-in effect on capital account opening and domestic financial liberalisation. China's accession to the WTO was regarded as a successful committed device for domestic reform by inviting external pressure and foreign competitors. China wants to repeat this success and utilise the RMB strategy to break domestic financial bottlenecks. Such an intention was quite subtle initially, but it has become more apparent, as many believe that financial openness and domestic financial liberalisation cannot be separated and should go hand-in-hand.

RMB for Trade and Investment

China gives priority within the RMB strategy to cross-border trade transactions, due to the increasing demand for RMB settlement from Chinese enterprises and their trade partners. In fact, before official ratification, the RMB had been accepted by China's neighbouring countries and circulated outside Mainland China for many years.

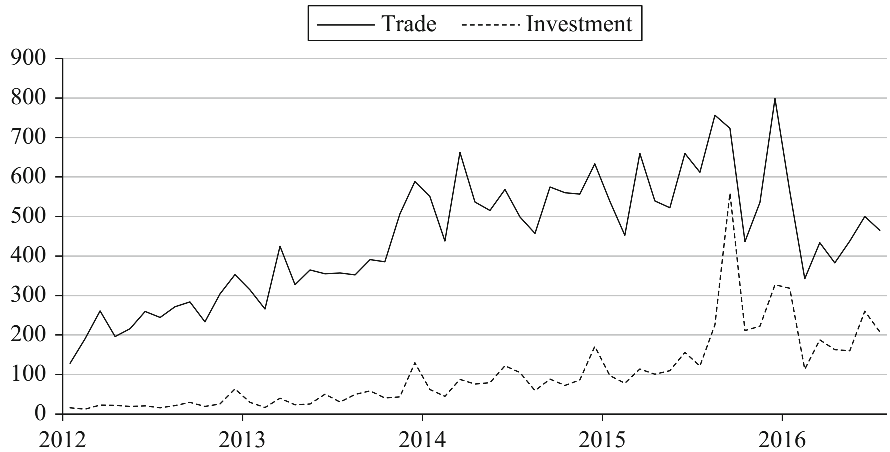
In July 2009, the PBC, together with other Mainland authorities, launched the Administrative Rules on Pilot Program of Renminbi Settlement of Cross-border Trade Transactions. Under the Pilot Program, the approved enterprises and banks could settle their trade transactions in RMB on a voluntary basis, and commercial banks could provide RMB settlement services. The designated areas included five Chinese cities: Shanghai, Guangzhou, Shenzhen, Dongguan and Zhuhai, and the specified areas outside China: Hong Kong, Macao and the ASEAN member countries (PBC 2009). In December 2010, the PBC and other authorities expanded the Pilot Program further by covering 20 designated Chinese provinces and all the areas outside China (PBC 2010). In March 2012, all the restrictions on RMB trade settlement were eliminated. Since 2009, the number of exporters using the RMB has increased rapidly. The volume of RMB settlements increased from zero to over 25% of China's total trade by the end of 2015. According to the Society for Worldwide Interbank Financial Telecommunication (SWIFT), the RMB has become one of the top ten payment currencies in the world.

Compared with trade settlement, the scale of RMB used in direct investment is limited (Fig. 1). In January 2011, China launched the Provisional Rules for the Pilot Program of RMB Settlement for Overseas Direct Investment, opening the gate for Chinese banks and enterprises using the RMB in their overseas direct investments. However, the use of the RMB for inward investment has been overwhelmingly larger than that for outward flows because of the persistent expectation of RMB appreciation during 2012–2014. Foreign firms preferred to claim the RMB assets under the pressure of currency appreciation in the hope of higher returns in the future.

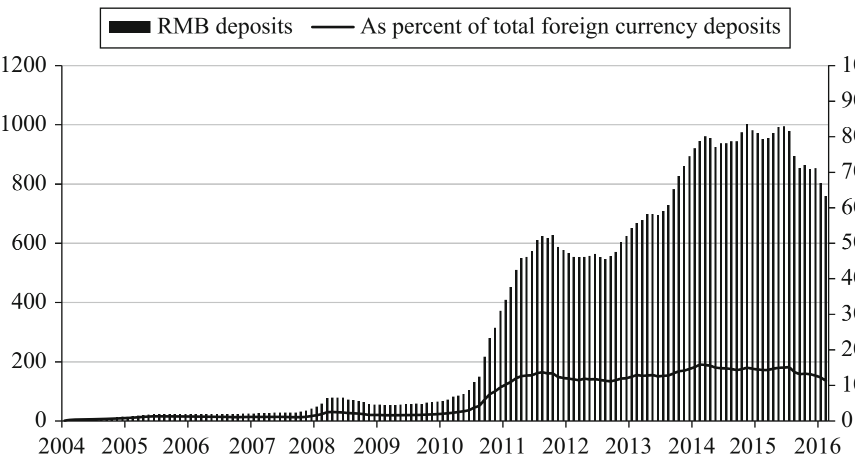
Offshore RMB Markets

The offshore RMB markets play an important role in the process of RMB internationalisation. In February 2004, the Mainland government permitted the banks in Hong Kong to provide RMB services – deposits, currency exchange, remittance, debit and credit cards, bond trading etc. That policy gave an initial push for Hong Kong to carry out RMB business (Fig. 2).

Since 2007, China has taken steps to develop 'Dim Sum bonds' – RMB-denominated bonds issued in Hong Kong. For instance, in June 2007, the PBC and the National Development and Reform Commission announced their



RMB Internationalisation, Fig. 1 RMB trade settlement and investment (2012.1–2016.7, in billion yuan) (Source: Wind database)



RMB Internationalisation, Fig. 2 RMB deposit in Hong Kong (2004.2–2016.2, in billion yuan) (Source: HKMA database and Wind database)

decision to allow policy and commercial banks to issue RMB-denominated bonds in Hong Kong. The China Development Bank issued the first RMB-denominated bond outside the Chinese Mainland, followed by the Export-Import Bank of China, the Bank of China and the Bank of Communications. The initial participation of these state-owned commercial and policy banks was a clear sign of the government's support for developing the RMB bond market overseas. Apart from the Chinese financial institutions, foreign financial institutions, corporate and multinational entities also began raising the RMB fund through issuance of Dim Sum bonds.

Given the limited size and the immaturity of the Chinese domestic bond market, the Chinese Mainland decided to take advantage of the well-developed market in Hong Kong for two reasons. The first is that the issuance of RMB bonds in Hong Kong helps to increase the RMB transactions and build up the pool of RMB funds outside the Mainland. The second is that the issuance of RMB bonds in Hong Kong is seen as the first step towards promoting the involvement of the RMB in the bond market outside the Chinese Mainland. It helps to quicken the pace of the opening of the Chinese Mainland's capital market as well as capital account convertibility. Hong Kong, in fact, has been playing the role of 'experimental ground' for RMB internationalisation – allowing the Mainland government to maintain limited convertibility and at the same time facilitating RMB liquidity in the overseas market.

Thanks to its free market, sufficient expertise and well-established financial infrastructure, Hong Kong has rapidly become the major RMB overseas pool, the largest RMB payment centre in the world, and a multi-currency financial platform for investors to raise RMB funds. The expectation of RMB appreciation was also beneficial for the growth of the RMB pool in Hong Kong, because the arbitrage incentives raised the market appetite for the RMB, which is one of the key variables for RMB attractiveness. However, the resultant cross-border speculative capital flows placed the Chinese authorities in a dilemma: the increase in RMB liquidity required further financial opening, but the ensuing volatile 'hot money' put domestic financial stability in danger.

After Hong Kong, London, Singapore and Taiwan became the other offshore RMB markets. Thanks to the establishment of RMB clearing centres worldwide, the offshore RMB business is expanding rapidly. London, serving as a global connecting hub, plays a crucial role in promoting RMB internationalisation. On 2 June 2016, China's Finance Ministry issued an offshore RMB bond in London, the first time that the Chinese government has sold a sovereign RMB offshore bond outside of China. Before that, such bonds had only been issued in Hong Kong.

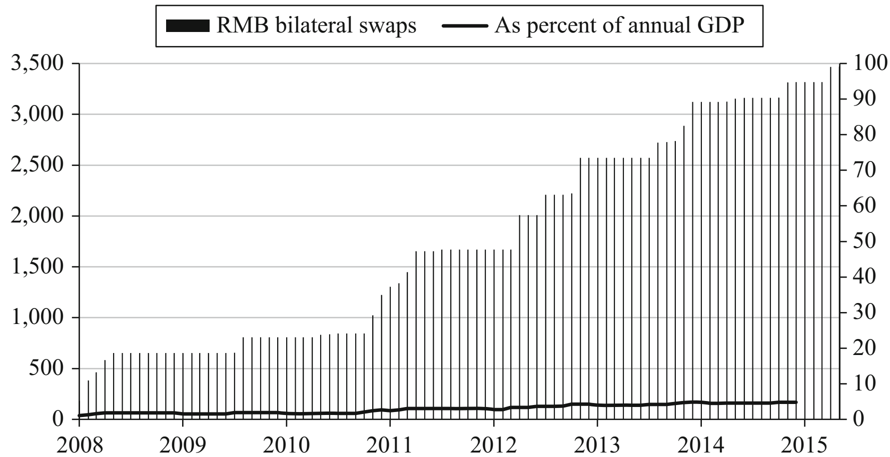
RMB as a Reserve Currency

There are several conditions for a currency to become a reserve currency, including the country having a large share of the global economy, a high degree of currency convertibility and financial openness, a well-developed financial market, stable intrinsic value, a credible central bank, and influential political and military power. Since the collapse of the Bretton Woods System in 1971, the US dollar has played the role of the major reserve currency, followed by the euro (Deutschmark before 1999), pound sterling and Japanese yen.

The RMB was not a reserve currency until the outbreak of the global financial crisis in 2008, when some central banks, such as the Bank of Nigeria, Bank of India and Bank of Japan, began diversifying their foreign reserves and considered including the RMB in their reserve assets. A major step took place when the IMF decided to add the RMB into the Special Drawing Rights (SDR) basket in November 2015. The new basket, to be effective on 1 October 2016, consists of five currencies: US dollar, euro, RMB, yen and pound sterling.

The RMB's inclusion in the SDR basket has multiple implications. First, the international reserve currency system will be more diversified than before with the participation of the RMB. Second, in order to keep the RMB assets accessible for their holders, China has to continue lifting its capital account restrictions. Third, China is eager to enhance the RMB's role through the IMF and considers the RMB's inclusion in the SDR basket as the gateway for the RMB to become a reserve asset. In fact, beginning from April 2016, the PBC released its foreign exchange reserve data denominated in the SDR, in addition to the US dollar. The PBC believes that 'as a currency basket, the SDR tends to be more stable than individual currencies in the basket. Having the SDR as a reporting currency for foreign exchange reserve data would help reduce valuation changes caused by frequent and volatile fluctuations of major currencies, hence provide a more objective measurement of the overall value of the reserve. This would also help enhance the role of the SDR as a unit of account' (PBC 2016a).

Another policy milestone for RMB internationalisation is the large scale of bilateral currency swap agreements signed between the PBC and other monetary authorities. In fact, the RMB as a reserve currency was used in bilateral swap agreements before its inclusion in the SDR basket. The purpose of the swap lines was initially



RMB Internationalisation, Fig. 3 RMB bilateral swaps (2008.12–2016.5, in billion yuan; percentage of China's annual GDP) (Source: various PBC announcements at http://www.pbc.gov.cn and author's calculations)

to give liquidity support and build confidence by sending positive signal to the market on the availability of adequate liquidity in times of crisis. Such agreements normally expire after three years, but can be extended by mutual consent. More importantly, the aims of the RMB swap lines are not only for liquidity support, but also for some extended purposes, such as boosting bilateral trade and investment by way of cooperation between the PBC and other monetary authorities. Also, the PBC expects such types of officially arranged swaps to be a great push for market appetite, making the RMB step up to the next level of its international use.

The first RMB bilateral swap agreement was signed between the PBC and Bank of Korea when Korean banks experienced liquidity shortage in the fall of 2008. Since then, many more bilateral currency swap agreements have been signed between the PBC and other central banks. As of May 2016, the PBC had signed 35 swap contracts (including renewed ones) with other central banks and monetary authorities, amounting to over three trillion yuan in total and accounting for 5% of China's annual GDP (Fig. 3). RMB bilateral swaps, along with other national, bilateral, regional and global liquidity arrangements, comprise important parts of the Global Financial Safety Net (GFSN) – one of the committed objectives of the G20 since 2011.

Capital Account Openness to Support the RMB's International Use

China has been following a gradual path in its capital account liberalisation. In December 1996, China accepted the IMF's article VIII and lifted foreign exchange restrictions in current account transactions. However, China has been very cautious. China has also adopted a general principle of 'crossing the river by feeling stones' and has delivered simple guidelines without a timetable concerning its currency convertibility under capital account transactions.

In 2012, a demarcation line was drawn for China's financial openness. Before that, China had retained the status quo with regard to capital control. However, after deciding to seize the window of opportunity for the RMB strategy resulting from the global financial crisis and the rise of its own economy, the Chinese government began to accelerate the speed of capital account openness. The PBC, which is the most liberal-minded among the Chinese governmental decision-making bodies, took the first step by laying out a timetable. It was published in the newspapers under the name of the PBC Project Team. According to the PBC Project Team (2012), China's plan was to achieve full currency convertibility by 2022. There were four stages with different risk ranks: at stage one, regarded as the lowest risk rank, China would lift restrictions on direct investments and encourage Chinese enterprises to 'go abroad'; at stage two, regarded as a middle risk rank, China would relax its control on commercial credits; at stage three, regarded as a high risk rank, China would open its domestic debt securities, equities and real estate markets to non-residents; and at stage four the items including money markets, financial institutional credit, resident capital transactions, collective securities, guarantees and derivative products were at the highest risk rank. China would retain controls without deadlines. Furthermore, in the Third Plenary Session of the Eighteenth Central Committee of the Communist Party, the leadership committed to achieving full capital liberalisation by 2020.

A major breakthrough in China's financial account openness was the establishment of the Shanghai Free Trade Zone (SFTZ) in September 2013. The purpose of the SFTZ was to serve as another experimental case for RMB convertibility and financial liberalisation, similar to Hong Kong in the early stage. But this time, the experiment was carried out in an onshore market, by allowing financial institutions and non-financial companies registered in SFTZ to conduct free RMB transactions under macro-prudential management, instead of capital control under the PBC and State of Administration of Foreign Exchange (SAFE). For instance, companies are allowed to set up a Free Trade (FT) account so that they can conduct free exchange between the RMB and foreign currencies; foreign banks are allowed to provide RMB syndicated loans to domestic enterprises; multinational corporations can transfer RMB funds between the parent company to its subsidiaries via capital account; and companies registered in SFTZ can borrow RMB outside the country. It was the first time that the relaxation of foreign exchange controls was tested in an onshore free economic zone.

However, critiques were initially centred on whether the policy makers could effectively prevent risks from spilling over to areas outside SFTZ, and whether the partial success of opening could be replicated and rolled out nationwide. In January 2016, the PBC expanded its framework of macro-prudential management on capital flows and currency transactions from SFTZ to three other free trade zones – Guangdong, Tianjin and Fujian – allowing 27 financial institutions and all the companies registered in those zones to enjoy free RMB transactions. In May 2016, the PBC decided to roll out its framework of macro-prudential management nationwide (PBC 2016b).

In accordance with the demand for tracking international use of the RMB and the changes of policies underlying the development, the PBC began to publish the Annual Report on RMB Internationalisation in 2015. The report provided a statistical summary of the use of the RMB for current account transactions and for capital account transactions and the RMB holdings by non-residents. It also outlined the policies implemented during the reporting period to support RMB internationalisation, such as those of openness in interbank markets, equity markets and bond issuance. For instance, the increase in the quotas for RMB in Qualified Foreign Investment Institutions (RQFII) and in Qualified Domestic Investment Institutions (RQDII) and the establishment of Shanghai-Hong Kong Connect were regarded as a policy push for further opening of financial transactions.

Exchange Rate Flexibility

The relationship between the function of an international currency and exchange rate regimes is unclear. For instance, since the establishment of the Bretton Woods system, the dollar has experienced different types of exchange rate regimes, from an adjustable one to a floating one. However, changes in the exchange rate regime did not change the status of the dollar as an international currency. As pointed out by Gao and Yu (2012), there is no firm answer to the question of whether a free-floating regime is a precondition for the RMB to become an international currency.

However, exchange rate regimes matter when capital account openness is considered as a key condition for RMB internationalisation. This is because, as the theoretical triangle of impossibility suggests, if China is to implement its aggressive timetable of capital account opening, it has to make its currency flexible in order to retain the autonomy of the central bank's monetary policy. Therefore the relationship between capital account liberalisation and exchange rate flexibility becomes the focal debate with regard to the policy sequence for boosting RMB internationalisation.

China has changed the exchange rate regime, since July 2005, from a dollar peg to a managed float one. The PBC has also broadened the fluctuation bands several times since then. At the same time, the RMB markets have grown rapidly because of the aggressive deregulation of financial transactions. As a result, there exist three RMB markets – the onshore RMB spot rate (CNY) and two offshore market rates in Hong Kong: non-deliverable forward rate (NDF) and offshore spot rate (CNH). The different pricing of the RMB in different markets entails arbitrage and puts domestic financial stability in danger. Under such circumstances, a flexible exchange can be used as a buffer to absorb external shocks, giving the monetary authority space for its domestic objectives.

On 11 August 2015, the PBC decided to relax its intervention in the RMB middle price. Such a policy move was regarded as hitting two birds with one stone: boosting China's exports by depreciating its currency and making the regime shift toward its long-run objective of exchange rate flexibility. The latter clearly helped to clear the hurdle before the deadline for the IMF's evaluation of the RMB's qualification for the SDR basket currency in 2015. However, the subsequent severe depreciation and instant capital outflows, which resulted from many downside factors, forced China to re-peg its currency against a basket. In fact, China faces a dilemma with regard to RMB internationalisation: to keep its commitment to capital account opening while at the same time maintaining financial stability.

Domestic Financial Market Development

Historical experience shows that a deep and liquid direct financing market is important for currency internationalisation, as it can narrow the bid–ask spread and lower transaction costs, hence increasing its attractiveness to international investors. However, China's domestic financial market has not been yet developed to the level matching the needs of the RMB's internationalisation (Gao 2013a). The major constraints are as follows. First, China's financial structure is typically a banking dominated one, where bank loans account for about 60–70% of China's aggregate financing. Normally, a market-based financial system with dominating direct financial instruments is more supportive than a bank-based one. Second, China's domestic financial markets are fragmented, and are subject to multiple governmental administrations. The transaction costs are relatively high for investors, compared with the international standard. Third, there is a lack of a reliable credit rating system, which is crucial for the development of a domestic bond market. Lastly, the development of China's domestic financial market is interlinked with reforms in other areas, such as the reform of the State-Owned Enterprises (SOEs), rule-of-law-based reform of governance and structural changes to the supply side of the economy. The rise of shadow banking and the increasing debt burden have also become barriers to further liberalising of domestic financial sectors.

A liberalised interest rate mechanism would allow market participants to take advantage of the quick response of interest rates resulting from international use of the RMB, a major benefit for the increase of liquidity and depth in China's financial market (Gao 2013b). China finalised its interest rate liberalisation in 2015, a milestone for RMB internationalisation. For a long time, the Chinese financial system was repressed under interest rate and credit controls, and the subsequent negative real rate depressed household consumption on the one hand and subsidised state-owned enterprises and investment on the other hand. This type of capital misallocation has been a major contribution to China's domestic economic imbalance.

Another milestone was the improvement of the financial infrastructure. China has made great efforts to facilitate RMB internationalisation by setting up designated Chinese banks and establishing many clearing centres worldwide.

The breakthrough was that, on 8 October 2015, China launched the RMB Cross-border Inter-bank Payment System (CIPS), a centralised, real-time, gross settlement system, adopting the ISO20022 message dashboard and compliant with Principles for Financial Market Infrastructures and other international regulations, and with full coverage of operation hours for the time zones where RMB business take place (PBC 2015). The operation of such a system greatly reduces the transaction costs and enhances the efficiency of cross-border RMB and offshore RMB business, and is a boost for the RMB internationalisation.

Conclusion

RMB internationalisation gives China the financial power to match its growing influence in the world economy. It is the logical outcome of changing economic weights, and as the RMB increasingly becomes acceptable in the world, the currency can serve as an alternative reserve, helping to overcome the problem of a shortage of global safe assets. The future of international monetary systems would be more diversified with the rise of the RMB. It would also require China to have more responsibility in global financial governance and to be more involved in the risk-sharing and rule-making processes.

The extent of RMB internationalisation will ultimately be determined by many conditions, including economic size, freedom of capital flows, credibility of the central banks and political factors. It is also a work in progress. RMB internationalisation depends in particular upon China's policy adjustments. China's financial opening is a major step forward, because it increases accessibility to the currency for free international transactions. However, a country's financial opening is never risk-free. The past debates discussed here reflect this reality. China's well-developed domestic financial market and flexible exchange rate are crucial for the smoothness of the process of RMB internationalisation.

Ultimately, the success of RMB internationalisation will result from the aggregation of market-driven decisions. It's fair to say that the policy-driven steps for RMB are necessary at the initial stage because the removal of existing institutional barriers can facilitate market transactions. However, in the later stages its international functions will mainly be determined by market forces.

China expects that with the RMB's participation the international monetary system can be more balanced and fairer than before. Such an expectation can only materialise if the resultant redistribution of financial powers meets the needs of all the major players in the world. There are also unanswered questions, such as whether a multipolar international monetary system is stable. If diversification of the reserve system is inevitable, the question of how to mitigate the risks for individual countries remains a subject to investigate.

See Also

► China, Economics in
► Currency Boards

Bibliography

Bénassy-Quéré, A., and J. Pisani-Ferry. 2011. *What international monetary system for a fast-changing world economy?* Presented at Asia-Europe Economic Forum, Paris, 10–11 January.

Chinn, M., and J. Frankel. 2005. *Will the euro eventually surpass the dollar as leading international reserve currency?* NBER Working paper. No. 11510.

Destais, C. 2012. China cannot achieve the internationalization of its currency without making it fully convertible. Money & Finance. CEPII.

Eichengreen, B. 2011. *Exorbitant privilege: The rise and fall of the dollar and the future of the international monetary system*. Oxford: Oxford University Press.

Frankel, J. 1999. *No single currency regime is right for all countries or at all times*. NBER Working paper series 7338.

Gao, H. 2010. Internationalization of the renminbi and its implications for monetary policy. In *Currency internationalization: International experiences and implications for the Renminbi*, ed. C. Shu and W. Peng, 209–220. Basingstoke: Palgrave Macmillan.

Gao, H. 2013a. Convertibility as a step for the RMB internationalization. *Economic Change and Restructuring* 46(1): 71–84.

Gao, H. 2013b. The RMB internationalization, capital account opening and chinese domestic financial reform. *Review of International Affairs* LXIV(1152): 5–24.

Gao, H., and U. Volz. 2012. Is China ready to open its capital account? 29 March. Available at http://www.eastasiaforum.org/2012/03/29/is-china-ready-to-open-its-capital-account/.Accessed 15 Aug 2016.

Gao, H., and Y. Yu. 2012. Internationalization of the renminbi. *BIS paper* 61: 105–124.

He, D., and R.N. McCauley. 2010. *Offshore markets for the domestic currency: Monetary and financial stability issues*. BIS Working papers. No. 320.

Ito, T. 2011. The internationalization of the RMB: Opportunities and pitfalls. Prepared for symposium *The future of the international monetary system and the role of the Renminbi*. Organised by the Council on Foreign Relations and China Development Research Foundation.

Kawai, M., and V. Pontines. 2014. *The renminbi and exchange rate regimes in East Asia*. ADBI Working paper No. 484.

Kawai, M., and S. Tagaki. 2011. *The RMB as a key international currency? lessons from the Japanese experience*. Notes prepared for the Asia-Europe Economic Forum, Paris, 10–11 January.

Kenen, P. 1983. The role of the dollar as an international currency. In *Occasional Papers 13*. New York: Group of Thirty.

Kenen, P. 2012. Currency internationalization: An overview in currency internationalization: Lessons from the global financial crisis and prospects for the future in Asia and the Pacific. *BIS Papers* 61: 9–18.

Krugman, P. 2009. *China's dollar trap*. Available at: http://nytimes.com/2009/04/03/ opinion/03krugman.html. Accessed 15 Aug 2016.

Li, D., and S. Liu. 2008. Dual advancement for RMB internationalization. *China Finance* 11: 1–16.

McCauley, R. N. 2015. Capital flowed out of China through BIS reporting banks in Q1 2015. BIS Quarterly Review 28–9.

McCauley, R. N., and C. Shu. 2016. Dollars and renminbi flowed out of China. *BIS Quarterly Review* 26–7.

Michalopoulos, G. 2006. The internationalization of the euro: Trend, challenges and risks. In *Global divergence in trade, money and policy*, ed. A. Volbert and H. Kotz, 195–214. Cheltenham: Edward Elgar.

Moss, F. 2012. The euro: Internationalized at birth. *BIS Papers* 61: 57–74.

PBC Project Team. 2012. The basic conditions are mature for accelerating China's capital account opening. China Securities Journal .

PBC. 2009. *Administrative rules on pilot program of renminbi settlement of cross-border rade transactions*. Available at: http://www.pbc.gov.cn/english/130721/2884989/index.html. Accessed 15 Aug 2016.

PBC. 2010. *Supporting pilot program for RMB settlement of Cross-border trade transactions to facilitate trade and investment*. Available at: http://www.pbc.gov.cn/english/130721/2849409/index.html. Accessed 15 Aug 2016.

PBC. 2015. *Milestone of RMB internalization: RMB Cross-border Inter-bank payment system starts operation*. Available at: http://www.pbc.gov.cn/english/130721/2963649/index.html. Accessed 15 Aug 2016.

PBC. 2016a. People's bank of China announced to implement macro prudential management of Ccoss border financing nationwide. Available at: http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3055696/index.html. Accessed 15 Aug 2016.

PBC. 2016b. SDR-denominated foreign exchange reserve data released by the PBC. Available at: http://www.pbc.gov.cn/english/130721/3045077/index.html. Accessed 15 Aug 2016.

Prasad, E., and L. Ye. 2012. *The renminbi's role in the global monetary system.* IZA Discussion paper No. 6335.

Qing, H., I. Korhonen, J. Guo, and F. Liu. 2015. The geographic distribution of international currencies and RMB internationalization. *International Review of Economics and Finance* 42(3): 442–458.

Sheng, S. 2013. Why China needs to accelerate capital account openness. Financial News..

Shu, C., D. He, and X. Cheng. 2015. One currency, two markets: the renminbi's growing influence in Asia-Pacific. *China Economic Review* 33: 163–178.

Subacchi, P., and Huang, H. 2012. The connecting dots of China's renminbi strategy: London and Hong Kong. *Chatham House* .

Takagi, S. 2012. Internationalizing the Yen, 1984–2003: Unfinished agenda or mission impossible? *BIS paper* 61: 75–92.

Yu, Y. 2011. Revisit of the road map of RMB internationalization. *RCIF Policy Brief* No. 2011.056.

Yu, Y. 2014. *How far can renminbi internationalization go?* ADBI Working paper No. 461.

Zhang, B., and Q. Xu. 2012. RMB internationalization under exchange rate and capital account control. *International Economic Review* 4: 63–76.

Zhang, B. 2011a. RMB internationalization: Sequencing upside down. *FTChinese, Financial Times*.

Zhang, M. 2011b. *The RMB internationalization: Onshore and offshore market*. RCIF Working paper No.2011W09.

Zhou, X. 2009. Reform the international monetary system. *BIS Review* 41/2009.

R

Robbins, Lionel Charles (1898–1984)

B. A. Corry

Keywords

Austrian economics; Beveridge, W. H.; British classical economics; Cambridge School; Economic liberalism; Economic science; Hayek, F. A. von; History of economic thought; Interpersonal utility comparisons; Keynes, J. M.;

Labour supply; Laissez-faire; Methodology of economics; Philosophy and economics; Representative firm; Robbins, L. C.; Scarcity; Torrens, R.; Value judgements

JEL Classifications

B31

Lionel Robbins, who in 1961 became Baron Robbins of Clare Market, was one of the major academic economists of the interwar period. He remained active after the Second World War but never really regained the centre of the stage that he had occupied. He was also a great public servant for his country, serving it well and loyally in many aspects of social, political and cultural life. He was truly a 'Renaissance man'.

Robbins was born in 1898 in Middlesex, the son of Rowland Richard Robbins – for many years President of the National Farmers' Union – and Rosa Marion Robbins. He spent one year reading for an Arts degree at University College London and then volunteered for war service with the Royal Artillery. He saw active service on the Western Front, was wounded and invalided back to England in 1918. He was an undergraduate at the London School of Economics and Political Science from 1920 to 1923, from which he graduated with a BSc (Econ.) degree, choosing political ideas as his major field of study, and having had the left-wing Harold Laski as his tutor. Beveridge employed him as a research assistant for a year, after which Robbins was a tutor in economics at New College, Oxford. He returned to teach economics at LSE from 1925 to 1927, then back to New College as Fellow (1928–1929) and finally, at the incredibly young age of 31, back to the Senior Professorship in Economics at LSE to succeed Allyn Young.

Apart from government service during the Second World War, Robbins remained at LSE as Head of the Economics Department until 1960 when, on accepting the Chairmanship of the *Financial Times*, the University of London forced him to resign his professorship – a move than brought Robbins great personal distress, although he retained his connection with LSE and taught courses there until a year or so of his death in 1984.

Outside academic and government advisory activity, Robbins had a distinguished record in arts administration, being connected with both the National Gallery and the Royal Opera House, but he may perhaps be best remembered, in such 'outside' activities, for his contribution to the structure of higher education in the United Kingdom. He chaired the committee – commonly referred to as the Robbins Committee on Higher Education – that proposed the criterion that all qualified applicants should receive a place, and financial support, to read for a degree. The acceptance of the 'Robbins Principle' led to a vast expansion of degree courses, especially in the social sciences in the 1960s and early 1970s in the UK.

Robbins's contributions to economics may be considered under four headings; economic theory, methodology and philosophy of economics, the theory of economic policy, and the history of economic thought.

Those who only knew Robbins later in his life often forget that he made his initial mark in economics as a theorist. Three contributions here are worth noting; he launched a sustained attack on Marshall's concept of the Representative Firm which was apparently so successful that it drove the concept out of the pages of microeconomic texts. Robbins basically argued that the understanding of the equilibrium neither of the firm nor of the industry was aided by introducing the Representative Firm, hence it should be eliminated from analysis. More recent work has shown a greater sympathy towards Marshall's construct and it seems clear now that Robbins failed to understand the exact dynamic problem that Marshall was trying to cope with and why the Representative Firm was an important contribution to this problem.

Robbins also pioneered the micro-analysis of the labour supply function. Although he did not explicitly use the division of a wage change into an income and substitution effect, he showed clearly why the sign on the response of hours to a real wage rate change would be ambiguous.

In macroeconomics Robbins was a firm exponent of the Austrian theory of the trade cycle and here he was greatly influenced by Frederick von Hayek, whom he brought to LSE from Vienna in 1928. The central feature of the Austrian analysis was that depression was due primarily to undersaving (or excess consumption) and these views, which Robbins expounded as an explanation of the 1930s depression in his book *The Great Depression,* led to a head-on collision between the senior LSE economists and the Cambridge School centred around Keynes. This rift was not finally healed until the wartime collaboration in Whitehall between Robbins and Keynes. After the war in the Marshall Lectures for 1946, published as *The Economic Problem in Peace and War*, Robbins announced his conversion to full employment policies via control of aggregate demand, although it is not clear that he became a Keynesian.

The second area where Robbins made a major contribution and where he wrote what is probably his best known work in economics was that of the methodology and philosophy of economics. His *Nature and Significance of Economic Science* was one of the most cited, if not most read, books on the subject in the period 1932–1960, and it influenced greatly economists' views about the nature of their discipline. There are several strands to the book, none original in themselves, but Robbins put them together in beautifully clear prose and in a very persuasive manner. The major themes were; first, that economic science could be clearly demarcated from those discussions of economic issues that involved value judgements – by which latter term Robbins meant evaluative statements of the form 'better or worse' where interpersonal comparisons of utility were involved. He also argued that there was a clear demarcation between economic science and other branches of social enquiry such as social psychology, sociology, politics and so on.

The second major theme was that the subject matter of economic science was not a particular activity (for example, Cannan's view that economics was the science of wealth), but rather an aspect of all human conduct. This aspect was the 'fact' of economic scarcity – a manifestation of unlimited ends on the part of individuals and society and means of satisfying those ends that were limited in supply. In words so often quoted in economics texts Robbins defined economic science as ' that science that studies the relationship between ends and means that have alternative uses' – a definition that is more than reminiscent of Menger's exposition of the economizing process.

These two aspects of the *Nature and Significance* were widely accepted by the world of academic economists and are still propagated. But they have always had their critics; in particular, the view that there is a body of scientific economics 'free from value' is much disputed.

The third aspect of the book – Robbins's views on the procedures for checking the validity of economic theory – was less fortunate in its effect on the development of the subject. Robbins appeared to argue that the central propositions of economics were derived from very basic, and obvious, assumptions and a process of logical deduction from these assumptions. Moreover, these deductions gave essentially qualitative predictions. Robbins expressed great scepticism about the feasibility and meaningfulness of quantitative work in economics, and by the implication of his message inhibited the development of econometric testing in economics.

Robbins's contributions to discussions of economic policy were basically consistent throughout his career, although the purity of his earlier thoughts was muddled as he grew older. His major policy theme was his advocacy of, what may be loosely termed, economic liberalism. Robbins decreasingly argued this on the grounds of some alleged theoretical or a priori superiority of market solutions over collectivist or interventionist plans, but rather as an empirical point that the liberal solution seemed best to combine liberty and efficiency. In his earlier writings, for example *The Economic Causes of War* (1939a) and *The Economic Basis of Class Conflict* (1939b) he adopted an extreme free trade position and it was this stance as much as macro-theory debate that led to his conflict with Keynes in the 1930s. His later work revealed a much greater readiness to allow ad hoc exceptions to strict economic liberalism – he espoused, among other measures, the Beveridge plan, grants for higher education, subsidies for the arts, control of the exports of works of art, overall macro-control for full employment. Probably the most rounded statement of his policy beliefs is to be found in his *Economic Problem in Peace and War.*

Finally, mention must be made of Lionel Robbins's contribution to the teaching and study of the history of economic thought. He, together with one or two other scholars of his generation – like his great friend, Jacob Viner – kept interest in the subject alive and flourishing when many economists regarded it, as they still do, as irrelevant to their studies. Much of his influence came via his masterly teaching of the subject and via the important theses that were produced under his supervision, as much as from his own specific contributions. He also aided the production of important series in the history of economic thought such as the LSE reprints and the collected works of Bentham and J.S. Mill.

Of his specific contributions, two are minor classics, his *Theory of Economic Policy in Classical Political Economy* (1952) and *Robert Torrens and the Evolution of Classical Economics* (1958). In the former work, Robbins argued very persuasively, if not entirely convincingly, that the British classical economists did not adhere to the Continental laissez-faire dogma but rather argued for freedom in economic relationships as a general principle with many ad hoc exceptions. He further tried to clear them of any anti-working class bias.

The book on Torrens is a perfect example of how to survey the collected works of a writer who, though not of the first rank of classical economists, is nonetheless a useful writer by whom to assess the general achievement of the classical school.

See Also

► Value Judgements

Selected Works

1932. *An essay on the nature and significance of economic science*. London: Macmillan.

1934. *The great depression.* London: Macmillan.

1937. *Economic planning and international order.* London: Macmillan.

1939a. *The economic causes of war.* London: Jonathan Cape.

1939b. *The economic basis of class conflict*. London: Macmillan.

1947. *The economic problem in peace and war*. London: Macmillan.

1952. *The theory of economic policy in english classical political economy.* London: Macmillan.

1958. *Robert Torrens and the evolution of classical economics*. London: Macmillan.

1971. *Autobiography of an economist*. London: Macmillan.

Robertson, Dennis (1890–1963)

M. Anyadike Danes

Keywords

Cooperative monetary economy; Cooperative non-monetary economy; Cyclical fluctuations; Keynes, J. M.; Liquidity preference; Rate of interest; Robertson, D.

JEL Classifications

B31

Dennis Robertson was born in 1890, the son of a clergyman and schoolmaster, and was educated at Eton and Trinity College, Cambridge. After taking a Part I in Classics and Part II in Economics he was elected a Fellow of Trinity College in 1914 and in 1930 became a Reader in the University of Cambridge. He left Cambridge in 1938 to become a Professor in the University of London but during most of his time in that post he was seconded to the Treasury on war-related work. Elected in 1944 to succeed Pigou in the Chair of Political Economy, he returned to the University of Cambridge, holding that position until his retirement in 1957. He died in Cambridge in 1963.

Economics in Cambridge when Robertson commenced working at it was dominated by

Marshall. Not by the man himself (although still alive he had retired in 1908) but by his analytical methods and by his *Principles of Economics*. It was quite natural that the topic selected by Robertson for his fellowship dissertation should involve a 'Marshallian' approach to a subject on which Marshall himself had written relatively little: the nature and causes of fluctuations in the general level of economic activity. As Robertson recorded in the introduction to the published version of this dissertation:

> In some of the more abstract portions of this essay I shall make use, without further explanation or apology, of the processes and terminology in common use among the school of economic thought associated in this country chiefly with the name of Dr Marshall. My reason is that after a study of many facts and theories I am deliberately of the opinion that one cause of the obscurity which still surrounds this problem is that in the attack upon it full and systematic use has never hitherto been made of the weapons supplied by this particular intellectual armoury. (1915, p. 11)

Although Robertson did not suspect it then, the refinement and further development of the ideas about cycles and growth in economic activity presented in this study were to occupy him for the next 20 years. Two different sorts of factors led him in this direction. The first was the need to develop a framework for designing an organized policy response to the large-scale dislocation of economic life which had followed the end of the First World War, while the second was a more specific, personal, influence. In the early 1920s Keynes commissioned him to write an introductory textbook (in the Cambridge Economic Handbook series) to be entitled, simply, 'Money'. The difficulties he encountered in attempting to provide an elementary account of monetary theory made Robertson particularly aware that, even in its more sophisticated variants, existing theoretical work provided an inadequate basis for dealing with the economic problems of the 1920s. The combined influence of these two resulted in a prolonged period of reflection and research, yielding a series of loosely related publications which recorded the development of a fairly comprehensive analytical scheme interrelating the problems of money, the trade cycle, economic growth, and the role of the state in promoting economic progress.

Robertson's approach to this analysis involved the development of successively more complicated, more 'realistic' models of economies each of which constituted a different, abstract, 'type'. All 'types' shared the characteristics that production was undertaken on the basis of 'rational' decision-making by competing producer 'groups', each making different products with a fixed labour force and a productive process involving fixed capital. Now although each type of economy was both a *production* and an *exchange* economy it was the possibility that these activities could be 'organized' in different ways that distinguished the different types. Production could be organized in two ways, cooperatively or non-cooperatively, while exchange could also be organized in two ways, direct exchange or monetary exchange. In total there were, then, four types of economies. The distinction between the two types of productive organization turned on the decision-making functions of the members of the groups: in a cooperative group decisions were *made* and *carried out* by the group members acting together, while in a non-cooperative group 'entrepreneurs' made decisions and 'workers' carried them out. In respect of the organization of exchange it was on the existence and use of money that the distinction rested, in one case exchange was carried out by 'direct barter', while in the other, money supplied through a (potentially) government-controlled banking system provided the means of exchange.

Robertson's basic analytical building block was the 'cooperative non-monetary economy', an economy where each competing industrial group made its employment and, thus, output decision cooperatively, and exchanged its output without the use of money. Although in such an economy no distinction was made between the members of the group, a distinction was made between two different categories of producer groups, those providing consumer goods and those producing capital goods. The first group, consumer goods producers exchanged some of their output with the second group for capital goods, thereby providing consumption goods for capital goods producers. Now an economy of this type, Robertson argued, would experience cyclical fluctuations in aggregate output deriving from the effect of gestation lags on the time pattern of the supply of capital goods and of the durability of capital goods on the time-pattern of demand for their replacement.

A non-cooperative non-monetary economy would, though, experience fluctuations of even greater severity than those felt in an otherwise identical cooperative economy. This proposition derived directly from the fact that in a non-cooperative economy production decisions were taken by entrepreneurs who hired workers to carry them out, and workers and entrepreneurs had differing 'interests'. These divergent interests were reflected most importantly in the different utility attached to leisure by the two classes. An entrepreneur, for example, would wish to expand output further in the boom and contract it further in the slump than the workers in his group; and since entrepreneurs were in control, their interests prevailed. Although the degree of fluctuation in the non-cooperative economy was more pronounced than in the cooperative, Robertson adopted it as the benchmark which defined the 'appropriate' degree of fluctuation to be aimed at by policymakers concerned with stabilization. He did so because he maintained that the failure to recognize that production was, in practice, organized non-cooperatively could lead to an attempt to reduce fluctuations too much. Such attempts, by altering the structure of incentives, could damage the longer-run growth possibilities of the economy.

The cooperative monetary economy construct was built directly on to the foundations provided by the cooperative non-monetary economy and this type of economy exhibited, therefore, a cyclical pattern in the production of fixed capital which generated cyclical fluctuations in output as a whole. Now the introduction of money also required a slight change of focus, since in the monetary case Robertson concentrated not on fixed capital but on the demand for circulating capital, essentially on the demand for consumption goods which were consumed by those engaged in the process of production. This concern with circulating capital was necessarily associated with the analysis of monetary economies because Robertson made the assumption (reflecting British banking practice) that it was with the finance of the acquisition of circulating capital that the banking system was concerned. His analysis then described the policies which, if adopted by the banking system, would lead to fluctuations being of no greater amplitude than in the corresponding non-monetary economy. A failure to implement such policies would lead to fluctuations in the price level, and thus in output, of greater magnitude than was 'appropriate'.

The difference made by the substitution of cooperation in the monetary type turned principally on the effect on decision-making of changes in income distribution. In particular, it was assumed that only entrepreneurial incomes adjusted quickly to changes in the price level, so that variations in the price level over the cycle were an additional source of influence on production decisions. The nature of this influence led entrepreneurs to expand their activities further in the boom (as rising prices increased their profits) and contract them further in the slump (as falling prices reduced their profits) than would have been the case in the corresponding cooperative economy. But these changes in income distribution were not permanent. In the course of the boom workers managed to restore real wages to pre-recovery levels, the expansion of output would be slowed, and in the slump, as entrepreneurs restored profits to their pre-depression levels, the contraction of output would be slowed. The end of the boom and the slump, though, if an 'appropriate' monetary policy were adopted, would be dictated by the behaviour of the underlying non-cooperative non-monetary economy. So non-cooperation in the monetary case had additional effects only on the amplitude of cyclical fluctuations.

Robertson also developed a set of tools to analyse the process of cyclical change in monetary economies. He divided time up into a sequence of market periods (during each of which the supply of goods was fixed) and then focused on the dynamics of the transfer of resources from current consumption by those already in employment to those newly employed to increase output during the expansion phase of the cycle. The mechanism generating this transfer was a price-level increase as the newly employed (whose wages had been borrowed from the banking system) outbid the existing employed on the goods market. Robertson's aim was to show how the magnitude of the price-level increase was determined and the nature of the monetary policies which could be adopted in order to minimize it. The rate of inflation was shown to depend upon the relationship between the rate at which the banking system made new loans to producers and the rate at which this new money was absorbed into the money-holdings of the existing employed. The faster the new money was absorbed, that is, the faster that the existing employed gave up their claims on current output, the smaller the rise in the price-level accompanying the transfer of resources from the public to the expanding producers. To the extent that this money was not immediately absorbed, the existing employed were *forced* to share current output with those producers by price-level changes. By minimizing these changes, then, the monetary authority through its control of the banking system would also be able to minimize the amount of 'forced' saving which accompanied the recovery. A similar approach was also applied to the non-cooperative case, but here policy design was more difficult because the inflation led to changes in the distribution of income between workers and entrepreneurs. Even so, monetary policy could play a useful role in reducing fluctuations to their 'irreducible' non-monetary amplitude.

The central concern of Robertson's analytical work was to provide an explanation of fluctuations in aggregate activity which was closely linked to a broader concern, that of remedying the adverse effects of such fluctuations. The identification of the use of capitalistic (though not necessarily capitalist) production methods as the source of fluctuations, though, left with a rather ambivalent attitude to possible remedies: capitalistic production methods always produced cycles, but also brought with them the possibility of economic progress. And he thought that there was a trade-off between these two, greater stability being associated with slower growth, less stability with faster growth:

> From some points of view the whole cycle of industrial change presents the appearance of a perpetual immolation of the present upon the altar of the future. During the boom sacrifices are made out of all proportion to the enjoyment over which they will ultimately give command: during the depression enjoyment is denied lest it should debar the possibility of making fresh sacrifices. Out of the welter of industrial dislocation the great permanent riches of the future are generated. (1926, p. 254)

He concluded that the choice between these two conflicting goals was ultimately a question of: 'ethics, rather than economics'.

The theoretical framework sketched above had emerged by the early 1930s. But its further development was interrupted by the publication in 1936 of Keynes's *General Theory of Employment, Interest and Money*. Robertson's response to this book was to examine how the General Theory might affect his vision of how the world worked. The central issue for Robertson was whether Keynes had provided a more satisfactory explanation than he had himself of the forces which determined the behaviour of the trend rate of growth of economic activity. The distinguishing feature of Keynes's approach identified by Robertson was in the treatment of the theory of the rate of interest. He interpreted as Keynes's central proposition the contention that there was an inherent tendency for the rate of interest to remain above the level consistent with the maintenance of full employment. And although Robertson was prepared to accept that an argument could, in principle, be made out along such lines he did not accept that Keynes had succeeded in doing so. In particular he maintained that while 'liquidity preference' might make the interest rate 'sticky' in the short period, with its downward movement resistant to monetary expansion, he rejected such an approach to the long-period theory of interest rate determination, summarizing the argument in the following way:

> . . . the rate of interest is what it is because it is expected to become other than it is; if it is not expected to become other than it is, there is nothing left to tell us why it is what it is. The organ which secretes it has been amputated, and yet it somehow still exists – a grin without a cat. ('Mr Keynes and the Rate of Interest' in *Essays in Monetary Theory,* 1940, p. 36)

Keynes's theoretical argument was, therefore, flawed. And the associated case for stabilizing the economy at a level other than that identified in Robertson's own analysis as 'appropriate' was consequently not proven.

The first repercussion of this reaction to the *General Theory* was an estrangement between Robertson and Keynes, virtually ending a close friendship which had lasted for more than 20 years (Robertson having been a student of Keynes, then a fellow teacher and collaborator in research). It then motivated Robertson's decision to leave Cambridge for London in 1938. Moreover, even after Keynes's death in 1946, strained and difficult relations with Keynes's disciples in Cambridge left him a somewhat isolated figure. The impact of Keynes's *General Theory* on Robertson's professional life was no less significant, the whole terrain of the area in which he worked was changed: from being on the creative frontier of the subject he felt himself forced into the role of commentator and critic. In the years after 1936 he wrote almost nothing new in what had been his specialist field. An explanation was provided in a letter to a friendly reviewer of one of his collections of essays who had called upon Robertson to prepare a monograph combining and extending his earlier analytical work, and to whom he wrote:

> . . . I'm afraid there is no chance of my responding to your challenge and trying to produce a full length synthetic Theory of Money or Fluctuations or What-you-will. I'm too old and too lazy! But even if I were younger and less lazy, I think history had made it impossible. I believe that once Keynes had made up his mind to go the way he did it was my particular function to . . . [elucidate and criticise the details of his work] . . . and to go on pegging away at them (as is still necessary). It will not be easy for *anyone* for another twenty years to produce a positive and constructive work which is not in large measure a commentary on Keynes, – that is the measure of his triumph. For me, it would now be psychologically impossible, and the attempt is not worth making. (Private letter of D.H. Robertson to T.J. Wilson, 31 October 1953.

Selected Works

1915. *A study of industrial fluctuation*. London: P.S. King & Son. Reprinted with a new introduction, in *Reprints of scarce works on political economy*. London: London School of Economics and Political Science, 1948.

1922. *Money*, Cambridge economic handbook. London: Nisbet & Co. Revised ed, 1924; new ed, 1948.

1923. *The control of industry*, Cambridge economic handbook. London: Nisbet & Co. Rev. ed, 1928; new ed (with S.R. Dennison), 1960.

1926. *Banking policy and the price level*. London: P.S. King & Son. Reprinted, 1926; reprinted with revisions, 1932; reprinted in the USA with a new preface. New York: Augustus M. Kelley, 1949.

1931a. *Economic fragments*. London: P.S. King & Son.

1931b. (With A.C. Pigou.) *Economic essays and addresses*. London: P.S. King & Son.

1940. *Essays in monetary theory*. London: P.S. King & Son.

1950. *Utility and all that*. London: George Allen & Unwin.

1956. *Economic commentaries*. London: Staples Press.

1957–9. *Lectures on economic principles*, 3 vols. London: Staples Press. Paperback edn in one volume, London: Fontana, 1963.

1960. *Growth, wages, money*, The Marshall lectures at the University of Cambridge. London: Cambridge University Press.

Bibliography

Fletcher, G. 2007. *Dennis Robertson: Essays on his life and work*. Basingstoke: Palgrave Macmillan.

Robinson Crusoe

M. V. White

Keywords

Capital; Commodity exchange; Consumer surplus; Individualism; Marginalist economics; Producer surplus; Rational behaviour; Robinson Crusoe; Supply and demand

JEL Classifications

B31

Written by Daniel Defoe, *Robinson Crusoe* was first published in 1719–20. By the end of the 19th century there were many references made to a Crusoe economy to illustrate the principles of supply and demand economic theory. Crusoe thus became a representative rational economic individual, allocating his available resources to obtain maximum satisfaction in the present or future.

The figure of Crusoe as the personification of supply and demand economic theory can be found in W.S. Jevons's *Theory* (1871), C. Menger's *Principles* (1871), P. Wicksteed's *Alphabet of Economic Science* (1888), E. Böhm-Bawerk's *Theory of Capital* (1890), A. Marshall's *Principles* (1891) and K. Wicksell's *Value, Capital and Rent* (1893). The principal uses of the device were to show how an isolated individual would allocate consumption items so as to maximize utility in a marginalist fashion and distribute labour effort between producing items for consumption or investment (creating 'capital'). Calculations were made according to the relative amounts of pleasure and pain immediately or ultimately involved in the various activities. Marshall also used Crusoe to illustrate producer and consumer surplus, while F.Y. Edgeworth's *Mathematical Psychics* (1881) introduced 'the black', Friday, when discussing issues in the theory of commodity exchange.

The role of a Crusoe economy was not simply to illustrate various components of supply and demand theory. It was also utilized to support the claim that the principles of rational behaviour, as defined by that theory, could be applied to any type of economy – from the isolated individual to 'modern civilization'. This point was made particularly clear in J.B. Clark's *The Distribution of Wealth* (1899). Similar references to a Crusoe economy can be found in textbooks today.

Two general characteristics of an economic Crusoe's actions are important to note. First, he must be able to calculate in a precise fashion making fine decisions between whether to work or rest, to consume or save/invest. Second, he has no resources other than those available in the island environment. Both characteristics mean the economic Crusoe bears no relation to the Crusoe in Defoe's novel. Defoe's Crusoe wastes time because he cannot calculate in a marginalist fashion; he cannot rationally allocate labour time because labour is as useful in one pursuit as another; and he would not have survived without items salvaged from the shipwreck. Other decisions, such as whether to consume or save, also preclude calculation on the basis of relative pleasure and pain (White 1982). Moreover, the relation between Crusoe and Friday is not based on voluntary reciprocal exchanges, as in the supply and demand parable, but rather on power and violence (Hymer 1980). The economic Crusoe could not, therefore, have been produced by relying on the letter of Defoe's text.

It is possible to find some references to Crusoe by English political economists during the 1830s, but these were sporadic and no attempts were made to utilize Crusoe in a systematic fashion. An economic Crusoe thus appears only after mid-century with references in F. Bastiat's *Economic Harmonies* (1850) and H. Gossen's *Entwickelung* (Gossen 1854). These references owed a good deal to the rewriting of Defoe's text within the literary genre of the 'Robinsonade'.

The Robinsonade literature dates from the early 18th century (Gove 1941) and includes voyage or shipwreck narratives, imaginary voyages to 'isolated lands' and more general discussions of colonial settlements which depicted various stages of societal development. The last group of Robinsonade texts bears some resemblance to the four-stage theory of societies produced during the Scottish Enlightenment, remnants of which can be found in the work of the classical political economists (Meek 1976). One such remnant was the illustrative device, used by A. Smith and D. Ricardo, of hunters exchanging commodities according to the labour embodied in them. While Marx was critical of this device, he noted it made sense in the context of the previous century's Robinsonades. However he considered the later discussion of Crusoe by Bastiat for example, was 'twaddle' because it depicted an individual 'outside society' (Marx 1857–8, pp. 83–5).

Bastiat's Crusoe relied on a different type of Robinsonade literature, particularly J.H. Campe's *Robinson the Younger* (1779/80). Campe rewrote Defoe's tale to show Crusoe's survival on the island was not dependent on the shipwreck items. Gossen also appealed to Campe's novel to illustrate his marginalist explanation of work and consumption decisions. By the mid-19th century, then, the 'individualist' Robinsonade was utilized by those theorists who conceptualized the economy as a series of voluntary exchanges, where the principles of economic activity were those of the individual writ large.

English supply and demand economists could also draw upon a discernible shift in the readings of Defoe's text by literary commentators after 1850. Earlier commentary had stressed the novel was useful for showing, especially to children and the 'working classes', the virtue of work and the need to accept the given social organization ordained by Divine Providence. Commentary after mid-century represented Crusoe more as an individual calculating costs and benefits in the manner of an English shop keeper. It was even argued Crusoe represented a ' national ideology' in that regard. The remarkable similarity between this Crusoe and the illustrative device of English supply and demand economic theory suggests the latter was able to appropriate the former as a recognizable referent.

The economic Crusoe served, in effect, as a useful defensive device against 'historical' critics of economic theory such as T.E. Cliffe Leslie and J.K. Ingram. Writing between the mid-1860s and early 1880s, the critics argued that there were no universal laws of economic behaviour since behaviour could change according to the type of society being considered. Supply and demand theorists, such as Jevons, rejected that criticism, claiming historical studies could only confirm the 'universal' laws of behaviour assumed in the theory (Jevons 1876, pp. 196–7). In this context, the economic Crusoe provided an apparently tangible reference point when supply and demand theory began its analysis with the actions of an 'isolated' or representative individual. Indeed, Gossen had used Campe's Crusoe in precisely that fashion when criticizing the German 'National Economists' in 1854 (Gossen 1854, pp. 45–7). The role of an economic Crusoe, as both illustrative and defensive device for supply and demand theory, was thus inscribed from its inception.

See Also

▶ Economic Man
▶ Rational Behaviour

Bibliography

Gossen, H.H. 1854. *Entwickelung der Gesetze des menschlichen Verkehrs und der daraus fliessenden Reglen für menschliches Handeln*. Brunswick: Vieweg.

Gove, P.B. 1941. *The imaginary voyage in prose fiction*. New York: Columbia University Press.

Hymer, S. 1980. Robinson Crusoe and the secret of primitive accumulation. In *Growth, profits and property*, ed. E.J. Nell. New York: Cambridge University Press.

Jevons, W.S. 1876. The future of political economy. In *The principles of economics and other papers*, ed. H. Higgs. London: Macmillan, 1905.

Marx, K. 1857–8. *Grundrisse.* Harmondsworth: Penguin, 1973.

Meek, R.L. 1976. *Social science and the ignoble savage*. Cambridge: Cambridge University Press.

White, M.V. 1982. Reading and rewriting. The production of an economic Robinson Crusoe. *Southern Review* 15(2): 115–142.

Robinson, Abraham (1918–1974)

W. A. J. Luxemburg

A logician, mathematician and applied mathematician, Abraham Robinson was one of the foremost proponents of applying the methods and results of mathematical logic, in particular model theory to mathematics. This point of view led Abraham Robinson around 1960 to the creation of Non-standard Analysis.

Today, under the general term 'Non-standard Analysis' mathematicians understand the study of mathematical structures with the use of their non-standard models. Non-standard models of mathematical structures are proper extensions of such structures that are also models of the properties of the original structure which can be expressed in terms of formulas of a formal language chosen in advance.

Although in a formal sense non-standard models of a mathematical structure share the same properties with the given structure as far as they are expressible in terms of the formulas of a formal language. The main advantage of using them is that such models, being proper extensions, contain entities that may be considered as ideal mathematical objects of the given structure whose potential existence can only be predicted in the original structure. In this sense, non-standard analysis follows very closely the traditional philosophy and method of mathematics of creating new mathematical entities by means of extensions such as, for instance, the real numbers arise from the counting numbers via algebraic extensions.

In the calculus, non-standard models of the real number system are totally ordered field extensions of the field of real numbers of a special nature in that, in a precise formal sense, these field extensions also have the same properties as the real number system. Being proper extensions, however, they contain entities which can be viewed as playing the role of the intuitively introduced infinitely small numbers by the original founders of the calculus. This original discovery of Robinson around 1960 not only generated the development of the use of non-standard models in mathematics and applied mathematics but also solved the three-century-old problem of Gottfield Wilhelm Leibniz of providing a rigorous foundation of the calculus with the use of infinitely small as well as infinitely large numbers.

Abraham Robinson was born in Waldenburg, Lower Silesia, on 16 October 1918. He was the second son of Abraham Robinsohn and Hedwig Lotte Robinsohn, born Bähr. The father was an active zionist and private secretary to David Wolffsohn, leader of the Zionist World Organization, and the family lived in Cologne. He died very shortly before Abraham's birth upon which the family moved to Waldenburg in Lower Silesia to live with Mrs Robinsohn's family. The first seven years of his life were spent in Waldenburg.

In 1925 the Robinsohn family moved to Breslau where Abraham entered a private school headed by Rabbi Simonson. The only brother of his father, Dr Isaac Robinsohn – a prominent physician and head of the Rothschild Hospital in Vienna specializing in radiology, who served as the guardian of the two brothers – had a profound influence on the young Abraham.

After the national elections in Germany on 5 March 1933 which brought Hitler to power, Mrs Robinsohn made plans to leave for Palestine, where the family arrived at Haifa on 9 April of that year. The family settled in Tel-Aviv, where the two brothers entered high school. In 1936, Abraham took up his university studies at the Hebrew University in Jerusalem. The lectures of Abraham Fraenkel introduced him to mathematical logic and set theory. After a few years Fraenkel declared that there was nothing more he could teach his promising student. In 1939, Abraham won a scholarship to the Sorbonne and arrived in France at the start of World War II. In June 1940 he fled to England where he joined the Free French Air Force, after an attempt to join the British Forces, and left it in June 1942 to become a Scientific Officer in the Ministry of Aircraft Production, with an assignment to the Royal Aircraft Establishment at Farnborough.

All through the war years Abraham continued his research in logic and mathematics and published various papers. The work at Farnborough generated his interest in applied mathematics which led to a number of important and lasting contributions to aerodynamics. Also during these years he met his wife Renée Koppel, a refugee from Vienna.

His PhD thesis of 1949, written under the direction of Professor P. Dienes of Birkbeck College, London, was entitled 'The Metamathematics of Algebraic Systems' and constituted a major breakthrough in applying model theory to algebraic structures. In 1951, Robinson became Associate Professor of Applied Mathematics at the University of Toronto, and in 1958 was promoted to Professor and Chairman of the Department of Applied Mathematics. In these years he continued his research not only in aerodynamics but above all in his most cherished field of mathematical logic, model theory. In 1957 he became chairman of the Department of Mathematics of the Hebrew University in Jerusalem.

After five years in Jerusalem he moved to the University of California at Los Angeles, and in the fall of 1967 to Yale University, where he died prematurely at the age of 56 on 11 April 1974. He is buried on a hillside in Har Menuchot Cemetery near Jerusalem.

During his years at Yale, Robinson came into contact with mathematical economics. He discovered soon that the ideas of non-standard analysis could be fruitfully applied to various problems in mathematical economics. Jointly with Donald Brown, he successfully analysed large exchange economies and Edgeworth's conjecture with the use of non-standard analysis.

Selected Works

1979. Selected Papers of Abraham Robinson with a Biography by George B. Seligman. Vol. 1: *Model theory and algebra*; Vol. 2: *Nonstandard analysis and philosophy*; Vol. 3: *Aeronautics*, ed. H.J. Keisler, S. Körner, W.A.J. Luxemburg, and A.D. Young. New Haven/London: Yale University Press.

Robinson, Edward Austin Gossage (1897–1993)

Z. A. Silberston

Keywords

Monopoly; Optimum firm size; Robinson E. A. G.; Robinson J. V.

JEL Classifications

B31

Austin Robinson was educated at Marlborough College and Christ's College, Cambridge. During the First World War he served as a pilot in the RNAS and the RAF. After finishing his studies at Cambridge he became a Fellow of Corpus Christi College, from 1923 to 1926. In 1926 he married Joan, daughter of Major-General Sir Frederick Maurice and later to become the eminent economist. From 1926 to 1928 Robinson was tutor to the Maharaja of Gwalior. He returned to Cambridge as a university lecturer in economics in 1929, and from then on was an important figure on the Cambridge economics scene. He became Professor of Economics in 1950. He retired in 1965 (and it so happened that Joan Robinson was appointed to his chair). After his retirement, he continued to play a prominent role in Cambridge economics, as well as on the national and international scene.

Austin Robinson's first book, *The Structure of Competitive Industry* (1931), established his reputation as an economist. This seminal work drew on Alfred Marshall's writings on industry, and considered in detail the problems involved in determining the optimum size of firm. But although it emphasized the importance of scale, and inspired much of the later empirical work in this area, it also recognized that low British productivity in manufacturing industry was not primarily the consequence of scale, but of attitudes towards work and competition. All subsequent writing on this subject owed a considerable debt to Robinson. He followed up his work on competitive industry with a book on *Monopoly* (1941), as well as with a number of articles, including work on Africa. He was a member of the group surrounding Keynes when he was formulating the *General Theory*, and wrote a review of it in *The Economist*, insisting on signing it (against the traditions of the paper) because of the exceptionally controversial nature of the subject.

Robinson's long association with the *Economic Journal* began in 1934, as Assistant Editor to Keynes, and was later to be followed by much editorial work. Robinson did distinguished service during the war. He was a member of the Economic Section, War Cabinet Office, from 1939 to 1942, and from 1942 to 1945 was Economic Adviser and Head of Programmes Division, Ministry of Production. This was followed by a period as Economic Adviser to the Board of Trade. He returned to Cambridge in 1946, but served a further period in government on the Economic Planning Staff from 1945 to 1947. He was joint editor of the *Economic Journal* from 1944 to 1970, and played a leading role in the profession in other ways, holding a number of important posts, including that of managing editor of the Royal Economic Society's edition of Keynes's works. He was much involved in the work of the new International Economic Association: he was President from 1959 to 1962 and editor of its publications for many years. A good deal of his subsequent writing and editorial work, much of it on the problems of developing countries, was carried out in the context of the work of the IEA.

Austin Robinson's career was a remarkable one. He combined writing, teaching, editorial work and administration with advising governments in both the developed and developing world. He played a leading role in the economics profession for an exceptionally long period, internationally as well as in Britain, and did so throughout with much distinction.

See Also

▶ Royal Economic Society

Selected Works

1931. *The structure of competitive industry*. London/Cambridge: Nisbet & Co./Cambridge University Press.

1941. *Monopoly*. London/Cambridge: Nisbet & Co./Cambridge University Press.

Robinson, Joan Violet (1903–1983)

Luigi L. Pasinetti

Keywords

Animal spirits; Beggar-thy-neighbour; Cambridge circus; Capital accumulation; Capital theory; Disguised unemployment; Effective demand; Employment theory; Euler's theorem; Firm, theory of; Harrod, R. F.; Historical time and logical time; Imperfect competition; Investment; Kahn, R. F.; Kalecki, M.; Keynes, J. M.; Logical time; Luxemburg, R.; Marshall, A.; Marshall–Lerner conditions; Marxist economics; Multiplier; Perfect competition; Price formation; Production function; Reswitching of techniques; Robinson, E. A. G.; Robinson, J. V.; Savings; Sraffa, P.; Two-sector models; Unorthodox economics; Wicksell, J. G. K.

JEL Classifications

B31

Joan Robinson (née Maurice) was born at Camberley, Surrey, on 31 October 1903. She died in Cambridge on 5 August 1983.

She is the only woman (with the possible, but controversial, exception of Rosa Luxemburg) among the great economists. In 1975, which was proclaimed Woman's Year, most economists in the United States expected that she would naturally be chosen for the Nobel Memorial Prize in Economics for that year. She had received triumphant acclaim, as a Special Ely Lecturer, at the

American Economic Association annual meeting three years earlier, in spite of the harsh hostility that her theories had always met in the United States. The American magazine *Business Week*, after sounding out the American economics profession, felt so sure of the choice as to anticipate the event by publishing a long article on her, presenting her explicitly as being ‘on everyone’s list for this year’s Nobel Prize in Economics’. But the Swedish Royal Academy missed that opportunity (and alas, never regained it). Ever since, in shop-talk among economists all over the world, Joan Robinson has become the greatest Nobel Prize winner that never was.

Basic Biography

Joan Robinson was the daughter of Major General Sir Frederick Maurice and of Helen Marsh (who was herself the daughter of a Professor of Surgery and Master of Downing College, Cambridge). Sir Frederick pursued a brilliant career in the British Army, but in 1918 he found himself at the centre of a public debate, and he gave up his army career on a point of principle. This was very much in the family tradition. Sir Frederick’s grandfather – Joan Robinson’s great-grandfather – was Frederick.

Denison Maurice, the Christian Socialist who lost his chair of theology at King’s College London, for his refusal to believe in eternal damnation.

Joan Robinson certainly had many of these traits: toughness and endurance of character, non-conformism and unorthodoxy of views, the absence of any reverential feeling or timidity, even in the face of the world’s celebrities, a passionate longing for the new and the unknown.

She was educated at St Paul’s Girls’ School in London. (Curiously enough, Richard Kahn was educated in the boys’ section of the same school.) In October 1922, she was admitted to the University of Cambridge, going up to Girton College, where she read economics at a time when the dominant figures in Cambridge were Marshall and Pigou. Marshall had retired (he died in 1924) but was extremely influential not only in Cambridge but in the whole of the British Isles. Pigou, his favourite pupil and chosen successor, was the Professor of Political Economy, at whose lectures Cambridge students absorbed the official *verbum* of Marshallian economics. Keynes was a sort of outsider, part-time in Cambridge and part-time in London, always involved with government policies, either at the Treasury or in public opposition. In those days he lectured on strictly orthodox monetary theory and policies. His lectures were not given regularly but were well attended.

The intellectual environment must have appeared solidly traditional. Joan graduated in 1925, as a good girl would: with second class honours.

In the following year (1926), she married E.A.G. Robinson (later Professor Sir Austin Robinson), who was six years her senior and at the time a junior Fellow of Corpus Christi College. Together they left Cambridge and set off for India, where they stayed for two years. Austin Robinson served as tutor of the Maharajah of Gwalior. Joan was there as Austin’s wife but did some teaching at the local school. When they returned, after their two-year Indian engagement, Austin Robinson took a permanent post as Lecturer in Economics at Cambridge, where they settled for life. They had two daughters.

It was on the return to Cambridge from India (summer 1928) that Joan Robinson began to do some College supervision of undergraduates, and then to do economics research in earnest. The Cambridge intellectual environment had changed dramatically. After Edgeworth’s death (1926), Keynes became the sole editor of the *Economic Journal* and was engaged on his *Treatise on Money* (Keynes 1930). Most of all, he had brought to Cambridge Piero Sraffa, the young Italian economist who had dared to launch a scathing attack on Marshallian economics (Sraffa 1926). Moreover, some new stars were rising in the firmament of Keynes’s entourage – Frank Ramsey, the brilliant mathematician; Ludwig Wittgenstein, the Austrian philosopher whom Keynes persuaded to come to Cambridge; and Richard Kahn, Keynes’s favourite pupil. It was with Richard Kahn that Joan Robinson began an intense intellectual partnership that lasted for her whole life.

On a strictly academic level, Joan Robinson slowly ascended the academic ladder: Junior Assistant Lecturer in 1931, Full Lecturer in 1937, Reader in 1949. It was suggested in Cambridge that the fact that her husband was in the same faculty kept her back at all stages of her academic career. She became full professor only on Austin Robinson's retirement, in 1965. Her association with the Cambridge colleges was more irregular. But she was, in succession, a Fellow of Girton College and of Newnham College. Yet whatever the formal position in the faculty or in the Cambridge colleges, she was for years one of the major attractions in Cambridge for many generations of undergraduates, not only in economics. In the post-war period, she was certainly the best-known member of the Cambridge economics faculty abroad. An indefatigable traveller, she did not limit her foreign visits to universities; she also wanted to know local customs and local conditions of life, even far away from urban centres. Her strong constitution and temperamental toughness helped her enormously. A friend from Makerere University, who took her, when she was already in her seventies, on a month's travel in tribal Africa was amazed at how much she could endure in terms of living in most primitive conditions with raw food, lack of facilities and exposure to harsh tropical weather, day and night.

It would be impossible to list here all the places she visited or the talks, seminars and public lectures she gave all over the world. She rarely stayed in Cambridge during the summer or term vacations or during her sabbatical years, though punctually and punctiliously returning there on the eve of the terms of her teaching. Asia was her favourite continent (especially India and China). But hundreds of students in North and South America, Australia, Africa and Europe also knew her at first hand.

In Cambridge she rarely missed her classes, lectures and seminars and she was a regular attendant of other people's seminars, especially visitors', never avoiding discussion and confrontation. Professor Pigou – a well-known misogynist – had included her in his category of 'honorary men'.

She was extremely popular with the students – a clear, brilliant, stimulating teacher. She was a person who inspired strong feelings – of love and hate. Her opponents were frightened by her, and her friends really admired, almost worshipped her. Her nonconformism in everyday life and even in her clothing (most of which she bought in India) was renowned.

She retired from her professorship in Cambridge on 30 September 1971. On retirement she did not agree to continue lecturing in Cambridge. (Later on, in the late 1970s, she gave in partially, giving a course of lectures on 'the Cambridge tradition'.) But her writing and lecturing abroad, at the invitation of economics faculties and students all over the world, continued unabated.

When, in the late 1970s, King's College (Keynes's college) finally dropped the traditional anachronistic ban on women and became co-educational, Joan Robinson, upon an enthusiastic and unanimous proposal by all economists of the college, became the first woman to be made an Honorary Fellow of King's College. (She had earlier become an Honorary Fellow of Girton College and of Newnham College.)

Towards the end of her life, she became very concerned and disappointed with the direction in which economic theory had turned and with the ease with which the younger economists could bend their elegant models to suit the new conservative moods and the selfish economic policies of politicians and governments. Her friends also noticed a sort of stiffening rigidity in her views that had not appeared before. This was unfortunate, as it contributed to increasing the hostility of her opponents towards her.

She suffered a stroke in early February 1983, from which she never recovered. She lay for a few months in a Cambridge hospital, and died peacefully six months later.

Distinctive Traits of Her Intellectual Personality

In order to understand better the nature of Joan Robinson's contributions to economic theory, it may be helpful to begin by considering explicitly a few characteristic traits of her intellectual personality.

Joan Robinson had a remarkable analytical ability. Since she did not normally use mathematics, this remarkable intellectual ability was of a nature that defies conventional description. In her early works she made use of geometrical representations, backed up by calculus (normally provided by Richard Kahn). In her mature works, her way of reasoning took up a more personal feature. Her style is difficult to imitate (as when she invites the readers to follow her in the construction of economic exercises) but very effective. The results are always impressive. Those who used to argue with her knew that she could grasp and keep in the back of her mind (to be brought out at the appropriate moment) a whole series of chain effects and interdependences which her interlocutors could hardly imagine.

She was not the type of person who could go on thinking in isolation. The way she could best express herself was by having somebody in constant confrontation.

She could put her views best either in opposition or in support of somebody else's position. This made her extraordinarily open to concepts and contributions coming from the people she encountered. The accurate historian of economic ideas will probably find in her works traces of almost every person she met. It is therefore important, in considering Joan Robinson's contributions, to keep in mind at least the most important economists who influenced her. These include her teachers (Marshall through Pigou, Keynes, Shove), her contemporaries (Sraffa, Kaldor, and Kalecki, through whom she went back to Marx, but especially Richard Kahn, who read, criticized and improved every single one of her works) and also a whole series of other (younger) people – pupils and students.

This raises the question of her originality. The prefaces to her books are packed with acknowledgements, sometimes heavy acknowledgements – consider, for example, the following excerpt from the *Economics of Imperfect Competition:*

> . . . this book contains some matter which I believe to be new. Of not all the new ideas, however, can I definitely say that 'this is my own invention'. I particularly have had the constant assistance of Mr R.F. Kahn . . . many of the major problems . . . were solved as much by him as by me. (Robinson 1933, p. v)

But one must remember what has been said above. In fact, Joan Robinson was a highly original thinker, but of a particular type. Besides the contributions to economic theory that are distinctly hers she had her own highly original way, even in small details, of presenting other authors' views, which she always did through a distinctly personal re-elaboration. Sometimes the re-elaboration is so personal as to sound parochial. But this trait is not exclusive to Joan Robinson. Cambridge parochialism is shared by almost all purely Cambridge-bred economists since Marshall (Keynes included). It sometimes creates unnecessary difficulties of communication with economists outside Cambridge (that is, with the overwhelming majority!) or introduces a few odd notes into an otherwise impeccable performance.

One can clearly detect an evolution in Joan Robinson's approach to economics that with age strengthened her innovative tendencies. It looks as if she was very cautious in her early years, preoccupied at first with building up solid analytical foundations. But as soon as she felt sure of her analytical equipment, she began to venture more and more into the exciting field of innovation. In her mature works her typical style became established. A sort of mixture of educational, temperamental and intellectual factors made her one of the leading unorthodox economists of the 20th century. Always impatient with dogmas, constantly fighting for new unorthodox ideas, relentlessly attacking established beliefs, she acquired a sort of vocation to economic heresies (see Robinson 1971). Her attitude reminds one of a dictum by Pietro Pomponazzi, the Italian Renaissance philosopher: 'It is better to be a heretic if one wishes to find the truth.'

Strongly related to this attitude is the social message that comes from her writings. Her 'box of tools' and her logical chain of arguments were not proposed for their own sake; they were always aimed at practical action, with a view to the world's most pressing problems – unemployment before the war, underdevelopment and the struggle of ex-colonial nations after the war (very noticeable is her special concern for Asia and her enthusiasm, at points rather naive, for Communist China). Consistently, she has been among the strongest assertors – second perhaps only to Gunnar Myrdal – of the non-neutrality of economic science and of the necessity of stating explicitly one's convictions and beliefs.

And yet, in spite of her bold attacks and her satirical mood, her literary style is surprisingly feminine – rich with fable-like parables, with down-to-earth examples from everyday life ('the price of a cup of tea . . .') and with similes from scenes and examples taken from nature (the *Accumulation of Capital* begins with the economic life of the robin). Her sparkling prose and her entertaining asides make Joan Robinson one of the most brilliant writers among economists and certainly one of the most enjoyable and delightful to read.

Her Scientific Achievements

Joan Robinson wrote numerous books and an enormous number of articles, most of which have been collected in her *Collected Economic Papers* (1951–79).

They fall neatly into three broad groups, corresponding to the three basic phases of her intellectual development. A first group belongs to the phase of her by now classic *Economics of Imperfect Competition* (1933). A second group belongs to the phase of explanation, propagation and defence of Keynes's *General Theory*. Finally, a third group of writings grew around the major work of her maturity, *The Accumulation of Capital* (1956). Other books and articles have originated from miscellaneous or wider interests or from the desire to provide students with economics exercises or with a non-orthodox economics textbook (Robinson and Eatwell 1973c). Altogether, they make an impressive list. Even neglecting her articles (most of which are reprinted in the books), her bibliography contains no fewer than 24 books.

The most widely known of Joan Robinson's works is still the first, *The Economics of Imperfect Competition* (1933). It was the book of her youth, which placed her immediately in the forefront of the development of economic theory. It is a work conceived in Cambridge, at the end of a decade characterized by an intense controversy on cost curves and the laws of returns (see Sraffa 1926, and the Symposium on the 'laws of returns' by Robertson et al. 1930). With this controversy in the background, Joan Robinson's book emerges in 1933 as a masterpiece in the traditional sense of the word. The restrictive conditions of perfect competition on which Marshall's theory was constructed are abandoned, and perfect competition is shown to be a very special case of what in general is a monopolistic situation. A whole new analysis of market behaviour is carried out on new, more general, assumptions; and yet the whole method of analysis, the whole approach – though refined and perfected – is still the traditional Marshallian one. Sraffa's criticism of the master is accepted, but is incorporated into the traditional fold by a generalization of Marshall's own theoretical framework. The outcome is extremely elegant and impressive. The whole matter of market competition is clarified. Marshall's ambiguities are eliminated, the various market conditions are rigorously defined, a whole technical apparatus (a 'box of analytical tools') is developed to deal with market situations in the general case (from demand and supply curves to marginal cost and marginal revenue curves). In a sense, therefore, rather than a radical critique, the *Economics of Imperfect Competition* might well be regarded as the completion and coronation of Marshallian analysis. This may help to explain why Joan Robinson herself came to like that book less and less, as her thought later developed on different lines. In 1969 she came to the point of writing a harsh eight-page criticism of it. Very courageously she published it, on the occasion of a reprint of the book, as a Preface to the second edition!

The book had appeared almost simultaneously with the *Theory of Monopolistic Competition* by Edward Chamberlin (1933); and the two books are normally bracketed together as indicating the decisive breakaway of economic theory from the assumptions of perfect competition. Chamberlin

R

always complained about this association. For although the two books represent the simultaneous discovery of basically the same thing, made quite independently by two different authors, they are in fact substantively different.

It may also be added that, looked at in retrospect, these two books do not appear so conclusive a contribution to the theory of the firm as they appeared to be in the 1930s. The behaviour of firms in oligopolistic markets and the policies of the large corporations have turned out to require more complicated analysis. At the same time, the assumption of perfect competition, far from being completely dead, has recently come back in different guises in the works of many theoretical economists. Yet there is no doubt that the two books remain there to represent a definite turning-point in the development of the theory of the firm – so much so as to be referred to as representing the 'monopolistic competition revolution' (Samuelson 1967). Very characteristically, Edward Chamberlin, after writing the *Theory of Monopolistic Competition*, spent the whole of his life in refining, completing and adding appendices to his masterpiece (no fewer than eight editions!). For Joan Robinson, the *Economics of Imperfect Competition* was only the first step on a very long way to a series of works in quite different and varied fields of economic theory.

It should be added that the *Economics of Imperfect Competition* was not Joan Robinson's only contribution to microeconomic theory in the 1930s. Her name appears again and again on the pages of the avant-garde economic journals of the time. From among her papers, explicit mention must be made at least of her remarkably lucid article on 'rising supply price' and of her contribution to clarifying the meaning of Euler's theorem as applied to marginal productivities, in the traditional theory of production (see her *Collected Papers*, vol. 1).

But something of extraordinary importance was happening in Cambridge in the 1930s. Keynes was in the process of producing his revolutionary work (Keynes 1936). Joan Robinson abandoned the theory of the firm and threw herself selflessly and entirely into the new paths opened up by him. This was a really brave decision, if one thinks that her first book had gained her great reputation in the economic profession. Very rarely do we find someone who, after striking success and becoming a leading figure in a certain field, pulls out of it and puts himself or herself into the shadow of someone else, be this someone else even of the stature of Keynes. Joan Robinson did precisely that. She was one of the members – actually an important member, as is revealed by the recent publication of her correspondence with Keynes (see Keynes 1973, 1979) – of that group of young economists known as the 'Cambridge Circus' (and including Kahn, Sraffa, Harrod, Meade, besides Austin and Joan Robinson) who regularly met for discussion, and played a crucial role in the evolving drafts of Keynes's *General Theory.*

It must be said that the new Keynes's ways were more congenial to her temperament. They were a break with tradition and this suited her nonconformist attitude; they dealt with the deep social problems of unemployment and this appealed to her social conscience. It is in this vein that she published her *Essays in the Theory of Employment* (1937a) and her *Introduction to the Theory of Employment* (1937b). These twin books were simply meant to be a help to the readers of Keynes's *General Theory.* In fact, they turned out to be much more than that. In particular, Joan Robinson contributes to the clarification of a major piece of Keynesian theory – the process through which investments determine savings – which had remained rather obscure from the *General Theory.* For her, this appeared important because it broke a crucial link in traditional theory, which presented the rate of interest as a compensation for the 'sacrifice' of supplying capital (that is, for saving). Joan Robinson stresses the role of investment as an independent variable, while total saving is shown as being determined passively by investment through the operation of the multiplier; the conclusion being that the rate of interest cannot be remunerating anybody's 'sacrifice'. Even more so in depression times, when thrift – a 'private virtue' – becomes a 'public vice'. Other concepts, introduced by Joan Robinson at the time, that were to remain permanently in the following economic literature on the theory of employment are those concerning what she called 'beggar-my-neighbour' policies, 'disguised unemployment' and the generalization of the Marshall–Lerner conditions on international trade, in terms of 'the four elasticities'.

Towards the end of the 1930s, Joan Robinson met Kalecki, and discovered that quite independently of, and in fact earlier than, Keynes he had come to the same conclusions. Kalecki had started from a Marxist background, against which Keynes was prejudiced. This led her to re-reading Marx and to re-thinking her own position vis-à-vis Marxian theory (Robinson 1942).

Joan Robinson's flirtation with Marx is very curious. It has all the charm of a meeting and all the clamour of a clash. She is no doubt attracted by Marx's general conception of society. She finds in Marx much that she approves of. But she finds his scientific nucleus embedded in, and in need of being liberated from, ideology. To obtain this, she says, one must work hard. Her writings on Marx are specifically aimed at 'separating the wheat of science from the chaff of ideology'. Needless to say, this has caused her a lot of trouble with the Marxists. It should be kept in mind that in Continental Europe discussions on Marx have a long and complex tradition of philological heaviness and ideological passion. Joan Robinson's discussion is short and simple. She is always looking at Marx as 'a serious economist'. Accordingly, she always tries to go straight to what she thinks is his economic analysis. Her insistence on the necessity of rescuing Marx, as a scholar and a first-rate analytical mind, has recently been vindicated, especially after the publication of Sraffa's book (1960; see also, for example, Samuelson 1971).

But the post-war period was opening up new vistas. With Keynes's *General Theory* in the background, Joan Robinson saw a formidable task ahead, consisting in nothing less than a reconstruction of economic theory. This led, after a decade of intense work, to the publication of her second major contribution to economic theory – *The Accumulation of Capital* (1956), the work of her maturity and the one that expresses Joan Robinson's genius at her best. Here she has chosen to move on new and controversial ground. While in her first book the direction – once established – was clear and she had to fill in the details, here the direction itself is not entirely clear and has to be continually adjusted. The details acquire less importance and may well be abandoned altogether and replaced with others at a second attempt. As a consequence, a lot of re-writing had to be done.

The Rate of Interest and Other Essays (1952), with its central essay devoted to a 'Generalization of the *General Theory*', turned out to be a sort of preparation. *The Accumulation of Capital* represents the central nucleus of what she perceived as a new framework for economic theory. Then the *Exercises in Economic Analysis* (1960a), the *Essays in the Theory of Economic Growth* (1962a) and a series of other articles fill in the gaps, clarify obscurities, and take the arguments further.

The 'Generalization of the *General Theory*' represents Joan Robinson's response to an interchange with Harrod, following Harrod's *Towards a Dynamic Economics* (1948) and also his earlier review of her *Essays in the Theory of Employment* (1937a). Joan Robinson breaks away from the limitations of the short run, but has not yet defined clearly her direction. Yet, once the process of 'generalization', that is, 'dynamization' of the *General Theory* is started, the author is compelled to recast the Keynesian arguments in terms of the more fundamental categories of capital accumulation, labour supply, technical progress and natural resources. Through this recasting, it became inevitable that she should go to the earlier methodological approach (common to Ricardo and Marx) of stating the problems in terms of social aggregates. The evidence of her intense searching may be found at the end of the book in a chapter of 'acknowledgements and disclaimers', where she describes in succession the way she has been influenced by, or has reacted to, Marx, Marshall, Rosa Luxemburg, Kalecki and Harrod.

The years of transition from the *Rate of Interest and Other Essays* (1952) to the *Accumulation of Capital* (1956) had been marked by a series of intense discussions in Cambridge, especially with Kahn, Sraffa, Kaldor and Champernowne. In the end, Joan Robinson emerged centring her attention on the problem of capital accumulation as the basic process in the development of a capitalist economy. She began with a scathing attack on the traditional concept of 'production function' (in a well-known article, now in her *Collected Papers*, vol. 2, which elicited a chain of angry responses: see, for example, Solow 1955–1956, and Swan 1956). Then she patiently proceeded to a reconstruction. A crucial step was her own way of rediscovering the Swedish economist Knut Wicksell.

The Accumulation of Capital (1956) bears the same title as Rosa Luxemburg's book, to whose translation into English Joan Robinson wrote an introduction (Luxemburg 1951). This was a great tribute to another woman economist. But we should not be misled. Joan Robinson's book belongs to an entirely different age and takes an entirely different approach. Set into a Keynesian framework extended to the long run, it takes its origin from a welding together of Harrod's economic dynamics and of Wicksell's capital theory. The main question Joan Robinson poses to herself is by now a typically classical one: what are the conditions for the achievement of a cumulative long-term growth of income and capital (what she characteristically christened a 'golden age'); and what is the outcome of this process, in terms of growth of gross and net output and of the distribution of income between wages and profits, given a certain evolution through time of the labour force and of technology? To answer these question Joan Robinson builds up a two-sector dynamic model with a finite number of techniques; and goes on to show the interactions of the relations between wages and profits, the stock of capital and the techniques of production, entrepreneurial expectations and the degree of competition in the economy, bringing in the effects of higher degrees of mechanization and both 'neutral' and 'biased' technical progress. The basic model and the basic answers are all worked out very quickly in the book. The rest is then devoted to relaxing the simplifying assumptions. The whole analysis is carried out *without* the use of mathematics. This is remarkable. Joan Robinson squeezes out of the model, one by one, all the answers that are needed. The non-use of mathematics has certain obvious disadvantages. Though the analysis need not necessarily be any less rigorous, in many passages it is not so easy to follow. It has, however, some advantages, which Joan Robinson is very ready and able to exploit. She succeeds, for example, in freeing herself from the symmetry that a mathematically formulated model normally imposes. In Joan Robinson's model, certain results are always more likely to happen than their symmetrical counterpart. Symmetry and formal elegance play no part; only relevance does, or at least it does in the way perceived by the author.

The overall result is, again, impressive. The oversimplified dynamic model of Harrod is enormously enriched by the introduction of the choice among a finite number of alternative techniques. At the same time the Wicksellian analysis of accumulation at a given technology is completed by the new analysis of a constant flow of inventions of various types. And this marriage of Harrod's model to Wicksellian analysis is made to fructify in a number of directions. So many and so rich are in fact these directions that Joan Robinson herself did not pursue all of them, as became evident from the abundant literature that followed.

To this literature, Joan Robinson contributed a whole series of essays and books (see her *Collected Papers,* vols. 2–5, and J. Robinson 1960b, 1962a), which represent clarifications and further elaborations. They also represent her way of recasting and adjusting her arguments in response to opposition from her critics and to comments, remarks and stimuli of any sort from her friends, as well as her way of coming to grips with results – not always or not entirely compatible with hers – coming from the works of other scholars, colleagues and pupils, who were broadly working on similar problems and with the same aims.

Meanwhile, proceeding on parallel lines, many other separate strands of thinking were emerging from her remarkable intellectual activity. At least a few must briefly be mentioned here.

First, a whole series of concepts and ideas were coming to fruition, which – though not belonging to her major fields of interest – came to complete her overall coverage of economic theory: writings on the theory of international trade (including her professorial inaugural lecture at Cambridge on *The New Mercantilism,* 1966a), on Marxian economics (at various stages in her career), and on the theory of economic development and planning, reproducing her lectures delivered during her world travels or coming from calm reflection, once she had returned home (see her *Collected Papers,* and also J. Robinson 1970b, 1979b).

Second, her deeply felt concern with economics students and economics teaching in general gave origin to books, such as Joan Robinson (1966b, 1971) and especially (with Eatwell) (1973c), which contributed to giving substance to, and disseminating all over the world, her strongly felt conviction that an overall approach to economic reality, alternative to that of traditional economics, does exist and is viable.

Third, the ideas, reflections, rationalizations, accumulated in the course of her life took the form of books such as *Economic Philosophy* (1962b) and *Freedom and Necessity* (1970a), which were concerned with wider issues than economics itself, attempting to give an overall conception of the world and a whole philosophy of life. These writings contribute, not marginally, to place Joan Robinson among the influential thinkers of this century. At the same time, they may well be enjoyed, by the general reader, even more than her masterpieces. From a purely literary point of view, they make delightful reading.

It should be added that there are, moreover, many themes which, while not being exclusively connected with any specific work of Joan Robinson's, recur time and again in her writings, so as to have become characteristically associated with her approach. Here are a few: (*a*) the concept of 'entrepreneurs' animal spirits' – an expression picked up from Keynes and developed as an important element contributing to explain investment in capitalist economies; (*b*) the conviction that Marshall's notions of prices and rate of profit, with reference to industry, are much more akin to Ricardo's notions than to Walras's; (*c*) a sharp distinction between 'logical' time and 'historical' time, both of which have a place in economic analysis but with different roles. On this point Joan Robinson's characterization of the evolution of an economy in historical time as concerning decisions to be taken between 'an irrevocable past and an uncertain future' is well known; (*d*) an equally sharp distinction between *comparisons* of equilibrium–growth positions and *movements* from one equilibrium-growth position to another, in dynamic analysis; (*e*) a tendency, especially in the later part of her life, to shift nearer and nearer to the positions of Kalecki, as opposed to those of Keynes, in interpreting the overall working of the institutions of capitalist economies, especially with reference to what she found as a more satisfactory integration in Kalecki of the concept of effective demand with the process of price formation.

Finally, one must mention specifically an issue which may well continue to give rise to controversial evaluations. This concerns the role that may be assigned to Joan Robinson in the well-known controversy on capital theory that flared up between the two Cambridges in the 1960s (see Pasinetti et al. 1966). One view on this issue is that Joan Robinson had the merit of anticipating the controversy by her (already mentioned) attacks on the neoclassical production function in the mid-1950s (see Harcourt 1972). Another view is that Joan Robinson, herself a victim of her emotional temperament, started her attacks on the traditional concepts too early and misplaced the whole criticism, by neglecting the really basic point (the phenomenon of reswitching of techniques; see Sraffa 1960) that in the end delivered the fatal blow to the neoclassical notion of production function. What one can say for certain is that a hint at the reswitching phenomenon does appear in the *Accumulation of Capital,* but is relegated to the role of a curiosum, in an entirely secondary section. Perhaps the phenomenon had been pointed out to her but she grossly underestimated its importance. What is curious is that she continued to underestimate it even after it was brought to the foreground (see her 'Unimportance of Re-switching' in *Collected Papers*, vol. 5).

But at this point the works of Joan Robinson merge into those of that remarkable group of Cambridge economists – notably, Piero Sraffa, Nicholas Kaldor and Richard Kahn, among others, besides Joan Robinson herself (on this, see the Preface to Pasinetti 1981) – who happened to be concentrated in Cambridge in the post-war period and who took up, continued and expanded the challenge that Keynes had launched on orthodox economic theory. This remarkable group of economists started a stream of economic thought which is obviously far from complete. Its basic features, however, are clear enough; they embody a determined effort to shift the whole focus of economic theorizing away from the problems of optimum allocation of given resources, where it had remained for almost a century, and move it towards the fundamental factors responsible for the dynamics of industrial societies. This shift of focus inevitably brings into the foreground the once central themes of capital accumulation, population growth, production expansion, income distribution, and thus technical progress and structural change.

It is perhaps too early to try to evaluate the relative role played by Joan Robinson as a member of this remarkable group of economists. The single components of the group have made contributions which are sometimes complementary, at other times overlapping, and at yet other times even partly contradictory. To mention only one major problem, Piero Sraffa's book appeared too late for Joan Robinson to be able to incorporate it into her theoretical framework; and the brave efforts she later made to this effect are not always convincing. They actually reveal here and there a sort of ambivalent attitude. At the same time, her *Accumulation of Capital* ventures into fields of economic dynamics which Sraffa does not touch at all. Quite obviously, the common fundamental thrust behind post-Keynesian analysis does not presuppose complete identity of views or complete harmony of approach.

Future developments will clarify issues and will reveal which of the lines of approach proposed are the most useful, fruitful or fecund. There can be little doubt, however, that if this theoretical movement is going to prove successful, quite a lot of rewriting will have to be done in economic theory. If, and when, this rewriting occurs, Joan Robinson's contributions are going to take a major place.

Selected Works

1933. *The economics of imperfect competition*. London: Macmillan. 2nd ed., 1969.

1937a. *Essays in the theory of employment*. London: Macmillan.

1937b. *Introduction to the theory of employment*. London: Macmillan.

1942. *An essay on Marxian economics*. London: Macmillan.

1951. *Collected economic papers,* vol. 1. Oxford: Basil Blackwell. (vol. 2, 1960b; vol. 3, 1965; vol. 4, 1973a; vol. 5, 1979a).

1952. *The rate of interest and other essays*. London: Macmillan.

1956. *The accumulation of capital*. London: Macmillan.

1960a. *Exercises in economic analysis*. London: Macmillan.

1960b. *Collected economic papers*. vol. 2. Oxford: Basil Blackwell.

1962a. *Essays in the theory of economic growth*. London: Macmillan.

1962b. *Economic philosophy*. London: C.A. Watts.

1965. *Collected economic papers*. vol. 3. Oxford: Basil Blackwell.

1966a. *The new mercantilism – An inaugural lecture*. Cambridge: Cambridge University Press.

1966b. *Economics – An awkward corner*. London: Allen & Unwin.

1970a. *Freedom and necessity*. London: Allen & Unwin.

1970b. *The cultural revolution in China*. London: Penguin Books.

1971. *Economic heresies: Some old-fashioned questions in economic theory*. London: Macmillan.

1973a. *Collected economic papers*. vol. 4. Oxford: Basil Blackwell.

1973b. ed. *After Keynes*. Papers presented to Section F (economics) of the 1972 annual meeting of the British Association for Advancement of Science. Oxford: Basil Blackwell.

1973c. (With J. Eatwell.) *An introduction to modern economics*. New York: McGraw-Hill.

1978. *Contributions to modern economics*. Oxford: Basil Blackwell.

1979a. *Collected economic papers*. vol. 5. Oxford: Basil Blackwell.
1979b. *Aspects of development and underdevelopment*. Cambridge: Cambridge University Press.
1980. *Further contributions to modern economics*. Oxford: Basil Blackwell.

Bibliography

Chamberlin, E. 1933. *The theory of monopolistic competition*. Cambridge, MA: Harvard University Press.
Harcourt, G. 1972. *Some Cambridge controversies in the theory of capital*. Cambridge: Cambridge University Press.
Harrod, R.F. 1948. *Towards a dynamic economics*. London: Macmillan.
Keynes, J.M. 1930. *A treatise on money.* 2 vols. London: Macmillan.
Keynes, J.M. 1936. *The general theory of employment, interest and money.* London: Macmillan.
Keynes, J.M. 1973, 1979. *The collected writings of John Maynard Keynes*, vols 13 and 14 (1973) and 29 (1979), ed. D.C. Moggridge. London: Macmillan for the Royal Economic Society.
Luxemburg, R. 1913. *The accumulation of capital*. Trans. A. Schwarzschild, with an Introduction by J. Robinson, London: Routledge & Kegan Paul, 1951.
Pasinetti, L.L. 1974. *Growth and income distribution: Essays in economic theory.* Cambridge: Cambridge University Press.
Pasinetti, L.L. 1981. *Structural change and economic growth: A theoretical essay on the dynamics of the wealth of nations*. Cambridge: Cambridge University Press.
Pasinetti, L.L., D. Levhari, P.A. Samuelson, M. Bruno, E. Burmeister, E. Sheshinski, M. Morishima, and P. Garegnani. 1966. Contributions to 'Paradoxes in Capital Theory – A Symposium'. *Quarterly Journal of Economics* 80: 503–583.
Robertson, D.H., P. Sraffa, and G. Shove. 1930. Increasing returns and the representative firm: A symposium. *Economic Journal* 40: 76–116.
Samuelson, P.A. 1967. The monopolistic competition revolution. In *Monopolistic competition: Studies in impact*, ed. R.M. Kuenne. New York: Wiley.
Samuelson, P.A. 1971. Understanding the Marxian notion of exploitation: A summary of the so-called transformation problem between Marxian values and competitive prices. *Journal of Economic Literature* 9: 339–431.
Solow, R.M. 1955–1956. The production function and the theory of capital. *Review of Economic Studies* 23 (2): 101–108.
Sraffa, P. 1926. The laws of returns under competitive conditions. *Economic Journal* 36: 535–550.
Sraffa, P. 1960. *Production of commodities by means of commodities: Prelude to a critique of economic theory.* Cambridge: Cambridge University Press.
Swan, T.W. 1956. Economic growth and capital accumulation. *The Economic Record* 32: 334–361.
Wicksell, K. 1901. In *Lectures on political economy*, ed. L. Robbins, vol. 1. London: Routledge & Kegan Paul. 1934.

Bibliographic Addendum

Essay collections on Robinson's contributions to economics include B. Gibson, ed., *Joan Robinson's Economics: A Centennial Celebration*, Cheltenham: Edward Elgar, 2005, and G.C. Harcourt, ed., *Joan Robinson: Critical Assessments of Leading Economists*, London: Routledge, 2002. A combination of biography and bibliography of Robinson's work is J. Cicarelli and J. Cicarelli, *Joan Robinson: A Bio-Bibliography,* Westport, CT: Greenwood Press, 1996.

Robust Control

Noah Williams

Abstract

Robust control is an approach for confronting model uncertainty in decision making, aiming at finding decision rules which perform well across a range of alternative models. This typically leads to a minimax approach, where the robust decision rule minimizes the worst-case outcome from the possible set. This article discusses the rationale for robust decisions, the background literature in control theory, and different approaches which have been used in economics, including the most prominent approach due to Hansen and Sargent.

Keywords

Ambiguity; Ambiguity aversion; Control theory; Error modelling; Kalman filtering; Lagrange multipliers; Linear quadratic control; Max-min expected utility; Minimax; Model uncertainty; Optimal control; Perturbation; Probability distribution; Risk aversion; Robust control; Uncertainty aversion; Unstructured uncertainty

R

JEL Classification

D4; D10

Robust control considers the design of decision or control rules that fare well across a range of alternative models. Thus robust control is inherently about model uncertainty, particularly focusing on the implications of model uncertainty for decisions. Robust control originated in the 1980s in the control theory branch of the engineering and applied mathematics literature, and it is now perhaps the dominant approach in control theory. Robust control gained a foothold in economics in the late 1990s and has seen increasing numbers of economic applications in the past few years. (For related surveys see Hansen and Sargent 2001; and Backus et al. 2005. For a more comprehensive view of the leading approach to robust control in economics, see Hansen and Sargent 2008.)

The basic issues in robust control arise from adding more detail to the opening sentence above – that a decision rule performs well across alternative models. To begin, define a model as a specification of a probability distribution over outcomes of interest to the decision maker, which is influenced by a decision or control variable. Then model uncertainty simply means that the decision maker faces subjective uncertainty about the specification of this probability distribution. A first key issue in robust control, then, is to specify the class of alternative models which the decision maker entertains. As we discuss below, there are many approaches to doing so, with the most common cases taking a benchmark *nominal model* as a starting point and considering perturbations of this model. How to specify and measure the magnitude of the perturbations are key practical considerations.

With the model set specified, the next issue is how to choose a decision rule and thus what it means for a rule to 'perform well' across models. In Bayesian analysis, the decision maker forms a prior over models and proceeds as usual to maximize expected utility (or minimize expected loss). Just as we defined a model as a probability distribution, a Bayesian views model uncertainty as simply a hierarchical probability distribution with one layer consisting of shocks and variables to be integrated over, and another layer averaging over models. In contrast, most robust control applications focus on minimizing the worst case loss over the set of possible models (a minimax problem in terms of losses, or max-min expected utility). Stochastic robust control problems thus distinguish sharply between shocks which are averaged over and models which are not. The robust control approach thus presumes that decision makers are either unable or unwilling to form a prior over the forms of model misspecification. Of course, decision makers must be able to specify the set of models as discussed above, but typically this involves bounding the set of possibilities in some way rather than fully specifying each alternative. Finally, there are some approaches which seek a middle ground between the average case and the worst case, for example by maximizing expected utility subject to a bound on the worst case loss. These have been less prominent both in control theory (Limebeer et al. 1994, is one example) and in economics (Tornell 2003, is one exception), and thus will not be discussed further. For the remainder of the article robust control will mean a minimax approach.

Robustness and Worst Case Analysis

Broadly speaking, the control theory literature has adopted the worst-case philosophy out of concerns for stability. A basic desideratum for robust control in practice is that the system remain stable in the face of perturbations, and since instability may be equated with infinite loss, minimizing the worst case outcomes will insure stability (when possible). Moreover, many engineering applications have specific performance objectives which must be maintained, and a cost function penalizing deviations is not clearly specified. However, in dealing with economic agents rather than controlled machines, decision theoretic criteria naturally come into play. In this sphere, robust control is closely related to the notions of Knightian uncertainty, ambiguity and uncertainty aversion, which are all roughly equivalent (although sometimes differing in formalization).

Starting with the observations of the classic Ellsberg (1961) paradox – that (some) decision makers prefer environments with known odds to those with uncertain probabilities – there has been a broad literature in decision theory which has weakened the Savage axioms to incorporate preferences which display such aversion to uncertainty or ambiguity. The most widely used characterization is due to Gilboa and Schmeidler (1989), who axiomatized ambiguity preferences with multiple priors. Decision-making with multiple priors can be represented as max-min expected utility: maximizing the utility with respect to the least favourable prior from a convex set of priors. More recently, Epstein and Schneider (2003) have extended the static environment of Gilboa and Schmeidler to a dynamic context, where the set of priors is updated over time. Hansen et al. (2006) formally established the links between robust control and ambiguity aversion, showing that the model set of robust control as discussed above can be thought of as a particular specification of Gilboa and Schmeidler's set of priors. Moreover, although the ambiguity preferences are characterized by posing particular counterfactuals which require multiple priors, once the least favourable prior is chosen, behaviour could be rationalized as Bayesian with that prior. Thus from a Bayesian viewpoint Sims (2001) views robust control as a means of generating priors, which then naturally leads to questioning whether the worst case prior accurately reflects actual beliefs and preferences. (See also Svensson 2001. Hansen et al. 2006, show how to back out the Bayesian prior which rationalizes robust decision-making.)

Finally, in many cases robust or ambiguity-averse preferences are similar to enhanced risk aversion, and in some cases they are observationally equivalent. This insight dates to Jacobson (1973) and Whittle (1981) in the control theory literature, and the relations between robust control and a particular specification of Kreps and Porteus (1978), Epstein and Zin (1989) and Duffie and Epstein (1992) recursive utility with enhanced risk aversion have been shown by Anderson et al. (2003), Hansen et al. (2006) and Skiadas (2003).

Control Theory Background

Since many of the ideas and inspiration for robust control in economics come from control theory, we give here just a broad outline of its development. More detail and different perspectives can be found in the books by Zhou et al. (1996), Başar and Bernhard (1995), and Burl (1999). Throughout the late 1960s and early 1970s optimal control came into its own, largely through the work of Kalman on linear quadratic (LQ) control and filtering. While this approach remains widely used today throughout economics, starting in the late 1970s and early 1980s the control theory literature started to change as theory and practice showed some of the limitations of the LQ approach. Although LQ control with full observation (the so-called linear quadratic regulator or LQR) was known to be robust to some types of model perturbations, Doyle (1978) showed that there are no such assurances in the case of partial observation (the so-called linear-quadratic-Gaussian or LQG case, which is an LQR control matched with a Kalman filter). Doyle's paper title and abstract are classic in the literature – title: 'Guaranteed Margins for LQG Regulators', abstract: 'There are none'.

Spurred by this and related work, control theorists started to move away from LQ control to look for a more robust approach. Zames (1981) was influential in the development of H_∞ control as a more robust alternative to LQ control. Loosely speaking, in LQ control the quadratic cost means that performance is measured with a 2-norm across frequencies. By contrast, H_∞ uses an ∞ $-$ norm that looks at the peak of the losses across frequencies. It is also interpretable as the maximal magnification of the disturbances to outputs of interest. While the early robust control literature used a frequency domain approach, in the late 1980s Doyle and others developed state space formulations (see Doyle et al. 1989, for example) which gave explicit solutions and allowed for alterative formalizations. For example, the H_∞ approach was given alternative justifications in terms of penalizing disturbances from the nominal model, which can be implemented as a dynamic game between a decision maker seeking to minimize losses and a malevolent agent seeking to maximize loss. (See Başer and Bernhard 1995, for a development of this approach.) Finally, the uncertainty sets in the H_∞ approach are unstructured – they represent perturbations of the model which are bounded but have no particular form. The implications of structured perturbations have been studied more recently. Some examples include parametric perturbations, unmodelled dynamics, or uncertainty only about particular channels or connections in a model. Applications with structured uncertainty use the structured singular value (also known as μ) rather than the H_∞ norm as a measure of performance. Although there are some important stability and performance criteria, in general constructing control rules is a more daunting task, and the theory is not as fully developed as the unstructured case.

The Hansen–Sargent Approach

In the economics literature, the most prominent and influential approach to robust control is due to Hansen and Sargent (and their co-authors), which is summarized in their monograph Hansen and Sargent (2008). This approach starts with a nominal model and uses entropy as a distance measure to calibrate the model uncertainty set. More specifically, the model set consists of those models whose relative entropy or Kullback–Leibler distance from the nominal model is bounded by a specified value. Note that this puts no structure on the uncertainty, but only restricts the alternative models to those which are difficult to distinguish statistically from the nominal model. In practice, a Lagrange multiplier theorem is typically used to convert the entropy constraint into a penalty on perturbations from the model. Then the solution of the control problem is found via a dynamic game implementation: the agent maximizes utility by his choice of control, while an evil agent minimizes utility by his choice of perturbation, while being penalized by the entropy of the deviations. Relative to the control theory literature such as Başar and Bernhard (1995), the main differences are that all models are stochastic, while control theory largely uses deterministic models. One exception is Petersen et al. (2000) who use a similar approach to consider uncertain stochastic systems. In addition, discounting is not typically considered in control theory, while it is natural in economics. In full information problems discounting has relatively little effect, but it raises important issues in problems with partial information (see Hansen and Sargent 2005a, b). Finally, the Hansen–Sargent approach naturally extends beyond the LQ setting laid out in Hansen et al. (1999), with some examples in Anderson et al. (2003), Cagetti et al. (2002) and Maenhout (2004).

To be more concrete, consider an LQ example where x_t is the state, i_t is the agent's control, and ε_t is an i.i.d. Gaussian shock. The nominal model is:

$$x_{t+1} = Ax_t + Bi_t + C\varepsilon_{t+1}, \tag{1}$$

and the agent's intertemporal preferences are:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(x_t' Q x_t + i_t' R i_t \right) \tag{2}$$

where $0 < \beta < 1$ and Q and R are negative definite matrices. The approach of Hansen and Sargent perturbs the nominal model with an additional 'misspecfication shock' w_{t+1} which is allowed to be correlated with the state x_t:

$$x_{t+1} = Ax_t + Bi_t + C(\varepsilon_{t+1} + w_{t+1}). \tag{3}$$

The shock w_{t+1} is used to represent alternative models. These models are made to be close to the nominal model in an entropy sense by imposing the bound:

$$E_0 \sum_{t=0}^{\infty} \beta^t w_{t+1}' w_{t+1} \le \eta \tag{4}$$

for some constant $\eta \ge 0$. The agent then maximizes (2) with respect to the worst case perturbed model (3) from the set (4). Using a Lagrange multiplier theorem, the constraint set can be converted to a penalty and the decision problem can by solved recursively by solving the Bellman equation for a two-player zero sum game:

$$V(x) = \max_i \min_w \{x'Qx + i'Ri + \beta\theta w'w + \beta E[V(Ax + Bi + C(\varepsilon + w))|x]\} \tag{5}$$

where $\theta > 0$ is a Lagrange multiplier on the constraint (4) and the expectation is over the Gaussian shock ε. Often this multiplier formulation is taken as the starting point, for example Maccheroni et al. (2006) characterize preferences of this form, with θ governing the degree of robustness. As $\theta \to \infty$ the penalization becomes so great that only the nominal model remains (thus $\eta \to 0$), and the decision rule is less robust. Conversely, there is typically a minimal value of θ beyond which the value is $V(x) = -\infty$. This gives the most robust decision rules, allowing for the largest uncertainty set.

Adding Structure to the Uncertainty Set

The approach discussed above uses unstructured uncertainty, and has been well developed and extended in different dimensions. We now discuss some alterative approaches which put more structure on the uncertainty set. There are many reasons to do so. It may be that some of the models that are close to the nominal model in a statistical sense may not be plausible economically. Alternatively, the decision makers may have a discrete set of models in mind, and bounding them all in one uncertainty set may include extraneous implausible models. Perhaps most substantively, the decision maker may be more confident some aspects of the model relative to others. Some examples of this include knowing the model up to the values of parameters, or being more certain about the dynamics of certain variables in the model. Not taking into account the particular structure may give a misleading impression of the actual uncertainty the decision makers face.

There are many ways of building in structured uncertainty, and the distinctions between cases are not always clear. For example, consider the same nominal model as above, but suppose that instead of the unstructured perturbations (3) the uncertainty is instead solely in the values of the parameters A and B. Thus we can represent the parametric perturbed models as:

$$x_{t+1} = \left(A + \widehat{A}\right)x_t + \left(B + \widehat{B}\right)i_t + C\varepsilon_{t+1} \tag{6}$$

for some matrices $\hat{A}$ and $\widehat{B}$ Of course it's possible to rewrite (6) as a version of (3) with:

$$w_{t+1} = \widehat{A}x_t + \widehat{B}i_t, \tag{7}$$

so in principle parametric perturbations are just a special case of the unstructured uncertainty. However what makes a substantive difference is how uncertainty is measured, that is whether we restrict w_{t+1} as in (4) or whether we restrict the parameters $\hat{A}$ and $\widehat{B}$ say by bounding them in a confidence ellipsoid around the nominal model. Moreover, as (7) makes clear the differences between the uncertainty measurements will depend on the actual control rule i_t in place. Onatski and Williams (2003) provide an example of a simple estimated model where the uncertainty specifications matter dramatically for outcomes. In particular, the optimal policy for the largest possible unstructured uncertainty set (that is for the minimal value of θ) leads to instability for relatively small parametric perturbations. Thus the particular structure and measurement of uncertainty can have important implications for decisions. (Petersen et al. 2000, modify the unstructured approach described above to deal with structured uncertainty by separating the entropy penalty for unstructured perturbations from a different penalization for structured perturbations.)

Some economic applications with structured uncertainty include the following:

1. The simplest cases are uncertainty sets with discrete possible models. Some examples include: Levin and Williams (2003), who consider both Bayesian and minimax approaches;

R

Cogley and Sargent (2005); and Svensson and Williams (2006) and who focus on a Bayesian approach, and the recent work of Hansen and Sargent (2006), who have built this type of structure into their robust approach.

2. Another common form is parameter uncertainty within a fully specified model. Brainard (1967) is the classic reference from a Bayesian perspective with many references in this line, while Giannoni (2002) and Chamberlain (2000) consider minimax approaches.
3. Somewhat more broad are cases with different parametric model specifications. For example this includes uncertainty about dynamics (lags and leads), variables which may enter, uncertainty about data quality, and other features which are built into parametric extensions of the nominal model. Examples include the model error modelling approach of Onatski and Williams (2003) and the empirical specifications of Brock et al. (2003).
4. Finally, the model sets may be nonparametric but structured in particular ways. For example, Onatski and Stock (2002) consider different structured types of uncertainty such as linear time-invariant perturbations, nonlinear time-varying perturbations, and perturbations which only enter particular parts of the model. Other examples include nonparametric specifications of uncertainty which differs across frequencies as in Onatski and Williams (2003) and Brock and Durlauf (2005).

See Also

▶ Ambiguity and Ambiguity Aversion
▶ Model Uncertainty
▶ Stochastic Optimal Control
▶ Uncertainty

Bibliography

Anderson, E., L.P. Hansen, and T. Sargent. 2003. A quartet of semi-groups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association* 1: 68–123.

Backus, D.K., B.R. Routledge, and S.E. Zin. 2005. Exotic preferences for macroeconomists. In *NBER macroeconomics annual 2004*, ed. M. Gertler and K. Rogoff. Cambridge, MA: MIT Press.

Başar, T., and P. Bernhard. 1995. *H_∞-optimal control and related minimax design problems*. Boston: Birkhauser.

Brainard, W. 1967. Uncertainty and the effectiveness of policy. *American Economic Review* 57: 411–425.

Brock, W.A., and S.N. Durlauf. 2005. Local robustness analysis: Theory and application. *Journal of Economic Dynamics and Control* 29: 2067–2092.

Brock, W.A., S.N. Durlauf, and K.D. West. 2003. Policy evaluation in uncertain economic environments. *Brookings Papers on Economic Activity* 2003(1): 235–301.

Burl, J.B. 1999. *Linear optimal control: H_2 and H_∞ methods*. Menlo Park: Addison-Wesley.

Cagetti, M., L.P. Hansen, T.J. Sargent, and N. Williams. 2002. Robustness and pricing with uncertain growth. *Review of Financial Studies* 15: 363–404.

Chamberlain, G. 2000. Econometric applications of maxmin expected utility theory. *Journal of Applied Econometrics* 15: 625–644.

Cogley, T., and T.J. Sargent. 2005. The conquest of U.-S. inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics* 8: 528–563.

Doyle, J.C. 1978. Guaranteed margins for LQG regulators. *IEEE Transactions on Automatic Control* 23: 756–757.

Doyle, J.C., K. Glover, P.P. Khargonekar, and B.A. Francis. 1989. State-space solutions to standard H_2 and H_∞ control problems. *IEEE Transactions on Automatic Control* 34: 831–847.

Duffie, D., and L.G. Epstein. 1992. Stochastic differential utility. *Econometrica* 60: 353–394.

Ellsberg, D. 1961. Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics* 75: 643–669.

Epstein, L., and M. Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113: 1–31.

Epstein, L., and S. Zin. 1989. Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57: 937–969.

Giannoni, M. 2002. Does model uncertainty justify caution? Robust optimal monetary policy in a forward-looking model. *Macroeconomic Dynamics* 6: 111–144.

Gilboa, I., and D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18: 141–153.

Hansen, L.P., and T.J. Sargent. 2001. Acknowledging misspecification in macroeconomic theory. *Review of Economic Dynamics* 4: 519–535.

Hansen, L.P., and T.J. Sargent. 2005a. Robust estimation and control under commitment. *Journal of Economic Theory* 124: 258–301.

Hansen, L.P., and T.J. Sargent. 2005b. *Robust estimation and control without commitment*, Working paper. Chicago: Department of Economics, University of Chicago.

Hansen, L.P., and T.J. Sargent. 2006. *Fragile beliefs and the price of model uncertainty*. Working paper, Department of Economics, University of Chicago.

Hansen, L.P., and T.J. Sargent. 2008. *Robustness*. Princeton: Princeton University Press.
Hansen, L.P., T. Sargent, and T. Tallarini. 1999. Robust permanent income and pricing. *Review of Economic Studies* 66: 873–907.
Hansen, L.P., T.J. Sargent, G.A. Turmuhambetova, and N. Williams. 2006. Robust control and model misspecification. *Journal of Economic Theory* 128: 45–90.
Jacobson, D.H. 1973. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control AC* 18: 1124–1131.
Kreps, D.M., and E.L. Porteus. 1978. Temporal resolution of uncertainty and dynamic choice. *Econometrica* 46: 185–200.
Levin, A.T., and J. Williams. 2003. Robust monetary policy with competing reference models. *Journal of Monetary Economics* 50: 945–975.
Limebeer, D.J.N., B.D.O. Anderson, and B. Hendel. 1994. A Nash game approach to mixed H_2/H_∞ control. *IEEE Transactions on Automatic Control* 39: 69–82.
Maccheroni, F., M. Marinacci, and A. Rustichini. 2006. Dynamic variational preferences. *Journal of Economic Theory* 128: 4–44.
Maenhout, P.J. 2004. Robust portfolio rules and asset pricing. *Review of Financial Studies* 17: 951–983.
Onatski, A., and J.H. Stock. 2002. Robust monetary policy under model uncertainty in a small model of the US economy. *Macroeconomic Dynamics* 6: 85–110.
Onatski, A., and N. Williams. 2003. Modeling model uncertainty. *Journal of the European Economic Association* 1: 1087–1122.
Petersen, I.R., M.R. James, and P. Dupuis. 2000. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control* 45: 398–412.
Sims, C.A. 2001. Pitfalls of a minimax approach to model uncertainty. *American Economic Review* 91: 51–54.
Skiadas, C. 2003. Robust control and recursive utility. *Finance and Stochastics* 7: 475–489.
Svensson, L.E. 2001. *Robust control made simple*. Working paper, Department of Economics, Princeton University.
Svensson, L.E.O., and N. Williams. 2006. *Monetary policy with model uncertainty: Distribution forecast targeting*. Working paper, Department of Economics, Princeton University.
Tornell, A. 2003. *Exchange rate anomalies under model misspecification: A mixed optimal/robust approach*. Working paper, Department of Economics, UCLA
Whittle, P. 1981. Risk sensitive linear quadratic Gaussian control. *Advances in Applied Probability* 13: 764–777.
Zames, G. 1981. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Transactions on Automatic Control* 26: 301–320.
Zhou, K., J. Doyle, and K. Glover. 1996. *Robust and optimal control*. Upper Saddle River: Prentice Hall.

Robust Estimators in Econometrics

P. Čížek and W. Härdle

Abstract

Econometric data are often obtained under conditions that cannot be well controlled, and so partial departures from the model assumptions in use (data contamination) occur relatively frequently. To address this, we first introduce concepts of robust statistics for qualifying and quantifying sensitivity of estimation methods to data contamination as well as important approaches to robust estimation. Later, we discuss how robust estimation methods have been adapted to various areas of econometrics, including time series analysis and general GMM-based estimation.

Keywords

Breakdown point; Data contamination; Generalized method of moments; Heteroscedasticity; Influence function; Instrumental variables; Least squares; Linear regression; Maximum likelihood; Maximum-bias curve; Measurement errors; Mestimation; Newton–Raphson procedure; Nonlinear models; Probability distribution; Qualitative and quantitative robustness; Quantile regression; Random variables; Robust econometrics; Robust statistics; S-estimation; Simultaneous equations models; Timeseries analysis

JEL Classifications

C14

Econometrics often deals with data under, from the statistical point of view, nonstandard conditions such as heteroscedasticity or measurement errors, and the estimation methods thus need either to be adapted to such conditions or to be at least insensitive to them. Methods insensitive to violation of certain assumptions – for example, insensitive to the presence of heteroscedasticity – are in a broad sense referred to as 'robust' (for example, robust to heteroscedasticity). On the other hand, there is also a more specific meaning of the word 'robust', which stems from the field of robust statistics. This latter notion defines robustness rigorously in terms of the behaviour of an estimator both at the assumed (parametric) model and in its neighbourhood in the space of probability distributions. Even though the methods of robust statistics have been used only in the simplest settings, such as estimation of location, scale or linear regression for a long time, they have motivated a range of new econometric methods, which we focus on in this article.

The concepts and measures of robustness are introduced first (section "Measures of Robustness"), followed by the most common types of estimation methods and their properties (section "Estimation Approaches"). Various econometric methods based on these common estimators are discussed in section "Methods of Robust Econometrics", covering tasks from time-series regression over GMM estimation to simulation-based methods.

Measures of Robustness

Robustness properties can be formulated within two frameworks: qualitative and quantitative robustness. *Qualitative robustness* is concerned with the situation in which the shape of the underlying (true) data distribution deviates slightly from the assumed model. It focuses on questions like stability and performance loss over a family of such slightly deviating distributions. *Quantitative robustness* is involved when the sensitivity of estimators to a proportion of aberrant observations is studied.

A simple example can make this clear. Suppose one has collected a sample on an individual's income (after say ten years of schooling) and one is interested in estimating the mean income. If $\{x_i\}_{i=1}^{n}$ denotes the logarithm of this data and we suppose that they have a cumulative distribution function (cdf) F, assumed to be N (μ,σ_2), the maximum likelihood estimator (MLE) is $\bar{x} = \int u dF_n(u) = T(F_n)$, where $F_n(u) = n^{-1}\sum_{i=1}^{n} I(x_i \leq u)$, and $\mu = \int u dF(u) = T(F)$. Qualitative robustness asks here the question: how well will μ be estimated if the true distribution is in some neighbourhood of F ? Quantitative robustness would concentrate on the question: will $T(F_n)$ be bounded if some observations $x_i \to \infty$? In fact, the latter question is easy to answer: if $x_i \to \infty$ for some i, $T(F_n) = \bar{x} \to \infty$ as well. So we can say here in a loose sense that $\bar{x}$ is not quantitatively robust.

Formalities

In the following we present a mathematical set-up that allows us to formalize our thoughts on robustness.

The notion of the sensitivity of an estimator T is put into theory by considering a model characterized by a cdf F and its neighbourhood $\mathcal{F}_{\varepsilon,G}$: distributions $(1 - \varepsilon)F + \varepsilon G$, where $\varepsilon \in (0,1/2)$ and G is an arbitrary probability distribution, which represents data contamination. Hence, not all data necessarily follow the pre-specified distribution, but the ε-part of data can come from a different distribution G. If $H \in \mathcal{F}_{\varepsilon,G}$, the estimation method T is then judged by how sensitive or robust the estimates $T(H)$ are to the size of $\mathcal{F}_{\varepsilon,G}$, or alternatively, to the distance from the assumed cdf F. Two main concepts for robust measures analyse the sensitivity of an estimator to infinitesimal deviations, $\varepsilon \to 0$, and to finite (large) deviations, $\varepsilon > 0$, respectively. Despite generality of the concept, easy interpretation and technical difficulties often limit our choice to point-mass distributions (Dirac measures) $G = \delta_x$, $x \in \mathbb{R}$, which simply represents an (erroneous) observation at point $x \in \mathbb{R}$. This simplification is also used in the following text.

Qualitative Robustness

The influence of infinitesimal contamination on an estimator is characterized by the *influence function*, which measures the relative change in estimates caused by an infinitesimally small amount ε of contamination at x (Hampel et al. 1986). More formally,

$$IF(x; T, F) = \lim_{\varepsilon \to 0} \frac{T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)}{\varepsilon}. \quad (1)$$

For each point x, the influence function reveals the rate at which the estimator T changes if a wrong observation appears at x. In the case of sample mean $\bar{x} = T(F_n)$ for $\{x_i\}_{i=1}^n$, we obtain

$$\begin{aligned} &IF(x; T, F_n) \\ &= \left[(1-\varepsilon)\int u dF_n(u) + \varepsilon \int u d\delta_x(u) - \int u dF_n(u)\right] / \varepsilon \\ &= \lim_{\varepsilon \to 0} \left[-\int u dF_n(u) + \int u d\delta_x(u)\right] = x - \bar{x}. \end{aligned}$$

The influence function allows us to define various desirable properties of an estimation method. First, the largest influence of contamination on estimates can be formalized by the *gross-error sensitivity*,

$$\gamma(T, F) = \sup_{x \in \mathbb{R}} IF(x; T, F), \quad (2)$$

which under robustness considerations is finite and small. Even though such a measure can depend on F in general, the qualitative results (for example, $\gamma(T,F)$ being bounded) are typically independent of F. Second, the sensitivity to small changes in data, for example moving an observation from x to $y \in \mathbb{R}$, can be measured by the *local-shift sensitivity*

$$\lambda(T, F) = \sup_{x \neq y} \frac{||IF(x; T, F) - IF(y; T, F)||}{||x - y||}. \quad (3)$$

Also, this quantity should be relatively small since we generally do not expect that small changes in data cause extreme changes in values or sensitivity of estimates. Third, as unlikely large or distant observations may represent data errors, their influence on estimates should become zero. Such a property is characterized by the *rejection point*,

$$\rho(T, F) = \inf_{r > 0} \{r : IF(x; T, F) = 0, ||x|| \geq r\}, \quad (4)$$

which indicates the non-influence of large observations.

Quantitative Robustness

Alternatively, the behaviour of the estimator T can be studied for any finite amount ε of contamination. The most common property looked at in this context is the estimator's bias $b(T, H) = E_H\{T(H)\} - E_F\{T(F)\}$, which measures a distance between the estimates for clean data, $T(F)$, and contaminated data, $T(H)$, $H \in \mathcal{F}_{\varepsilon,G}$. The corresponding *maximum-bias curve* measures the maximum bias of T on $\mathcal{F}_{\varepsilon,G}$ at any ε:

$$B(\varepsilon, T) = \sup_{x \in \mathbb{R}} b\{T, (1-\varepsilon)F + \varepsilon\delta_x\}. \quad (5)$$

Although the computation of this curve is rather complex, Berrendero and Zamar (2001) provide general methodology for its computation in the context of linear regression.

The maximum-bias curve is not only useful on its own, but allows us to define further scalar measures of robustness. The most prominent is the *breakdown point* (Hampel 1971), which is defined as the smallest amount ε of contamination that can cause an infinite bias:

$$\varepsilon^*(T) = \inf_{\varepsilon \geq 0} \{\varepsilon : B(\varepsilon, T) = \infty\}. \quad (6)$$

The intuitive aim of this definition specifies the breakdown point $\varepsilon^*(T)$ as the smallest amount of contamination that makes the estimator T useless. Note that in most cases $\varepsilon^*(T) \leq 0.5$ (He and Simpson 1993). This definition and the upper bound, however, apply only in simple cases, such as location or linear regression estimation (Davies and Gather 2005). The most general definition of breakdown point formalizes the idea of 'useless' estimates in the following way: an estimator is said to break down if, under contamination, it is not random anymore, or, more precisely, it can achieve only a finite set of values (Genton and Lucas 2003). This definition is based on the fact that estimates are functions of observed random samples and are thus random quantities themselves unless they fail. Although the latter definition includes the first one, the latter one may generally depend on the underlying model F, for example in time-series context.

Estimation Approaches

Denote by F_n an empirical distribution function (edf) corresponding to a sample $\{x_i\}_{i=1}^n \in \mathbb{R}$ drawn from a model based on probability distribution F. Most estimation methods can be defined as an extremum problem, minimizing a contrast $\int h(z, \theta)dF(z)$ over θ in a parameter space, or as a solution of an equation, $\int g(z, \theta)dF(z) = 0$ in θ. The estimation for a given sample utilizes finite-sample equivalents of these integrals, $\int h(z, \theta) dF_n(z)$ and $\int g(z, \theta)dF_n(z)$, respectively.

Consider the pure location model $X_i = \mu + \sigma\varepsilon_i$, $i = 1, \ldots, n$, with a known scale σ and $\varepsilon \sim F$. The cdf of X is then $F\{(x - \mu)/\sigma\}$. With a quadratic contrast function $h(x, \theta) = (x - \theta)^2$, the estimation problem is to minimize $\int(x - \theta)^2\, dF\{(x - \mu)/\sigma\}$ with respect to θ. For known F, this leads to $\theta = \mu$ and one sees that, without loss of generality, one can assume $\mu = 0$ and $\sigma = 1$. For the sample $\{x_i\}_{i=1}^n$ characterized by edf F_n, the location parameter μ is estimated by

$$\hat{\mu} = \arg\min_{\theta} \int (x-\theta)^2 dFn(x) = n^{-1}\sum_{i=1}^{n} x_i = \bar{x}.$$

Note that for $g(x, \theta) = x - \theta$, the parameter μ is the solution to $\int g(x, \theta)dF(x) = 0$. The estimator may therefore be alternatively defined through $\mu = T(F) = \int udF(u)$.

As indicated in the introduction, this standard estimator of location performs unfortunately rather poorly under the sketched contamination model. Estimating a population mean by the least squares (LS) or sample mean $\bar{x} = T(F_n)$ has the following properties. First, the influence function (Eq. (1))

$$\begin{aligned} IF(x;T,F) &= \lim_{\varepsilon\to 0}\frac{T\{(1-\varepsilon)F+\varepsilon\delta_x\}-T(F)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0}\frac{\left\{(1-\varepsilon)\int udF(u)+\varepsilon x\right\}-\int udF(u)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0}\varepsilon^{-1}\left\{-\varepsilon\int udF(u)+\varepsilon x\right\} \\ &= x-\int udF(u) = x-T(F). \end{aligned}$$

Hence, the gross-error sensitivity (Eq. 2) $\gamma(T, F) = \infty$, the local-shift sensitivity (Eq. 3) $\lambda(T, F) = 0$, and the rejection point (Eq. 4) $\rho(T, F) = \infty$. Second, the maximum-bias (Eq. 5) is infinite for any $\varepsilon > 0$ since

$$\begin{aligned} &\sup_{x\in\mathbb{R}} ||T\{(1-\varepsilon)F+\varepsilon\delta_x\}-T(F)|| \\ &\quad = \sup_{x\in\mathbb{R}} ||-\varepsilon T(F)+\varepsilon x|| = \infty. \end{aligned}$$

Consequently, the breakdown point (Eq. 6) of the sample mean $\bar{x} = T(F_n)$ is zero, $\varepsilon^*(T) = 0$.

Thus, none of robustness measures characterizing the change of T under contamination of data (even infinitesimally small) is finite. This behaviour, typical for LS-based methods, motivated alternative estimators that have the desirable robust properties. In this section, the *M*-estimators, *S*-estimators and τ-estimators are discussed as well as some extensions and combination of these approaches. Even though there is a much wider range of robust estimation principles, we focus on those already studied and adopted in various areas of econometrics.

M-estimators

To achieve more flexibility in accommodating requirements on robustness, Huber (1964) proposed the *M-estimator* by considering a general extremum estimator based on $\int \rho(z, \theta)dF(z)$, thus minimizing $\int \rho(z, \theta)dF_n(z)$ in finite samples. Providing that the first derivative $\psi(z, \theta) = \partial\rho(z, \theta)/\partial\theta$ exists, an *M*-estimator can be also defined by an implicit equation $\int \psi(z, \theta)dF_n(z) = 0$.

This extremely general definition is usually adapted to a specific estimation problem such as location, scale or regression estimation. In a univariate location model, $F(z)$ can be parameterized as $F(z - \theta)$ and hence one limits $\rho(z, \theta)$ and $\psi(z, \theta)$ to $\rho(z - \theta)$ and $\psi(z - \theta)$. In the case of scale estimation, $F(z) = F(z/\theta)$ and consequently $\rho(z, \theta) = \rho(z/\theta)$ and $\psi(z, \theta) = \psi(z/\theta)$. In linear regression, $z = (x, y)$ and a zero-mean error term $\varepsilon = y - x^{\mathrm{T}}\theta$. Analogously to the location case, one can then consider $\rho(z, \theta) = \rho(y - x^{\mathrm{T}}\theta)$ and $\psi(z, \theta) = \psi(y - x^{\mathrm{T}}\theta)x$, or more generally, $\rho(z, \theta) = \rho(y - x^{\mathrm{T}}\theta, x)$ and $\psi(z, \theta) = \psi(y - x^{\mathrm{T}}\theta, x)$

Robust Estimators in Econometrics, Table 1 Examples of ρ and ψ functions used with *M*-estimators

	$\rho(t)$	$\psi(t)$
Least squares	t_2	$2\,t$
Least absolute deviation	$\lvert t\rvert$	Sign (t)
Quantile estimation	$\{\tau - I(x<0)\}x$	$\tau - I(x<0)$
Huber: for $\lvert t\rvert \le c$	t^2	$2\,t$
for $c < \lvert t\rvert$	$c\lvert t\rvert$	c sign(t)
Huber: for $\lvert t\rvert \le c$	t^2	$2\,t$
for a $< \lvert t\rvert \le b$	$a\lvert t\rvert$	a sign(t)
for b $< \lvert t\rvert \le c$	$\frac{ac}{c-b}t - \frac{a}{c-b}t^2\text{sign}(t)$	$a(c-\lvert t\rvert)/(c-b)$
for c $< \lvert t\rvert$	$a\lvert t\rvert$	0
Biweight (Tukey)	$-(c^2-t^2)^3 I(\lvert t\rvert \le c)/6$	$t(c^2-t^2)^2 I(\lvert t\rvert \le c)/6$
Sine (Andrews)	$-c\, cos(x/c)I(\lvert t\rvert \le \pi c)$	$sin\,(x/c)I(\lvert t\rvert \le \pi c)$

(*GM*-estimators). Generally, we can express $\rho(z,\theta)$ as $\rho\{\eta(z,\theta)\}$, $\psi\{\eta(z,\theta)\}$, where $\eta(z,\theta) \sim F$.

Some well-known choices of univariate objective functions ρ and ψ are given in Table 1; functions $\rho(t)$ are usually assumed to be non-constant, non-negative, even and continuously increasing in $|t|$. This documents flexibility of the concept of *M*-estimators, which include LS and quantile regression as special cases.

On the other hand, many of the ρ and ψ functions in Table 1 depend on one or more constants $a, b, c \in \mathbb{R}$. If an estimator T is to be invariant to the scale of data, one can apply the estimator to rescaled data, that is, to minimize $\int \rho\{(z-\theta)/s\}\, dF_n(z)$ or to solve $\int \psi\{(z-\theta)/s\}dF_n(z) = 0$ for a scale estimate s like the median absolute deviation (MAD). Alternatively, one may also estimate parameters θ and scale s simultaneously by considering $\rho(z, \{\theta, s\}) = \rho\{(z-\theta)/s\}$ or

$$\psi(z,\{\theta,s\}) = \{\psi_1(z,\{\theta,s\}), \psi_s(z,\{\theta,s\})\}.$$

Let us now turn to the question how the choice of functions ρ and ψ determines the robust properties of *M*-estimators. First, the influence function of an *M*-estimator can generally depend on several quantities such as its asymptotic variance or the position of explanatory variables in the regression case, but the influence function is always proportional to function $\psi(z,b)$. Thus, the finite gross-error sensitivity, $\gamma(T, F) < \infty$, requires bounded $\psi(t)$ (which is not the case with LS). Similarly, the finite rejection point, $\rho(T, F) < \infty$, leads to $\psi(t)$ being zero for all sufficiently large t (the *M*-estimators defined by such a ψ-function are called redescending). Hampel et al. (1986) shows how, for a given bound on $\gamma(T,F)$, one can determine the most efficient choice of ψ function (for example, the skipped median, $\psi(t) = \text{sign}(t)I(|t| < K), K > 0$, in the location case).

More formally, the optimality of *M*-estimators in the context of qualitative robustness can be studied by the *asymptotic relative efficiency* (ARE) of an estimator $\hat{\theta}^1$ relative to another estimator $\hat{\theta}^2$:

$$ARE\left(\hat{\theta}^1, \hat{\theta}^2\right) = \frac{\text{as.var}\left(\hat{\theta}^1\right)}{\text{as.var}\left(\hat{\theta}^2\right)}. \tag{7}$$

For example, at the normal distribution with $\hat{\theta}^1$ and $\hat{\theta}^2$ being the least absolute deviation (LAD) and LS estimators, ARE equals $2/\pi \approx 0.64$. Under the Student cdf t_5, the ARE of the two estimators climbs up to $\approx$0.96. For Huber's M-estimator, we see that its limit cases are the median for $c \to 0$ and the mean for $c \to \infty$. At the normal distribution and for $c = 1.345$, we have ARE of about 0.95. This means that this *M*-estimator is almost as efficient as MLE, but does not lose so drastically in performance as the standard mean under contamination because of the bounded influence function.

Whereas the influence function of M-estimators is closely related to the choice of its objective function, the global robustness of

M-estimators is in a certain sense independent of this choice. Maronna et al. (1979) showed in linear regression that the breakdown point of M-estimators is bounded by 1/p, where p is the number of estimated parameters. As a remedy, several authors proposed one-step M-estimators that are defined, for example, as the first step of the iterative Newton–Raphson procedure, used to minimize $\int \rho(z, \theta) dF(z)$, starting from initial robust estimators $\hat{\theta}^0$ of parameters and $\hat{s}^0$ of scale (see Welsh and Ronchetti 2002, for an overview). Possible initial estimators can be those discussed in sections "S-estimators" and "Multivariate Regression". For example for an M-estimator of location $\hat{\theta}$ defined by a function $\psi(x, \theta) = \psi(x - \theta)$, its one-step counterpart can be defined at sample $\{x_i\}_{i=1}^n$ by

$$\hat{\theta} = \hat{\theta}^0 + \hat{s}^0 \sum_{i=1}^n \psi\left(\frac{x_i - \hat{\theta}^0}{\hat{s}^0}\right) \Big/ \sum_{i=1}^n \psi'\left(\frac{x_i - \hat{\theta}^0}{\hat{s}^0}\right),$$

where $\hat{\theta}^0$ and $\hat{s}^0$ represent initial robust estimators of location and scale like the median and MAD, respectively. Such one-step estimators, under certain conditions on the initial estimators, preserve the breakdown point of the initial estimators, and at the same time have the same first-order asymptotic distribution as the original Mestimator (Simpson et al. 1992; Welsh and Ronchetti 2002). Further development of such ideas includes an adaptive choice of parameters of function ψ in the iterative step (Gervini and Yohai 2002).

S-estimators

An alternative approach to *M*-estimators achieving high breakdown point (HBP) was proposed by Rousseeuw and Yohai (1984). The S-estimators are defined by minimization of a scale statistics $s^2(z, b) = s\{\eta(z, b)\}$ defined as the *M*-estimate of scale,

$$\int \rho[\eta(z, b)/s\{\eta(z, b)\}] dF_n(z) = K = \int \rho(t) dF(t),$$

at the model distribution *F*; the functions ρ and η are those defining *M*-estimators in section "M-estimators". More generally, one can define *S*-estimators by means of any scaleequivariant statistics s^2, that is, $s\{c\eta(z, b)\} = |c| s\{\eta(z, b)\}$. Under this more general definition, *S*-estimators include as special cases LS and LAD estimators. Further, they encompass several well-known robust methods including least median of squares (LMS) and least trimmed squares (LTS): whereas the first defines the scale statistics $s^2\{\eta(z, b)\}$ as the median of squared residuals $\eta(z, b)$, the latter uses the scale defined by the sum of *h* smallest residuals $\eta(z, b)$. In order to appreciate the difference to *M*-estimators, it is worth pausing for a moment and to present LMS, the most prominent representative of *S*-estimators, in the location case:

$$\arg\min_{\theta} \operatorname{med}\left\{(x_1 - \theta)^2, \ldots, (x_n - \theta)^2\right\}.$$

Due to its definition, the *S*-estimators have the same influence function as the *M*-estimator constructed from the same function ρ. Contrary to *M*-estimators, they can achieve the highest possible breakdown point $\varepsilon^* = 0.5$. For example, this is the case of LMS and LTS. For Gaussian data, the most efficient (in the sense of ARE (Eq. 7)) among the *S*-estimators with $\varepsilon^* = 0.5$ is, however, the one corresponding to $K = 1.548$ and ρ being the Tukey biweight function (see Table 1). Given the HBP of *S*-estimators, their maximum-bias behaviour is of interest too. Although it depends on the function ρ and constant *K* (Berrendero and Zamar 2001), Yohai and Zamar (1993) proved that LMS minimizes maximum bias among a large class of (residual admissible) estimators, which includes most robust methods.

An important shortcoming of HBP *S*-estimation is, however, its low ARE: under Gaussian data, efficiency relative to LS varies from zero per cent to 27 per cent. Thus, *S*-estimators are often used as initial estimators for other, more efficient methods. Nevertheless, if an *S*-estimator is not applied directly to sample observations, but rather to the set of all pairwise differences of sample observations, the resulting generalized *S*-estimator exhibits higher relative efficiency for Gaussian data, while preserving its robust properties (Croux et al. 1994; Stromberg et al. 2000).

τ-estimators

The *S*-estimators improve upon *M*-estimators in terms of their breakdown-point properties, but at the cost of low Gaussian efficiency. Although one-step *M*-estimators based on an initial *S*-estimate can remedy this deficiency to a large extent, their exact breakdown properties are not known. One alternative approach, proposed by Yohai and Zamar (1988), extends the principle of *S*-estimation in the following way. Assuming that ρ_1 and ρ_2 are non-negative, even, and continuous functions, the *M*-estimate $s^2(z, \theta) = s^2\{\eta(z, \theta)\}$ scale can be defined as in the case of *S*-estimation,

$$\int \rho_1[\eta(z,\theta)/s\{\eta(z,\theta)\}]dF_n(z) = K = \int \rho_1(t)dF(t).$$

Next, the τ-estimate of scale is defined by

$$\tau^2(z,\theta) = s^2\{\eta(z,\theta)\}\int \rho_2[\eta(z,\theta)/s\{\eta(z,\theta)\}] dF_n(z)$$

and the corresponding τ-*estimator* of parameters θ is then defined by minimizing the τ-estimate of scale, $\tau^2(z, \theta)$.

As a generalization of *S*-estimation, the τ-estimators include *S*-estimators as a special case for $\rho_1 = \rho_2$ because then $\tau^2(x, \theta) = \theta s^2(z, \theta)$. On the other hand, if $\rho_2(t) = t^2$, $\tau^2(z, \theta) = \int \eta^2(z, \theta) dF_n(z)$ is just the standard deviation of model residuals. Compared with *S*-estimators, the class of τ estimators can improve in terms of relative Gaussian efficiency because its breakdown point depends only on function ρ_1, whereas its asymptotic variance is a function of both ρ_1 and ρ_2. Thus, ρ_1 can be defined to achieve the breakdown point equal to 0.5 and ρ_2 consequently adjusted to reach a pre-specified relative efficiency for Gaussian data (for example, 95 per cent).

Methods of Robust Econometrics

The concepts and methods of robust estimation discussed in sectinos "Measures of Robustness" and "Estimation Approaches" are typically proposed in the context of a simple location or linear regression models, on the assumption of independent, continuous and identically distributed random variables. This, however, rarely corresponds to assumptions typical for most econometric models. In this section, we therefore present an overview of developments and extensions of robust methods to various econometric models. As the *M*-estimators are closest to the commonly used LS and MLE, most of the extensions employ *M*-estimation. The HBP techniques are not that frequently found in the economics literature (Zaman et al. 2001; Sapra 2003) and are mostly applied only as a diagnostic tool.

In the rest of this section, robust estimation is first discussed in the context of models with discrete explanatory variables, models with time-dependent observations, and models involving multiple equations. Later, robust alternatives to general estimation principles, such as MLE and generalized method of moments (GMM), are discussed. Before doing so, let us mention that dangers of data contamination are not studied only from the theoretical point of view. There is a number of studies that check the presence of outliers in real data and their influence on estimation methods. For example, there is evidence of data contamination and its adverse effects on LS and MLE in the case of macroeconomic time series (Balke and Fomby 1994; Atkinson et al. 1997), in financial time series (Sakata and White 1998; Franses et al. 2004), marketing data (Franses et al. 1999), and many other areas. These adverse effects include biased estimates, masking of structural changes, and creating seemingly nonlinear structures, for instance.

Discrete Variables

To achieve a HBP, many robust methods such as LMS often eliminate a large portion of observations from the calculation of their objective function. This can cause nonidentification of parameters associated with categorical variables. For example, having data on income $\{y_i\}_{i=1}^n$ of men and women, where gender is indicated by $\{d_i\}_{i=1}^n \in \{0, 1\}$, one can estimate the mean income of men and women by a simple regression model $y_i = a + bd_i$. If a HBP method such as LMS or LTS is used to estimate the model and it

R

eliminates a large portion of observations from the calculation (for example, one half of them), the remaining data could easily contain only income of men or only income of women, and consequently the mean income of one of the groups could not be then identified. Even though this seems unlikely in our simple example, it becomes more pronounced as the number of discrete variables grows (see Hubert and Rousseeuw 1997, for an example).

A common strategy employs a robust estimator with a HBP for a model with only continuous variables, and using this initial estimate, the model with all variables is estimated by an *M*-estimator. Such a combined procedure preserves the breakdown point of the HBP estimator: even though a misclassified values of categorical explanatory variables can bias the estimates, this bias will be bounded in common models as the categorical variables are bounded as well. See Hubert and Rousseeuw (1997) and Maronna and Yohai (2000), who combine an initial *S*-estimator with an *M*-estimator.

Time Series

In time series, there are several issues not addressed by the standard theory of robust estimation because of time-dependency of observations. First, the asymptotic behaviour of various robust methods has to be established; see Koenker and Machado (1999) and Koenker and Xiao (2002) for L_1 regression; Künsch (1984) and Bai and Wu (1997) for *M*-estimators and Sakata and White (2001), Zinde-Walsh (2002) and Čížek (2006) for various *S*-type estimators. In these cases, the results are usually established for general nonlinear models.

Second, the effects of data contamination are more complex and widespread due to time dependency: an error in one observation is transferred, by means of a model, to others close in time. The possible effects of outliers in time series are elaborated by Chen and Liu (1993) and Tsay et al. (2000), for instance. The first work also offers a sequential identification of outliers (an alternative procedure based on τ-estimators is offered by Bianco et al. 2001). Consequently, the robust properties in time series differ from those experienced in cross-sectional data. For example, the breakdown point is asymptotically zero in the case of *M*-estimators (Sakata and White 1995) and can be much below 0.5 for various *S*-estimators (Genton and Lucas 2003).

A further issue specific to time series is testing for stationarity of a series. Effects of outliers are in this respect similar to those of neglected structural changes. To differentiate between random outliers and real structural changes, robust tests for change-point detection have been proposed by Gagliardini et al. (2005) and Fiteni (2002, 2004); the last of these papers uses τ-estimation. The asymptotics of *M*-estimators under unit-root assumption and the corresponding tests have been established, for example, by Lucas (1995), Koenker and Xiao (2004), and Haldrup et al. (2005). An early reference is Franke et al. (1984).

Multivariate Regression

An important application of robust methods in economics concerns the multivariate regression case. This is relatively straightforward with exogenous explanatory variables only, see Koenker and Portnoy (1990), Bilodeau and Duchesne (2000) and Lopuhaä (1992) for the *M*-, *S*- and τ-estimation, respectively. Estimating general simultaneous equations models has to mimic either three-stage LS or full information MLE (Maronna and Yohai 1997). Whereas Koenker and Portnoy (1990) follow with the weighted LAD the first approach, Krishnakumar and Ronchetti (1997) use *M*-estimation together with the second strategy.

General Estimation Principles

There are naturally many more model classes for which one can construct robust estimation procedures. Since most econometric models can be estimated by means of MLE or GMM, it is however easier to concentrate on robust counterparts of these two estimation principles. There are other estimation concepts, such as nonparametric smoothing, that can employ robust estimation (Härdle 1982), but they go beyond the scope of this article.

First, recent contributions to robust MLE can be split to two groups. One simply defines a weighted maximum likelihood, where weights are computed from an initial robust fit (Windham 1995; Markartou et al. 1997). Alternatively, some erroneous observations can be excluded completely from the likelihood function (Clarke 2000; Marazzi and Yohai 2004). This approach requires existence of an initial robust estimate, and thus it is not useful for models for which there are no robust methods available. The second approach is motivated by the *S*estimation, namely, LTS, and defines the maximum trimmed likelihood as an estimator maximizing the product of the *h* largest likelihood contribution, that is, those corresponding only to *h* most likely observations (Hadi and Luceño 1997). This estimator has been studied mainly in the context of generalized linear models (Müller and Neykov 2003), but its consistency is established in a much wider class of models (Čížek 2004).

Second, more widely used GMM has also attracted attention from its robustness point of view. A special case – instrumental variable estimation – has been studied, for example, by Wagenvoor and Waldman (2002) and Kim and Muller (2007). See also Chernozhukov and Hansen (2006) for instrumental variable quantile regression. More generally, Ronchetti and Trojani (2001) have proposed an *M*-estimation-based generalization of GMM, studied its robust properties, and designed corresponding tests. This work became a starting point for others, who have extended the methodology of Ronchetti and Trojani (2001) to robustify simulation-based methods of moments (Genton and Ronchetti 2003; Ortelli and Trojani 2005).

See Also

- ▶ Adaptive Estimation
- ▶ Categorical Data
- ▶ Computational Methods in Econometrics
- ▶ Generalized Method of Moments Estimation
- ▶ Maximum Likelihood
- ▶ Measurement Error Models
- ▶ Time Series Analysis
- ▶ Two-Stage Least Squares and the *K*-Class Estimator

Acknowledgment This work was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 ‘Economic Risk’.

Bibliography

Atkinson, A., S. Koopman, and N. Shephard. 1997. Detecting shocks: Outliers and breaks in time series. *Journal of Econometrics* 80: 387–422.

Bai, Z.D., and Y. Wu. 1997. General M-estimation. *Journal of Multivariate Analysis* 63: 119–135.

Balke, N., and T. Fomby. 1994. Large shocks, small shocks, and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics* 9: 181–200.

Berrendero, J., and R. Zamar. 2001. Maximum bias curves for robust regression with non-elliptical regressors. *Annals of Statistics* 29: 224–251.

Bianco, A., M. Ben, E. Martínez, and V. Yohai. 2001. Regression models with ARIMA errors. *Journal of Forecasting* 20: 565–579.

Bilodeau, M., and P. Duchesne. 2000. Robust estimation of the SUR model. *The Canadian Journal of Statistics* 28: 277–288.

Chen, C., and L.-M. Liu. 1993. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88: 284–297.

Chernozhukov, V., and C. Hansen. 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132: 491–525.

Čížek, P. 2004. General trimmed estimation: Robust approach to nonlinear and limited dependent variable models. Discussion Paper No. 2004/130. CentER: Tilburg University.

Čížek, P. 2006. Least trimmed squares in nonlinear regression under dependence. *Journal of Statistical Planning and Inference* 136: 3967–3988.

Clarke, B. 2000. An adaptive method of estimation and outlier detection in regression applicable for small to moderate sample sizes. *Probability and Statistics* 20: 25–50.

Croux, C., P. Rousseeuw, and O. Hossjer. 1994. Generalized S-estimators. *Journal of the American Statistical Association* 89: 1271–1281.

Davies, L., and U. Gather. 2005. Breakdown and groups. *Annals of Statistics* 33: 988–993.

Fiteni, I. 2002. Robust estimation of structural break points. *Econometric Theory* 18: 349–386.

Fiteni, I. 2004. τ-estimators of regression models with structural change of unknown location. *Journal of Econometrics* 119: 19–44.

Franses, P., T. Kloek, and A. Lucas. 1999. Outlier robust analysis of longrun marketing effects for weekly scanning data. *Journal of Econometrics* 89: 293–315.

R

Franses, P., D. van Dijk, and A. Lucas. 2004. Short patches of outliers, ARCH and volatility modelling. *Applied Financial Economics* 14: 221–231.

Franke, J., W. Härdle, and R. Martin. 1984. *Robust and nonlinear time series analysis*. Berlin: Springer.

Gagliardini, P., F. Trojani, and G. Urga. 2005. Robust GMM tests for structural breaks. *Journal of Econometrics* 129: 139–182.

Genton, M., and A. Lucas. 2003. Comprehensive definitions of breakdown-points for independent and dependent observations. *Journal of the Royal Statistical Society* Series B 65: 81–94.

Genton, M., and E. Ronchetti. 2003. Robust indirect inference. *Journal of the American Statistical Association* 98: 67–76.

Gervini, D., and V. Yohai. 2002. A class of robust and fully efficient regression estimators. *Annals of Statistics* 30: 583–616.

Hadi, A., and A. Luceño. 1997. Maximum trimmed likelihood estimators: A unified approach, examples, and algorithms. *Computational Statistics and Data Analysis* 25: 251–272.

Haldrup, N., A. Montans, and A. Sanso. 2005. Measurement errors and outliers in seasonal unit root testing. *Journal of Econometrics* 127: 103–128.

Hampel, F. 1971. A general qualitative definition of robustness. *Annals of Mathematical Statistics* 42: 1887–1896.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel. 1986. *Robust statistics: The approach based on influence function*. New York: Wiley.

Härdle, W. 1982. Robust regression function estimation. *Journal of Multivariate Analysis* 14: 169–180.

He, X., and D. Simpson. 1993. Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Annals of Statistics* 21: 314–337.

Huber, P. 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35: 73–101.

Hubert, M., and P. Rousseeuw. 1997. Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference* 57: 153–163.

Kim, T.-H., and C. Muller. 2007. Two-stage Huber estimation. *Journal of Statistical Planning and Inference* 137: 405–418.

Koenker, R., and J. Machado. 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* 94: 1296–1310.

Koenker, R., and S. Portnoy. 1990. M-estimation of multivariate regressions. *Journal of the American Statistical Association* 85: 1060–1068.

Koenker, R., and Z. Xiao. 2002. Inference on the quantile regression process. *Econometrica* 70: 1583–1612.

Koenker, R., and Z. Xiao. 2004. Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99: 775–787.

Krishnakumar, J., and E. Ronchetti. 1997. Robust estimators for simultaneous equations models. *Journal of Econometrics* 78: 295–314.

Kunsch, H. 1984. Infinitesimal robustness for autoregressive processes. *Annals of Statistics* 12: 843–863.

Lopuhaä, H. 1992. Multivariate τ-estimators. *The Canadian Journal of Statistics* 19: 307–321.

Lucas, A. 1995. An outlier robust unit root test with an application to the extended Nelson–Plosser data. *Journal of Econometrics* 66: 153–173.

Marazzi, A., and V. Yohai. 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference* 122: 271–291.

Markatou, M., A. Basu, and B. Lindsay. 1997. Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference* 57: 215–232.

Maronna, R., O. Bustos, and V. Yohai. 1979. Bias- and efficiency-robustness of general *M*-estimators for regression with random carriers. In *Smoothing techniques for curve estimation*, ed. T. Gasser and M. Rossenblatt. Berlin: Springer.

Maronna, R., and V. Yohai. 1997. Robust estimation in simultaneous equations models. *Journal of Statistical Planning and Inference* 57: 233–244.

Maronna, R., and V. Yohai. 2000. Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89: 197–214.

Müller, C., and N. Neykov. 2003. Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference* 116: 503–519.

Ortelli, C., and F. Trojani. 2005. Robust efficient method of moments. *Journal of Econometrics* 128: 69–97.

Ronchetti, E., and F. Trojani. 2001. Robust inference with GMM estimators. *Journal of Econometrics* 101: 37–69.

Rousseeuw, P., and V. Yohai. 1984. Robust regression by means of *S*-estimators. In *Robust and nonlinear time series analysis*, ed. J. Franke, W. Härdle, and R. Martin. Heidelberg: Springer.

Sakata, S., and H. White. 1995. An alternative definition of finite-sample breakdown point with application to regression model estimators. *Journal of the American Statistical Association* 90: 1099–1106.

Sakata, S., and H. White. 1998. High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* 66: 529–567.

Sakata, S., and H. White. 2001. S-estimation of nonlinear regression models with dependent and heterogeneous observations. *Journal of Econometrics* 103: 5–72.

Sapra, S. 2003. High-breakdown point estimation of some regression models. *Applied Economics Letters* 10: 875–878.

Simpson, D., D. Ruppert, and R. Carroll. 1992. On one-step GM estimates and stability of inferences in linear regression. *Journal of the American Statistical Association* 87: 439–450.

Stromberg, A., O. Hossjer, and D. Hawkins. 2000. The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association* 95: 853–864.

Tsay, R., D. Pena, and A. Pankratz. 2000. Outliers in multivariate time series. *Biometrika* 87: 789–804.

Wagenvoor, R., and R. Waldman. 2002. On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics* 106: 297–324.
Welsh, A., and E. Ronchetti. 2002. A journey in single steps: Robust one-step Mestimation in linear regression. *Journal of Statistical Planning and Inference* 103: 287–310.
Windham, M. 1995. Robustifying model fitting. *Journal of the Royal Statistical Society* Series B 57: 599–609.
Yohai, V., and R. Zamar. 1988. High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association* 83: 406–413.
Yohai, V., and R. Zamar. 1993. A minimax-bias property of the least a-quantile estimates. *Annals of Statistics* 21: 1824–1842.
Zaman, A., P. Rousseeuw, and M. Orhan. 2001. Econometric applications of highbreakdown robust regression techniques. *Economics Letters* 71: 1–8.
Zinde-Walsh, V. 2002. Asymptotic theory for some high breakdown point estimators. *Econometric Theory* 18: 1172–1196.

Rodbertus, Johann Karl (1805–1875)

M. Falkus

Rodbertus is chiefly remembered as a pioneer of the theory of state socialism. Born on 12 August 1805, at Griefswald on the Baltic, he came from a wealthy and intellectual background (his father was a professor). Rodbertus studied law at the universities of Gottingen and Berlin. After a period of service with the Russian Government he settled in 1836 at his country estate in Jagetzow in Pomerania and concentrated on the study of social and economic issues.

Rodbertus was strongly influenced by the writings of Sismondi and, like Sismondi, he was a distinguished historian. He drew analogies between modern capitalism and ancient serfdom, and he was the first continental writer to consider explicitly the grievances and claims of the working-classes. For Rodbertus, labour was the only true source of productive wealth. His fundamental proposition, taken from Ricardo, was that the working-classes would always receive only a subsistence wage: the 'iron law' of wages. Hence, any growth of national income would inevitably increase the share of rent and profits, whereas the proportion going to wage-earners would fall. This in turn would produce recurrent economic crises as consumption would fail to match output due to lack of demand. Rodbertus may thus be considered as a continental forerunner of underconsumptionist theorists.

Rodbertus considered that the permanent misery inflicted upon the working classes should be alleviated by the state. The state should ensure that the proportion of national wealth going to workers should rise alongside that going to capitalists; the state should fix minimum and maximum wages, fix the length of a normal working day, and determine the amount of work that could be done. Yet he was conservative in politics just as he was socialist in economics. He wished to maintain established laws of capital and land ownership, and to preserve the monarchy. He believed in gradualism, arguing that it might take five hundred years to educate the population for socialism, and he put his trust in the benevolence of existing state institutions. Indeed, he warned against the dangers of directing energies into movements for political liberalism, arguing that 'the tyranny of misery is felt much more deeply than the misery of tyranny'.

Rodbertus further refined his underconsumptionist approach to economic crises by relating such crises to imperialism. He suggested that falling home demand and glutted markets would lead capitalists to seek new markets in non-industrial countries. In turn the superior military strength of the industrial countries would threaten the independence of these new regions as capitalists sought to open them up to trade.

Despite Rodbertus' undoubted claims as a forerunner of theories of underconsumption, imperialism, and state socialism, he cannot be considered a major figure in the history of social and economic thought. Neither in terms of the intellectual force in his theories nor of his influence on others has his work been profound. He failed to link his underconsumptionist theories with any explanation of why capitalists' income should not create demand; he ignored the effect of continuing demand for labour or long-term real wages; and he had no explanation of why

R

economic crises should recur periodically. Ferdinand Lassalle took up Rodbertus' 'iron law' of wages, but both Marx and Engels rejected his theories (with the consequence that Rodbertus has never found a favoured place in Marxist histories of socialism). Moreover Rodbertus had little influence outside Germany, and few of his works have been translated into other languages. According to Landauer, Rodbertus 'was one of the lost prophets who have been so frequent in the history of socialism'.

Rodbertus had a brief career in public life: he was elected to the Prussian National Assembly in 1848 and, for two weeks, became Minister of Worship and Education. But he retired from the Assembly in 1849 and returned to writing and study. He died in Jagetzow on 6 December 1875.

Selected Works

1842. *Zur Erkenntnis unsurer Staatswirthschaftlichen Zustände*. Neubrandenburg.
1850–51. *Sociale Briefe an von Kirchmann*.
1872–85. *Zur Beleuchtung der Socialen Frage*.
1884. *Das Kapital* (unfinished).

References

Dietzel, H. 1886–8. *Karl Rodbertus, Darstellung seines Lebens und seiner Lehre*, 2 vols. Jena: G. Fischer, 1923.
Gonner, E.C.K. 1890. *The social philosophy of Rodbertus*. London: Macmillan.
Plekhanov, G.V. 1923. *Ekonomicheskaia teoriia Karla Rodbertusa – Iagetsova*. Moscow: Sochinenii I.

Rogers, James Edwin Thorold (1823–1890)

O. Kurer

Keywords
Economic history; English Historical School; Rogers, J. E. T.

JEL Classifications
B31

Rogers was educated at King's College London, and Magdalen College, Oxford. From 1859 until his death he held the first Tooke Professorship of Statistics and Economic Science at King's College London. In 1862 he was elected Drummond Professor of Political Economy in the University of Oxford, a post he lost in 1868 largely because of his outspoken radical views, but to which he was re-elected in 1888. He was ordained but abandoned the clerical profession. From 1880 to 1886 he served as a rather inconspicuous member of the House of Commons.

His chief work is his monumental *History of Agriculture and Prices,* where he did much to turn economic history into the field of distribution and attempted to use more exact methods in economic historical investigations on a large scale. His work is marred by his casual deductions. He argued for a high standard of living of the English labourer during the Middle Ages and explained the subsequent deterioration by legislative interference by the landowners controlling the government.

Politically, he was greatly influenced by his friend and brother-in-law Richard Cobden. He was firmly opposed to extensive government intervention. He did however support trade unions as providing the remedy for nearly all social ills. His advocacy of laissez-faire separates him from the rest of the English Historical School, his allies in attacking theoretical economics in looking to economic history as a realistic foundation for the proper understanding and solution of contemporary social and economic problems.

Selected Works

1884. *Six centuries of work and wages: The history of English labour.* London: Swan Sonnenschein.
1886–1902. *A history of agriculture and prices in England. From the year after the Oxford parliament* (*1259*) *to the commencement of the continental war* (*1793*), 7 *vols.* Oxford: Clarendon.

1888. *The economic interpretation of history.* New York: Putnam.

1892. *The industrial and commercial history of England,* ed. A.G.L. Rogers. New York: Putnam. Published posthumously.

Bibliography

Ashley, W.J. 1889. James E. Thorold Rogers. *Political Science Quarterly* 4: 381–407.

De Marchi, N.B. 1976. On the early dangers of being too political an economist: Thorold Rogers and the 1868 election to the Drummond Professorship. *Oxford Economic Papers* 28: 364–380.

Hewins, W.A.S. 1897. James Edwin Thorold Rogers. In *Dictionary of national biography*. Oxford: Oxford University Press.

Wood, J.C. 1983. *British economists and the empire, 1860–1914*. Beckenham/Kent: Croom Helm.

Roos, Charles Frederick (1901–1958)

Karl A. Fox

Keywords

Econometric Society; Econometrics; Mathematical economics; Roos, C. F.

JEL Classifications

B31

Born on 18 May 1901, in New Orleans, Roos completed his Ph.D. in mathematics at Rice Institute in 1926. Influenced directly by his supervisor Evans (1922, 1924, 1930) and indirectly by Volterra, his main interests in graduate work were the calculus of variations, integral equations, and applications of those areas of mathematics to problems in dynamic economics.

Although he published several brilliant articles (Roos 1925, 1927a, b, c, 1928, 1930), Roos found no journal which would readily accept manuscripts in which he combined economics, mathematics and sometimes statistics at suitably advanced levels (cf. Roos 1934, p. xiii). Spurred by similar frustrations, Frisch and Roos jointly took the initiative which led to creation of the Econometric Society in 1930 (of which Roos became President in 1948) and publication of its journal, *Econometrica,* from 1933 on.

In 1930 Roos set out to write a treatise on dynamic economics; he published an important book under that title in 1934. It was reviewed enthusiastically by Tintner (1936) and uncomprehendingly by Freeman (1935). *Dynamic Economics* (1934) is a brilliant combination of mathematical economic theory and applied econometrics. Roos's mathematical approach inspired Tintner to write a dozen articles on dynamic economic theory (for example, Tintner 1938).

Roos held a series of administrative positions during 1931–7 and published a major book on *NRA Economic Planning* (1937). In 1938 he founded an econometric consulting firm in New York and directed it until his death. Examples of his later work are Roos and von Szeliski (1939a, b) and Roos (1955, 1957). He died in Milwaukee on 7 January 1958.

Hotelling (1958) describes Roos as 'a unique and outstanding figure', while Davis (1958) presents a complete list of his writings.

Selected Works

1925. A mathematical theory of competition. *American Journal of Mathematics* 47: 163–175.

1927a. Dynamical economics. *Proceedings of the National Academy of Sciences* 13: 145–150.

1927b. A dynamical theory of economic equilibrium. *Proceedings of the National Academy of Sciences* 13: 280–285.

1927c. A dynamical theory of economics. *Journal of Political Economy* 35: 632–656.

1928. A mathematical theory of depreciation and replacement. *American Journal of Mathematics* 50: 147–157.

1930. A mathematical theory of price and production fluctuations and economic crises. *Journal of Political Economy* 38: 501–522.

1934. *Dynamic economics: Theoretical and statistical studies of demand, production and prices,* Cowles Commission monograph no. 1. Bloomington: Principia Press.
1937. *NRA economic planning,* Cowles Commission monograph no. 2. Bloomington: Principia Press.
1939a. (With V. von Szeliski.) *The dynamics of automobile demand.* Detroit: General Motors Corporation.
1939b. (With V. von Szeliski.) The concept of demand and price elasticity; the dynamics of automobile demand. *Journal of the American Statistical Association* 34: 652–666.
1955. Survey of economic forecasting techniques. *Econometrica* 23: 363–695.
1957. *Dynamics of economic growth: The American economy, 1957–1975.* New York: Econometric Institute.

Bibliography

Davis, H.T. 1958. Charles Frederick Roos. *Econometrica* 26(4), 580–589. Contains a complete bibliography of Roos's writings (91 items).
Evans, G.C. 1922. A simple theory of competition. *American Mathematical Monthly* 29(10): 371–380.
Evans, G.C. 1924. The dynamics of monopoly. *American Mathematical Monthly* 31: 77–83.
Evans, G.C. 1930. *Mathematical introduction to economics*. New York: McGraw-Hill.
Freeman, H.A. 1935. Review of C.F. Roos, dynamic economics. *American Economic Review* 25: 520.
Hotelling, H. 1958. C.F. Roos, econometrician and mathematician. *Science* 128: 1194–1195.
Tintner, G. 1936. Review of dynamic economics. *Journal of Political Economy* 44: 404–409.
Tintner, G. 1938. The theoretical derivation of dynamic demand curves. *Econometrica* 6: 375–380.

Röpke, Wilhelm (1899–1966)

Josef Molsberger

German economist and social philosopher, a leading figure of German Neo-liberalism, Röpke was born on 10 October 1899 at Schwarmstedt (Hannover) and died on 12 February 1966 at Geneva. Obtaining the *Dr.rer.pol.* (1921) and *Habilitation* (1922) at the University of Marburg, he became professor of economics at Jena (1924), Graz (1928) and again Marburg (1929). A liberal adversary of National Socialism, Röpke was ousted from office for political reasons in 1933 and went into exile. He was professor at the University of Istanbul (1933–37) and at the Graduate Institute for International Studies, Geneva (1937–66).

Röpke's scholarly work was centred on applied economics, rather than pure theory, and on the economic order, including the political, social and philosophical foundations of the market economy. His numerous publications reflect the chronology of major problems in German and international economic policy. As a member of the Brauns Commission to advise the German government (1931) and in his 1932 book he advocated, at the right moment, a 'Keynesian' policy before Keynes: a government investment programme, financed by credit expansion, to provide the *Initialzündung* (initial ignition) for overcoming the depression. Röpke later became a critic of the inflationary bias in Keynesianism.

Rejecting laissez-faire no less than central planning, Röpke conceived an economic order that supposed government not only to set the rules of the game (*Ordnungspolitik*) but included also decentralization, deconcentration, environmental policy, and a 'liberal interventionism' backing up market forces by adjustment assistance and not interfering with the price mechanism ('market-conformity principle'). As one of the intellectual architects of the 'Social Market Economy' and an adviser to Minister Ludwig Erhard, Röpke exerted considerable influence on post-war German economic policy.

Selected Works

1929. Staatsinterventionismus. In *Handwörterbuch der Staatswissenschaften*, Supplementary volume, 4th ed. Jena: Gustav Fischer.
1931. Praktische Konjunkturpolitik: Die Arbeit der Brauns-Kommission. *Weltwirtschaftliches Archiv.*

1932. *Krise und Konjunktur*. Leipzig: Quelle & Meyer, Trans. as *Crises and cycles*. London: W. Hodge, 1936.
1937. *Die Lehre von der Wirtschaft*, 12th ed. Bern/Stuttgart: Paul Haupt, 1979. Trans. as *Economics of the free society*. Chicago: Regnery, 1963.
1942a. *Die Gesellschaftskrisis der Gegenwart*, 6th ed. Bern/Stuttgart: Paul Haupt, 1979. Trans. as *The social crisis of our time*. Chicago: Regnery, 1950.
1942b. *International economic disintegration*. London.
1944. *Civitas Humana: Grundfragen der Gesellschafts- und Wirtschaftsreform*, 4th ed. Bern/Stuttgart: Paul Haupt, 1979. English trans. as *Civitas Humana*. London: W. Hodge, 1949.
1945a. *Internationale Ordnung*. Erlenbach-Zürich: Eugen Rentsch.
1945b. *Die deutsche Frage*. Erlenbach-Zürich: Eugen Rentsch. Trans. as *The German question*. London: G. Allen & Unwin, 1946.
1947. Repressed inflation. *Kyklos* 1(3): 242–253.
1950a. *Ist die deutsche Wirtschaftspolitik richtig?* Stuttgart: Kohlhammer.
1950b. *Mass und Mitte*, 2nd ed. Bern/Stuttgart: Paul Haupt, 1979.
1954. *Internationale Ordnung – Heute*, 3rd ed. Bern/Stuttgart: Paul Haupt, 1979.
1958. *Jenseits von Angebot und Nachfrage*, 5th ed. Bern/Stuttgart: Paul Haupt, 1979.
1959. In *Gegen die Brandung* (Collected articles), ed. A. Hunold. Erlenbach-Zürich/Stuttgart: Eugen Rentsch. Trans. as *Against the tide*. Chicago: Regnery, 1969.
1962. *Wirrnis und Wahrheit* (Collected articles). Erlenbach-Zürich/Stuttgart: Eugen Rentsch.
1979. In *Ausgwählte Werke in sechs Bänden*, ed. F.A. von Hayek, H. Sieber, E. Tuchtfeldt, H. Willgerodt. Bern/Stuttgart: Paul Haupt. (Comprising Röpke (1937), 12th ed; (1942a), 6th ed; (1944), 4th ed; (1954), 3rd ed; (1958), 5th ed; (1950b), 2nd ed.) Complete list of works 1920–1968 and bibliography in *In memoriam Wilhelm Röpke*, ed. E. Hoppmann. Marburg: N.G. Elwert, 1968.
Supplement 1967–1976 to list of works and bibliography in *Marburger Gelehrte in der ersten Hälfte des 20. Jahrhunderts,* ed. I. Schnack. Marburg: N.G. Elwert, 1977.

Roscher, Wilhelm Georg Friedrich (1817–1894)

B. Schefold

Keywords
Cameralism; German Historical School; History of economic thought; Methodology of economics; Self-interest; Stages of economic development

JEL Classifications
B31

Roscher was born in Hannover into a well-established civil service family. He studied history and political science in Göttingen and Berlin. In 1840 he became lecturer in both subjects at Göttingen, in 1843 he was appointed extraordinary professor of political economy, and in the next year was promoted professor. In 1848 he transferred to Leipzig, where he taught for the rest of his life. Roscher had a Protestant background and was deeply religious, adhering to a rather 'primitive form of religious belief' (Max Weber 1903–6).

Roscher may be considered as one of the most important German economists of his time. He was one of the founders and the leading exponent of the German 'older' Historical School. He did not develop any new theory: his main contribution to political economy lay in the field of method. He adhered to what he called the 'historical-physiological method', as opposed to the 'idealistic method' (1842; 1854–94, vol. 1, pp. 43–56). This inductive method intended to provide a description of the actual course of economic development and of real economic life. Thus,

R

Roscher tried to analyse laws of economic development by comparing the history of different people and nations and showing analogies in stages of their development. The emphasis was on historical relativism: economic behaviour depended to a large extent on the specific national and historic conditions of the different people and nations. This implied that a nation had to be regarded as a whole, as an 'organic unity', and not as the mere sum of individuals.

This was opposed to what Roscher called the 'idealistic method', which intended to provide an ideal picture, logically derived from abstract principles, of the functioning of the economic system. An example of this was the classical economists' deduction of economic laws from a system of hypotheses. Although Roscher emphasized that in economic analysis there existed generally no definite causal relationships but reciprocal ones, he did not reject the existence of 'laws of motion' within economic life. However, these laws were distinct from laws of natural science in that they dealt with free human beings gifted with reason and hence with changing motives for action (1854–94, vol. 1, pp. 26–9). Roscher was closer to the theoretical system of the classics than the exponents of the 'younger' historical school. He tended to regard it as the appropriate system of analysis of the current stage of economic development. He only modified and supplemented it with a careful historical analysis, but he may still be regarded as being in the classical tradition.

The first volume of Roscher's main work, *System der Volkswirtschaft* (1854–94: 1854) still looked very much like a traditional textbook. It analysed essentially the same topics as the classical economists – production, distribution and prices. Roscher was already strongly influenced by supply and demand approaches, but still determined the exchange value of a commodity by its cost of production. His theory of rent was Ricardian and his thinking about population development followed Malthusian patterns. Differing from classical textbooks, Roscher supplemented the theoretical analysis with a historical description – the reader finds the history of rent, interest and wages, of population development, of the prices of necessary and luxury commodities, and of luxury in general. Roscher accepted the classical notion of individual self-interest as a central axiom of modern economic behaviour, but he did not follow the classical patterns in deriving his economic principles from this assumption. As a result of his religious beliefs, he included human conscience as a regulating mechanism into his analysis of the role of self-interest (1854–94, vol. 1, pp. 20–3).

The other four volumes of the *System der Volkswirtschaft* (1859; 1881; 1886; 1894), which may be perceived as his main contribution to applied economics, were even more historically oriented and focused on agriculture, trade and industry, public finance, social policy and poor relief.

Roscher classified economic development into stages of maturity. The economic factors that govern the development of nations were land, labour and capital which subsequently dominated the different stages (1861, ch. 1). Later, Roscher presented a more detailed analysis of stages of political and societal development (1892) on the basis of a classification of the different forms of government during history: early patriarchal kingdom, aristocracy of knights and priests, absolute monarchy, democracy. The latter then degenerated into a plutocracy, which is followed by a military dictatorship Roscher called 'Caesarismus'. Roscher did not systematically attempt an integration of his theory of political development and the stages of economic evolution.

He wrote several contributions on the history of economic thought. His compendium on the history of political economy in Germany (1874) was his most outstanding work and has remained important. Roscher may be regarded as the most eminent historian of cameralism and early German political economy. His treatise on economic problems of the location of large towns (1871) was an original contribution to economic theory.

Roscher supported German imperialism. In order to secure raw materials and markets for German goods, as well as to relieve the national labour market and prevent social unrest, he advocated an expansive German colonial policy, especially towards Eastern Asia, where he saw Germany's colonial future (1885). He was a conservative but he remained all his life unaffiliated to any political party or group.

See Also

► Historical School, German

Selected Works

1842. *Leben, Werk und Zeitalter des Thukydides*. Göttingen.

1854–94. *System der Volkswirtschaft*. Stuttgart: Cotta. Vol. 1: *Die Grundlagen der Nationalökonomie*, 1854. Trans. from 13th edn by J.J. Lalor as *Principles of political economy*, 2 vols, New York, 1878. Vol. 2: *Nationalökonomik des Ackerbaues und der verwandten Urproduktionen*, 1859. Vol. 3: *Nationalökonomik des Handels und Gewerbefleißes*, 1881. Vol. 4: *System der Finanzwissenschaft*, 1886. Vol. 5: *System der Armenpflege und der Armenpolitik*, 1894.

1861. *Ansichten der Volkswirtschaft aus dem geschichtlichen Standpunkt*. Leipzig/Heidelberg: Winter.

1871. *Betrachtungen über die geographische Lage der grossen Städte*. Leipzig.

1874. *Geschichte der Nationalökonomik in Deutschland*. Munich: Oldenbourg.

1885. *Kolonien, Kolonialpolitik und Auswanderung*. Leipzig: Winter Part II, ch. 1. Translated from 3rd edn by E.H. Baldwin and E.G. Bourne as *The Spanish Colonial System*, New York, 1904.

1892. *Politik: Geschichtliche Naturlehre der Monarchie, Aristokratie und Demokratie*. Stuttgart.

Bibliography

Cunningham, W. 1894–5. Why had Roscher so little influence in England? *Annals of the American Academy of Political and Social Sciences* 5: 317–334.

Weber, M. 1903–6. Roscher und Knies und die logischen Probleme der Historischen Nationalökonomie. In *Gesammelte Aufsätze zur Wissenschaftslehre*. Tübingen: Mohr, 1922.

Rosen, Sherwin (1938–2001)

Robert Tamura

Abstract

Sherwin Rosen made fundamental contributions in equilibrium theory, human capital theory, income distribution theory and investment theory. One characteristic feature of Rosen's work is the minimal use of heterogeneity of individuals. His work explains price disparities, differential earnings, and investment cycles. Underlying differences in characteristics produce price differences. Human capital theory is enriched by characterization of accumulation beyond schooling. Skewed income distributions arise from outcomes of tournaments, superstars from economies of scale or hierarchical complementarities. Rational investment cycles occur when the capital stock is large relative to investment, and when the breeding stock is large relative to the overall stock.

Keywords

Agglomeration economy; American Academy of Arts and Sciences; American Economics Association; Baby boom; Compensating differentials; Differentials; Division of labour; Earnings generating function; Econometric Society; Economies of scale; Endogenous comparative advantage; Equilibrium theory; Fertility; Game theory; Giffen goods; Hedonic models; Human capital; Income and consumption inequality; Income distribution theory; Information aggregation; Investment cycles; life-cycle earnings; marginal rate of substitution; Midwest Economic Association; National Academy of Sciences; non-convexity; occupational choice; optimal knowledge accumulation; price differences; Rosen, S.; Roy model; Society of Labor Economists; superstars; symmetric ignorance; value of life

R

JEL Classifications

B31; D33; D91; D92; J24; J31; J33

Rosen was born in Chicago on 29 September 1938. He died in Chicago on 17 March 2001. He earned his BS in economics from Purdue University in 1960. He obtained his graduate economics degrees from the University of Chicago: his MA in 1962 and his Ph.D. in 1966. His first appointment was as assistant professor of economics at the University of Rochester in 1964. Promoted to associate professor in 1968 and full professor in 1970, he became the Kenan Professor of Economics in 1975. Rosen returned to the University of Chicago in 1977, and became the Bergman Professor of Economics in 1983. From 1992 until his death he served as the Edwin A. and Betty L. Bergman Distinguished Service Professor. In addition, he served as department chairman during 1988–94. He was a Senior Research Associate of the National Bureau of Economic Research from 1968 and a Senior Research Fellow of the Hoover Institution during 1983–96 before becoming a Senior Fellow in 1997.

Rosen was elected a fellow of the Econometric Society in 1976, and a fellow of the American Academy of Arts and Sciences in 1984. He became a member of the National Academy of Sciences in 1998. He was President of the Midwest Economic Association during 1996–7, President of the Society of Labor Economists in 2000, and President of the American Economics Association for 2001.

Rosen was a prolific scholar and one of the leading economists of his generation. His contributions spanned many fields, including equilibrium theory, human capital theory, income distribution theory and investment theory. His research provided the theoretical underpinnings of labour economics, urban economics and health economics. A unifying aim of his research is to explain differential market outcomes. Price differences of goods can be explained by their differential amounts of characteristics. These price differences could be driven by differences in preferences arising from wealth differences or differences in technologies available to firms. For example, cars sell for different prices because they contain different attributes, and workers earn different amounts across jobs because the jobs have different characteristics. Earnings may differ if workers differ in their human capital. Life-cycle earnings are explained from the characterization of human capital accumulation beyond formal schooling. Returns to higher education are best modelled as arising from revealed preference of workers, both college-educated and non-college educated. Earnings may differ between identical workers because they are in different job classifications. Small differences in worker productivity can manifest themselves in large earnings differences if there are production scale economies (creating superstars, for example), strong complementarities, or increasing returns in skill use. Finally, differences in returns and investments can arise from predictable future demand shifts or unpredicted contemporaneous demand shocks. Rational investment cycles are likely if investment is small relative to the stock of capital, and if the seed capital is a large proportion of the stock of capital.

Rosen was the author of two books, *A Disequilibrium Model of Demand for Factors of Production* (Nadiri and Rosen, 1973) and Markets and Diversity (2005), and editor of three collections: *Studies in Labor Markets* (1981), *Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure* (Rosen and Winship, 1988), and *Implicit Contract Theory* (1994).

Equilibrium Theory

Rosen's 1974 article 'Hedonic prices and implicit markets: product differentiation in pure competition' is the quintessential example of his work in equilibrium theory. Consider the following labour market application. Rosen's analysis allows for a job to be characterized by N dimensions, but for clarity we focus on only two, its wage and its dirtiness. Some jobs are dirtier than others. They provide meaner working conditions including unheated and/or non-air-conditioned workplaces, less pleasant coworkers, few or no promotion possibilities, high unemployment risk, large variability in hours demanded by the employer, inflexible hours of work, fewer vacations, worse fringe benefits like poor or no health insurance or disability insurance, poor pensions, and so on. Consider aggregating all of these features into a single measure called ‘dirt’. A worker likes wages and dislikes dirt. Employers can offer any combination of wages and dirt as a package to prospective workers. Assume that (*a*) worker preferences are convex; hence a worker’s dislike for dirt increases with the level of dirt on the job, and he or she requires ever larger increases in wages to accept an additional unit of dirt as the level of dirt increases; and (*b*) firm production technologies are convex; firms require greater wage reductions for a unit reduction in dirt as the job becomes less dirty. There are three extreme cases. First, if all workers are identical in wealth and preferences, and all employers have access to the same technology of production, then there is a single equilibrium point. This occurs at the tangency of the representative worker’s iso-utility curve and the representative employer’s iso-profit curve. With free entry, competition drives the equilibrium to the zero profit iso-profit curve. In the second case, suppose all workers are identical in wealth and preferences, but employers have different technologies. For example, mining firms find it more costly than software design firms to provide cleaner work environments. In equilibrium the economist will observe a locus of points, which traces out the representative worker’s iso-utility curve, and each observed increase in dirt is associated exactly with the worker’s compensating differential to accept the increased dirtiness. Finally, suppose workers have different preferences, say because of wealth differences. Assume that dirt is an inferior good. Let all firms have access to a single technology. In equilibrium the economist observes a locus of points, which traces out the representative firm’s zero iso-profit curve. The second example identifies preferences, the third example identifies technology. Of course, the world is not so stark or clean for an economist. Preferences are heterogeneous, workers have differing skill levels, firms have different technologies. Thus, econometrically the problem is one of finding controls that allow for identification (see Ekeland et al. 2004). One important application to the labour market is Murphy (a Rosen student) and Topel (1987).

Rosen’s (1974) paper serves as the benchmark for thinking about how markets link customers of multiple characteristic goods and services with the suppliers of these complex goods and services. One important application of this model is by Roback (1982), a Rosen student. Her model examines the compensating differentials in worker wages and land rents arising from differences in location-specific amenities, say, climate or population density. Another application of this hedonic approach is the examination of the increased wages that firms pay to workers in order to induce them to accept greater risks to their health or, in particular, their lives. Rosen and Thaler (1976) allow variation in earnings due to variation in on-the-job risks to life, controlling for productivity (schooling and experience) and in other job characteristics to identify the reservation price of mortality risk for the typical worker. This allows for the calculation of the economic value of a life. Rosen (1988) revisits this arena by examining the valuation placed on increasing longevity. These two papers served as inspiration for an entire sub-field of health economics, highlighted by Murphy and Topel (2003).

Rosen (1978) examines the assignment solution of workers to tasks within an organization, in a world with a fixed number of inputs and many worker types. Rosen shows that the division of labour corresponding to the optimum assignment determines the marginal rates of substitution between worker types or between job categories. Thus the division of labour determines the extent of product and factor market substitutions in the economy. This paper provides an application of economics to the optimal determination of job types, or the efficient bundling of activities into a job. Rosen (1982b) extends the analysis. It is further generalized in Tamura (1992); with a continuum of intermediate tasks, and N different worker types, each of measure 1, output can be shown to come from the following reduced form:

$$Y = \left\{ \sum_{i=1}^{N} h_i^{\rho} \right\}^{\frac{1}{\rho}} \quad (1)$$

where type i workers have h_i units of human capital and $0 < \rho < 1$. With each individual a set of measure zero, each worker is paid the marginal product of his or her human capital and, given the constant returns to scale in the distribution of human capital, output is completely exhausted. However, since $\rho < 1$, there is an agglomeration economy in participation. Earnings for an individual of type j are the product of the marginal product of human capital of type j workers, w_j, and the amount of human capital of type j workers, h_j, or:

$$\begin{aligned} y_j = w_j h_j &= \left\{ \sum_{i=1}^{N} h_i^{\rho} \right\}^{\frac{1-\rho}{\rho}} h_i^{\rho-1} h_j \\ &= \left\{ \sum_{i=1}^{N} h_i^{\rho} \right\}^{\frac{1-\rho}{\rho}} h_i^{\rho} \end{aligned} \quad (2)$$

Assume that the human capital of worker type j grows at rate λ_j. Suppose workers of type j have more human capital than workers of type i. If the growth rates of human capital differ across type, then the relative earnings of these two worker types will change. In particular, notice:

$$\frac{y_{jt}+1}{y_{it}+1} = \left(\frac{\lambda_j h_{jt}}{\lambda_i h_{it}} \right)^{\rho} = \left(\frac{\lambda_j}{\lambda_i} \right)^{\rho} \frac{y_{jt}}{y_{it}} \quad (3)$$

Thus earnings become more (less) unequal if $\lambda_j > (<) \lambda_i$. Nothing about differences in firm investments in technical change is required, merely differences in the abilities of workers to continue to accumulate human capital. As Rosen (1972a, b) notes, higher education can help individuals become better learners. Thus rising wage inequality can be the result of rising task specialization of the more skilled. Hence the works of Acemoglu (1998, 2002) can be thought of as arising from underlying primitives of differential worker abilities to learn.

Rosen also made fundamental contributions with Li, Mussa, and Suen. Mussa and Rosen (1978) provide an equilibrium analysis of the product quality choice of monopolists. Li and Rosen (1998) examine the effect of breaches of contracts, unravelling, on optimal assignments of workers to firms when worker quality is uncertain. Li, Rosen and Suen (2001) examine the role of committees in information aggregation. If individuals have idiosyncratic information, committees help to aggregate the information. However, if committee members have conflicting preferences, the only equilibrium truth-telling rules are binary: yes or no, promote or do not promote, hire or not hire, keep or fire.

Human Capital Theory

Rosen applied his characteristics approach in order to make fundamental contributions to human capital theory. Rosen (1972a) models jobs as producing both output and learning opportunities for workers. Jobs differ in their learning opportunities. These opportunities are costly to firms; they produce less market output in return for producing more skills for workers in the future. With identical workers and many firms in equilibrium, young workers seek out the firms with the best learning opportunities. Workers accept lower earnings to pay the firm for the learning opportunities associated with their job. As they gain experience and skill, but have fewer years of work remaining, they switch to jobs offering less rapid learning and greater production. The theory produces occupational switching and the typical age-earnings profile. With heterogeneity in ability, the model is capable of producing a distribution of outcomes by age. The most able learners choose jobs with the most rapid learning possibilities, while less able learners choose to forgo those jobs entirely. Rosen (1972b) displays an early grasp of dynamic programming. He formulates the optimal accumulation of knowledge from learning by producing by analysing the excess valuation of production over and above current profits for the acquisition of higher future profits. He formulates a model of optimal knowledge accumulation as an explanation for technological progress. Curiously, Rosen notes that a stationary solution to an infinite horizon problem is not possible. However modern endogenous growth models in fact do take his first functional form in the paper. As long as output grows at a constant rate, knowledge growth will continue at a constant rate. Rosen presages Romer (1986) and Lucas (1988) by arguing that knowledge creation is likely to have important spillovers across workers and industries.

In a contribution to a Feschrift volume for his advisor H. Gregg Lewis, on the occasion of his retirement from the University of Chicago, Rosen (1976) applies a novel twist on the problem of life-cycle earnings. He considers the standard formulation of time t wealth value of human capital:

$$W(t) = \int_t^N y(s) e^{-r(s-t)} ds \tag{4}$$

where N is an exogenously specified retirement age, $y(s)$ is the earnings at age s, and the individual faces a constant interest rate, r. Differentiating (4) with respect to t and rearranging produces:

$$rW(t) - \dot{W}(t) = y(t) \tag{5}$$

The standard interpretation is to consider the first term as the potential earnings, the second term as the dollar cost of human capital investment and the right-hand side as observed earnings. In his words (1976, p. S46), 'The method adopted here is to go behind the scenes of (5) and use the theory to parameterize $y(t)$ directly in the form of restrictions on the unobservable $W(t)$.'

Rosen considers the accumulation of human capital as a self-directed process as in Ben-Porath (1967); $y = Rk(1 - s)$, where s is the proportion of knowledge spent on learning. Rosen considers two possible tractable formulations on the earnings generating function: (a) the learning function depends on total capital resources spent on accumulation, $\dot{k} = h(sk)$, which produces $y = Rk - h^{-1}(\dot{k})$, and (b) the learning function is linear in the stock of knowledge, $\dot{k} = h(s)k$, which produces $y = Rk\left[1 - h^{-1}\left(\frac{\dot{k}}{k}\right)\right]$ Rosen chooses the latter functional form. The reader familiar with endogenous growth models will immediately see that his preferred specification is the Ak model of Jones and Manuelli (1990), Rebelo (1991) and Lucas (1988). Rosen also assumes that children are born with a fixed fraction of their parents' capital, such that at age 0, $k^{t+1}(0) = \gamma k^t(0) > k^t(0)$; thus he formulates Lucas (1988) without the human capital externalities, but with perpetual growth! Despite the difficulty imposed by finite time horizon models, Rosen derives closed form solutions for the optimal rate of human capital accumulation, $s(t)$, as well as for $k(t)$ and $y(t)$. Unfortunately, economists appear to have misgivings about working with hyperbolic sines, cosines and cotangents! Rosen (1976) produces the standard life-cycle shapes of observed earnings, potential earnings and human capital investments. From his explicit analytic solutions, Rosen is able to estimate the structural parameters of this model using census data. His estimates are broadly consistent with empirical results on rates of returns to schooling. More interestingly, he conducts counterfactual experiments about the nature of college. Schooling could be purely vocational in substance, increasing the knowledge of the future worker. It could also make the individual permanently more productive at future learning. Rosen posits different pairs of learning efficiencies and initial knowledge immediately after college completion that make the college graduate indifferent to college or work after high school graduation. He conducts the same counterfactual for high school graduates. Clearly, this thought experiment is one that foreshadows his seminal work with Robert Willis.

Willis and Rosen (1979) present a version of the Roy (1951) model for educational choice. Individuals can choose between stopping after high school graduation and continuing on to college. In their model there exists comparative advantage. Revealed preference implies that those workers that stopped after high school chose optimally to ignore college because their own rate of return to college education would be less than their cost of funds. Revealed preferences of college graduates imply the opposite. Now, if in addition high school graduates have an absolute advantage in high school occupations relative to what college graduates could earn in those jobs as high school graduates, estimated rates of return to college would be biased. The estimated rate of return to college would be below the true return to the college graduate, but more than the prospective return to the high school graduate. This revealed preference of educational-occupational selection indicates that there might not exist much ability bias in estimated rates of return to college. After nearly 30 years of empirical work, this is the dominant view in the economics profession (see Ashenfelter and Krueger, 1994; Ashenfelter and Zimmerman, 1997; Ashenfelter and Rouse, 1998).

Rosen (1983) identifies an increasing returns feature to human capital. The key point is that the marginal cost of creating human capital is independent of the intensity of use of human capital. That is to say, human capital investment is like a sunk cost. Once acquired, the marginal cost of using human capital is zero. The more intensely an individual uses his or her human capital, the greater the return to the human capital. Identical workers have an incentive to specialize their human capital or to endogenously choose their comparative advantage. This endogenous comparative advantage is a further extension of Willis and Rosen (1979). Thus, more specialized workers spend their careers in large markets that more fully utilize their skills. The largest metropolitan areas will be home to the most specialized human capital, in any field. Medical specialists will agglomerate in large metro areas and not smaller cities or rural areas; see Baumgardner (1988a, b), a Rosen student. The increasing returns to utilization and endogenous comparative advantage model he envisions are explored in Tamura (1992, 1996, 2002, 2006).

Income Distribution Theory

Rosen made seminal contributions to understanding the functional distribution of earnings in the economy. Underlying his work is a search for the answer to the fundamental question: ‘Why are earnings so skewed?’ Furthermore, his work operates under the constraint that the answer should arise from a minimal amount of heterogeneity in underlying individual talent. Ideally, *ex ante* identical individuals would produce the observed skewed earnings distribution. One can view Rosen’s work in human capital theory specifically as producing answers to this question with close to this ideal assumption of identical initial human capital endowments. In addition to his human capital programme, his research in this category includes Lazear and Rosen (1981) and his solo authored works (1981; 1982a; 1986a; 1997a). In Lazear and Rosen (1981) and Rosen (1986a), workers are paid as a result of internal relative comparisons. Assume that worker effort is not observable. If individual worker productivity is measured with noise, but a large proportion of that noise is common for all workers of the firm or for workers at similar levels within the firm, the use of relative productivity in order to determine compensation is efficient. This is because, by using relative comparisons, the effects of the unmeasured noise tend to be eliminated or greatly mitigated. For all workers, the wage bill must equal the value produced by the workers. However, workers are paid in relation to their place in the tournament, and hence paid in line with their job title. Increasing the spread between job levels or ranks raises the effort level of workers in the tournament. The larger the total wage bill, with the number of workers at the firm held constant, the greater is the average effort level, and the greater the average ability of workers. In noisier environments, the spread between winning and losing workers must be greater than the earnings spread in more predictable environments. This larger spread is required because increasing noise dissipates the return to worker effort. Thus, in order to elicit the same level of effort, noisier industries must have greater earnings disparity.

Lazear and Rosen (1981) deals with a single contest. However, internal hierarchies are tournaments with many rounds. As a worker successfully progresses up the job ladder, there are fewer and fewer rounds left to play. In order to maintain the efficient level of effort, the prize gap must increase. Hence the gap between the CEO and the second in command of the firm must be larger than the gap between the second in command and his or her direct subordinates. Even if the CEO is only marginally better than the second in command, this larger prize must be given in order to provide the correct incentives throughout the organization. The pay gap serves to motivate not only the CEO but all workers in the internal hierarchy, and especially those close to the CEO in rank. As Lazear (2003, p. 13) notes, '(t)he theory helps explain why there is a larger spread in earnings between the top and bottom in new industries than in old ones.' As a consequence newer industries often pay workers in stocks or stock options in order to enlist greater effort levels. When the winners of these new companies in new industries are anointed, the stock options induce huge pay differentials within and across firms in this industry. Rosen (1986a) also shows that the single elimination tournament among players with heterogeneous talents is more likely to be the efficient tournament design than a round robin format. It promotes survival of the fittest at a more rapid rate. Rosen (1986a) shows this in an environment of 'symmetric ignorance', or the 'veil of ignorance', in which all players know only the common distribution from which all players' talents are drawn, including their own. Through Bayesian updating, survival in each round provides information about the ability of the contestant. These papers were among the first to apply game theory to labour economics. Furthermore, in the conclusion, Rosen (1986a) identifies the interesting area of further research, namely, the effects of player optimism or pessimism. His conjectures again presage the seminal works of Benabou and Tirole (2002, 2003, 2004) on micro models of behavioural attitudes.

On the question of skewed income distribution from small initial differences, Rosen (1981) shows that markets where costs of reproduction are trivial overwhelmingly choose to reward the individual who is perceived to be the best, even when the best is only trivially better than the second-best performer. Hence the entertainment industry with low-cost reproduction of movie prints greatly increases the skewness of the earnings distributions of actors, producers and directors in comparison with the earnings distributions of these same labour inputs in the days of the travelling show, or the Broadway theatre. Adding books, LPs, video tapes, CDs, DVDs, and so on continues to lower the cost of 'owning' a performance. Hence, an individual perceived by the market to be the best will harvest the overwhelming bulk of the demand. The individual performance is captured or recorded once, and then can be replicated at near zero marginal cost. An additional example is the falling costs of journal publication, producing rising skewness in the earnings distribution in academics.

This research leads directly to Rosen (1982a). The CEO can supply the same effort level working for a family firm with \$1 million of revenue or a publicly traded firm with \$100 billion of revenue. The marginal return to talent, however, greatly varies between the two. Hence, those workers with the lowest disutility of effort, or the greatest productivity of effort, will be more valuable working for organizations with greater sales. Essentially, managers are distributing their efforts across a greater scope of inputs, just as superstar performers spread their efforts to ever larger groups of customers. Once again, marginally better managers will earn significantly more than slightly less able managers because they work with a much larger scale of complementary inputs.

Rosen's (1997a) presidential address to the Society of Labor Economists shows that there is an endogenous reason for income inequality among *ex ante* identical workers. His model relies on non-convexity of preferences. These can arise from a variety of primitives: Friedman (1953) provides one; Bergstrom (1986) utilizes state dependent utility functions; Becker, Murphy and Werning (2005) use status; Becker, Murphy and Tamura (1990) and Tamura (1994) produce one with non-convexities in human capital. With these assumptions, Rosen demonstrates that the equilibrium and efficient outcome includes occupation lotteries or specialized investment in order to convexify the non-convex portion of utility. The winners get to enjoy higher utility, and the losers enter a lower level of utility. *Ex ante*, individuals are better off for entering into the lottery.

Investment Theory

Another area that receives considerable attention from Rosen is investment theory. Rosen applies his customary analytical insights to understanding the dynamics of investment, particularly in areas where the 'time to build' aspect is significant (see, for example, Kydland and Prescott, 1982). This occurs in Rosen (1983), where the costs of acquiring human capital are separable from the rate of intensity of use. Furthermore, Rosen (1987) focuses on the role that investment in anticipation of future demand plays in price and quantity dynamics. These issues are explored in more detail in Rosen and Topel (1988), who produce a rational explanation for boom–bust cycles, observed in the hog market and elsewhere. With a rising supply price, rational individuals build in anticipation of demand. When investment is a small fraction of the stock of the durable good, anticipated future demand shocks produce contemporaneous price changes. If an anticipated large permanent increase in demand will occur five years into the future, then investment will occur today. The immediate rise in investment and the slow shifting out of supply leads to a reduction in current rentals to the service flow. Investment continues rising because the value of the durable good continues to rise as the number of periods before the permanent demand shift shrinks. Until the demand shock appears, rental rates continue to fall; this is the bust phase of the cycle. When the demand shift arrives, rental rates jump, but less than they would have with no anticipatory investment; this is the boom phase of the cycle. Rosen and Topel produce a bust-boom cycle that is created because of the anticipated nature of the demand shock. Unanticipated shocks would produce even more dramatic changes in the rental rate of the durable, but no boom–bust characteristics. These insights are evident in Rosen (1992) and, from two dissertations he supervised, Siow (1984) and Zarkin (1985).

One might ask when known future demand shocks would arise. Two examples are the baby boom and Disney World. The baby boom, starting in 1946 and continuing through 1964, produced above-trend rates of fertility in American women. It is known that by age six a child must be enrolled in school. Hence, with generally little uncertainty, college students in the mid-1940s would have foreseen an increase in the demand for primary school teachers starting in 1951, for secondary school teachers in 1959, and for college faculty in the 1960s. In the second example, Walt Disney announced the construction of Disney World in Orlando, Florida, in the mid-1960s. It opened to the public only in 1970; but the model predicts increased construction of hotels, housing, schools, shopping areas, and so on in anticipation of the future increase in population. Examples of unexpected shocks would be the space race induced by Sputnik, the Soviet satellite, in the late 1950s, the space-science bust of the late 1960s, and the unexpected end of the baby boom.

Murphy, Rosen and Scheinkman (1994) apply the dynamic model of investment to the cattle industry. The long gestation cycle of cows (eight months) and their relatively short reproductive life (eight to ten years) implies that the breeding stock is likely to be a very large portion of the overall cattle herd. Thus, demand shocks are likely to greatly affect the breeding stock and hence the industry's ability to respond to future demand shocks. The authors show that their model does an excellent job of fitting the data from 1875 to 1990, despite the change in technology arising from corn feeding as opposed to range feeding, introduced in the 1930s and 1940s, which halved the time of the beef production cycle.

Rosen (1999) re-examines the Irish potato famine. He disproves the idea that potatoes were a Giffen good. As in the cattle industry, seed potatoes are a large proportion of the crop. Rosen argues that rational expectations of Irish potato farmers, who assumed that the potato blight was a temporary and not a permanent productivity shock, sealed their doom, since they did not consume their seed stock. When the blight turned out to be permanent, their exposure to imminent starvation was *ex post* predictable and tragic.

Conclusion

A measure of Rosen's influence on the economics profession can be seen by the number of published academic tributes to him (see, for example, Hartog, 2002; Lazear, 2003; Sanderson, 2001). In addition to his fertile research, Rosen possessed a great talent for synthesis, not only in his own work, as testified by Lazear (2003), but in entire fields. This is evident by his seminal contributions in this regard for human capital theory in 'Human capital: a survey of empirical research' (1977), 'Implicit contracts: a survey' (1985), 'The theory of equalizing differences' (1986b), 'Public employment taxes and the welfare state in Sweden' (1997b) and 'Theories of the distribution of earnings' (Neal and Rosen, 2000). Rosen was influential in much of Lazear's work (1995) on personnel economics.

Sherwin Rosen married Sharon Girsburg from Chicago. They were the embodiment of the marriage covenant, a beacon to all who knew them. They shared their love for 40 years, and had two daughters, Jennifer and Adria. Sherwin Rosen was a beloved professor at the University of Rochester and the University of Chicago. He was treasured by his colleagues and affectionately admired by graduate students. His concern for the success of his junior colleagues and of graduate students was legendary. His keen insight lit the seminars and classes, and his infectious laughter filled the hearts of his colleagues and graduate students. His work continues to illuminate the way for the economics profession, and his memory inspires and warms his former colleagues and students.

See Also

► Compensating differentials
► Hedonic prices
► Human capital
► Personnel economics
► Roy model
► Superstars, economics of
► Value of life

Selected Works

1972a. Learning and experience in the labour market. *Journal of Human Resources* 7: 326–342.

1972b. Learning by experience as joint production. *Quarterly Journal of Economics* 86: 366–382.

1973. (With M. Nadiri.) *A Disequilibrium Model of Demand for Factors of Production*. New York: NBER, Columbia University Press.

1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82: 34–55.

1976. A theory of life earnings. *Journal of Political Economy* 84: S45–S68.

1976. (With R. Thaler.) The value of saving a life: Evidence from the labor market. In *Household production and consumption: NBER studies in income and wealth*, vol. 40, ed. N. Terleckyj. New York: Columbia University Press.

1977. Human capital: A survey of empirical research. In *Research in Labor Economics*, vol. 1, ed. R. Ehrenberg. Amsterdam: North-Holland.

1978. Substitution and the division of labor. *Economica* 45: 235–250.

1978. (With M. Mussa.) Monopoly and product quality. *Journal of Economic Theory* 18: 301–317.

1979. (With R. Willis.) Education and self selection. *Journal of Political Economy* 87: S7–S36.

1981. The economics of superstars. *American Economic Review* 71: 845–858.

1981. ed. *Studies in Labor Markets*. Chicago: NBER, University of Chicago Press.

1981. (With E. Lazear.) Rank-order tournaments as optimal labour contracts. *Journal of Political Economy* 89: 841–864.

1982a. Authority, control and the distribution of earnings. *Bell Journal of Economics* 13: 311–323.

1982b. Further notes on the division of labour and the extent of the market. Working paper. Chicago: University of Chicago.

1983. Specialization and human capital. *Journal of Labor Economics* 1: 43–49.

R

1985. Implicit contracts: A survey. *Journal of Economic Literature* 23: 1144–1175.

1986a. Prizes and incentives in elimination tournaments. *American Economic Review* 76: 701–715.

1986b. The theory of equalizing differences. In *Handbook of labor economics*, vol. 1, ed. O. Ashenfelter and R. Layard. Amsterdam: North-Holland.

1987. Dynamic animal economics. *American Journal of Agricultural Economics* 69: 547–557.

1988. The value of changes in life expectancy. *Journal of Risk and Uncertainty* 1: 285–304.

1988. (With R. Topel.) Housing investment in the United States. *Journal of Political Economy* 96: 718–740.

1988. (With C. Winship, eds.) Organizations and institutions: Sociological and economic approaches to the analysis of social structure. Special issue of *American Journal of Sociology* 94.

1992. The market for lawyers. *Journal of Law and Economics* 35: 215–246.

1994. ed. *Implicit contract theory*. London: Edward Elgar.

1994. (With K. Murphy and J. Scheinkman.) Cattle cycles. *Journal of Political Economy* 102: 468–492.

1997a. Manufactured inequality. *Journal of Labor Economics* 15: 189–196.

1997b. Public employment taxes and the welfare state in Sweden. In *The Welfare State in Transition*, ed. R. Freeman, B. Swedenborg and R. Topel. Chicago: NBER, University of Chicago Press.

1998. (With H. Li.) Unraveling in matching markets. *American Economic Review* 88: 371–387.

1999. Potato paradoxes. *Journal of Political Economy* 107: S294–S313.

2000. (With D. Neal.) Theories of the distribution of earnings. In *Handbook of Income Distribution*, vol. 1, ed. A. Atkinson, and F. Bourgignon. Amsterdam: North-Holland.

2001. (With H. Li and W. Suen.) Conflicts and common interests in committees. *American Economic Review* 91: 1478–1497.

2002. Markets and diversity. *American Economic Review* 92: 1–15.

2005. *Markets and Diversity*. Cambridge, MA: Harvard University Press.

Bibliography

Acemoglu, D. 1998. Why do new technologies complement skills? Directed technical change and wage inequality. *Quarterly Journal of Economics* 113: 1055–1089.

Acemoglu, D. 2002. Directed technical change. *Review of Economic Studies* 69: 781–810.

Ashenfelter, O., and A. Krueger. 1994. Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84: 1157–1173.

Ashenfelter, O., and C. Rouse. 1998. Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113: 253–284.

Ashenfelter, O., and D. Zimmerman. 1997. Estimates of the returns to schooling from sibling data: fathers, sons, brothers. *Review of Economic Statistics* 79: 1–9.

Baumgardner, J. 1988a. The division of labor, local markets, and worker organizations. *Journal of Political Economy* 96: 509–527.

Baumgardner, J. 1988b. Physicians' services and the division of labor across local markets. *Journal of Political Economy* 96: 948–982.

Becker, G., K. Murphy, and R. Tamura. 1990. Human capital, fertility and economic growth. *Journal of Political Economy* 98: S12–S37.

Becker, G., K. Murphy, and I. Werning. 2005. The equilibrium distribution of income and the market for status. *Journal of Political Economy* 113: 282–310.

Benabou, R., and J. Tirole. 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics* 117: 871–915.

Benabou, R., and J. Tirole. 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies* 70: 489–520.

Benabou, R., and J. Tirole. 2004. Willpower and personal rules. *Journal of Political Economy* 112: 848–887.

Ben-Porath, D. 1967. The production of human capital and the life cycle in earnings. *Journal of Political Economy* 75: 352–365.

Bergstrom, T. 1986. Soldiers of fortune. In *Essays in Honor of Kenneth J. Arrow*, ed. W. Heller, R. Starr, and D. Starrett, vol. 2. New York: Cambridge University Press.

Ekeland, I., J. Heckman, and L. Nesheim. 2004. Identification and estimation of hedonic models. *Journal of Political Economy* 112: S60–109.
Friedman, M. 1953. Choice, chance and the personal distribution of income. *Journal of Political Economy* 61: 277–290.
Hartog, J. 2002. Desperately seeking structure: Sherwin Rosen (1938–2001). *Economic Journal* 112: 519–531.
Jones, L., and R. Manuelli. 1990. A convex model of equilibrium growth: Theory and policy implications. *Journal of Political Economy* 98: 1008–1038.
Kydland, F., and E. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50: 1345–1370.
Lazear, E. 1995. *Personnel Economics*. Cambridge: MIT Press.
Lazear, E. 2003. Sherwin Rosen. In *Biographical memoirs of the National Academy of Sciences*, vol. 83. Washington, DC: National Academies Press. Online. Available at http://www.nap.edu/readingroom/books/biomems/srosen.pdf. Accessed 10 Aug 2005.
Lucas, R. Jr. 1988. On the mechanics of economic development. *Journal of Monetary Economics* 22: 3–42.
Murphy, K., and R. Topel. 1987. Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. In *Unemployment and the structure of labor markets*, ed. K. Lang and J. Leonard. New York: Basil Blackwell.
Murphy, K., and R. Topel, eds. 2003. *Measuring the gains from medical research: An economic approach*. Chicago: University of Chicago Press.
Rebelo, S. 1991. Long-run policy analysis and long-run growth. *Journal of Political Economy* 99: 500–521.
Roback, J. 1982. Wages, rent and the quality of life. *Journal of Political Economy* 90: 1257–1278.
Romer, P. 1986. Increasing returns and long-run growth. *Journal of Political Economy* 94: 1002–1037.
Roy, A. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146.
Sanderson, A. 2001. Sherwin Rosen, 1938–2001. *Journal of Sports Economics* 2: 211–212.
Siow, A. 1984. Occupational choice under uncertainty. *Econometrica* 52: 631–645.
Tamura, R. 1992. Efficient equilibrium convergence: Heterogeneity and growth. *Journal of Economic Theory* 58: 355–376.
Tamura, R. 1994. Fertility, human capital and the 'wealth of families'. *Economic Theory* 4: 593–603.
Tamura, R. 1996. Regional economies and market integration. *Journal of Economic Dynamics and Control* 20: 825–845.
Tamura, R. 2002. Human capital and the switch from agriculture to industry. *Journal of Economic Dynamics and Control* 27: 207–242.
Tamura, R. 2006. Human capital and economic development. *Journal of Development Economics* 79: 26–72.
Zarkin, G. 1985. Occupational choice: An application to the market for public school teachers. *Quarterly Journal of Economics* 100: 409–446.

Rosenstein-Rodan, Paul Narcyz (1902–1985)

Richard S. Eckaus

Keywords

Development economics; Disguised unemployment; Rosenstein-Rodan, P. N.

JEL Classifications

B31

Rodan was one of the founders and first leaders of the field of development economics. His formative intellectual years were in the Austrian School of economics at the University of Vienna. He moved to the Department of Political Economy at University College London, in 1931.

Rodan's early essays in economics show a preoccupation with themes which reappeared throughout his professional career: the interaction and complementarity of economic processes (1933) and their temporal patterns (1934). Rodan's seminal article on developing countries (1943) argued that complementarities and externalities in demand and production created a need for the programming of investment. The arguments were subsequently extended to justify the need for an across-the-board 'big push' for a successful start to the development process (1963). He was among the first to apply the concept of 'disguised unemployment', described by Joan Robinson (1936), to developing countries as a persisting rather than cyclical problem.

R

Rodan first became actively engaged in development policy during his tenure at the World Bank from 1947 to 1954. In 1954 he moved to the Department of Economics at the Massachusetts Institute of Technology, where he produced an influential article (1961) which demonstrated that feasible levels of assistance to developing countries would substantially improve their growth performance. After retirement from MIT in 1968 he moved to the University of Texas and then to Boston University in 1972, where he established and worked in the Center for Latin American Development Studies until his death. Rodan was an active policy adviser to international agencies and governments of many countries and served on the Panel of Experts, the 'Nine Wise Men' of the Alliance for Progress, from 1961 to 1966.

Selected Works

1933. La complementarita, prime delle tre fase del progresso della teoria economica pura. *Riforma Sociale* 44: 257–308.
1934. The role of time in economic theory. *Economica* NS 1: 77–97.
1943. Problems of industrialization of eastern and south-eastern Europe. *Economic Journal* 53: 202–211.
1957. Disguised unemployment and underemployment in agriculture. *Monthly Bulletin of Agricultural Economics and Statistics* 6(7–8): 1–6.
1961. International aid for underdeveloped countries. *Review of Economics and Statistics* 43: 107–138.
1963. Notes on the theory of the 'big push' in economic development. In *Proceedings of a conference of the international economics association*, ed. H.S. Ellis. London: Macmillan.

Bibliography

Robinson, J. 1936. Disguised unemployment. *Economic Journal* 46: 225–237.

Rosenthal, Robert W. (1944–2002)

Barton L. Lipman

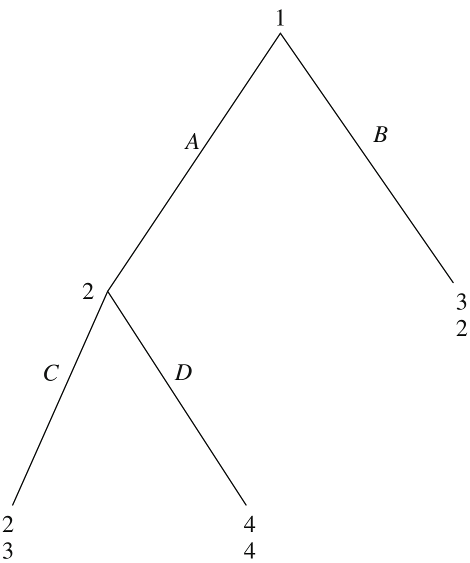
Keywords
Backward induction; Centipede game; Iterated dominance; Mixed strategy equilibria; Multi-unit auctions; Quantal response equilibrium; Rationality; Repeated games; Revelation principle; Rosenthal, R. W.; Social norms; Sovereign debt

JEL Classification
B31

Robert W. Rosenthal (1944–2002) was an economic theorist whose thoughtful papers inspired a wide range of new ideas. As Radner and Ray (2003) point out, Rosenthal (1978) gives one of the first formal statements of the revelation principle, a result noted in Myerson's first paper (1979) on the subject. Rosenthal (1979) initiated the study of repeated games with varying opponents, a modelling device used by Milgrom et al. (1990), Kandori (1992), and others to study social norms and other issues. He also wrote influential papers on pricing (Rosenthal 1980, 1982), multi-unit auctions (Krishna and Rosenthal 1996), purification of mixed strategy equilibria (Radner and Rosenthal 1982; Aumann et al. 1983), sovereign debt (Fernandez and Rosenthal 1990), analysis of experimental data (Brown and Rosenthal 1990), and many other topics.

He is arguably best-known for his 1981 *Journal of Economic Theory* paper in which he discussed what Binmore (1987) named the 'centipede game'. Like its older cousin, the Prisoner's Dilemma, the centipede game beautifully summarizes a fundamental and intriguing strategic problem. Like the game which inspired but was overshadowed by it, Selten's (1978) chain store paradox, the centipede calls into question one of the most basic principles of game theory, namely, backward induction.

Consider the game shown in Fig. 1. In this game, backward induction predicts that 1 plays *A* and 2 plays *D*. The reasoning seems very compelling. If 2 is rational, then, faced with a choice between a payoff of 3 and a payoff of 4, he obviously chooses 4. Hence 2 will play *D*. If 1 knows that 2 is rational, 1 knows that 2 will play *D*. Hence, if 1 is rational, he will choose *A* to get 4 instead of playing *B* which would yield 3. Thus the hypothesis that each player is rational and knows the other is rational seems to predict the backward induction solution. In longer games, there will be longer chains of reasoning, of course. However, the reasoning above has led many to conclude that backward induction is the implication of rationality and common knowledge of rationality.



Rosenthal, Robert W. (1944–2002), Fig. 1

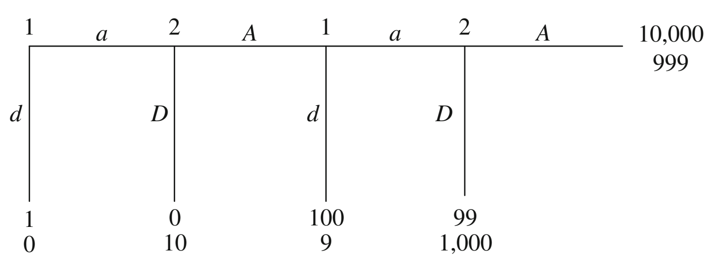
A version of Rosenthal's centipede is shown in Fig. 2. Here backward induction predicts that 1 chooses *d* at his first choice node, ending the game right away. Now the reasoning seems more suspect. If 2 is rational, he should choose *D* at the end rather than *A*. If 1 anticipates this, he should choose *d* at his last decision node. Similar reasoning shows that 2 should choose *D* at his first decision node and that 1 should choose *d* at his first node. Yet it is clear that each player must be virtually certain about his opponent's choice at the next move to justify choosing down rather than across, a certainty that seems extremely implausible in practice.

Many writers have argued that Rosenthal's centipede shows the paradoxical nature of backward induction. Consider player 1 at his second decision node. Here he is supposed to be certain that player 2 will choose *D* at the following node, justifying his own choice of *d*. Yet he also knows that 2 should have chosen *D* at the previous node and did not. If 2 failed to be rational in the past, why should 1 remain confident that he will be rational in the future? If 1 does have doubts, perhaps he should play *a* – a move which would make 2 glad to have deviated from 'rationality' at the preceding move!

Rosenthal's work led to a major debate on the question of backward induction. See, for example, Aumann (1995, 1998), Binmore (1987, 1996) and Reny (1993). See also Glazer and Rosenthal (1992) for a conceptually related critique of the use of iterated dominance in implementation theory.

Interestingly, McKelvey and Palfrey's (1992) experiments with the centipede game led them to develop the notion of quantal response equilibrium (McKelvey and Palfrey 1995), an idea which

Rosenthal, Robert W. (1944–2002), Fig. 2



echoes Rosenthal's own analysis. Rosenthal suggested that players might make 'mistakes' where these mistake probabilities would be decreasing in the cost of the mistakes, an idea he explored further in Rosenthal (1989). Quantal response equilibrium is another formulation of this idea.

See Also

- ▶ Behavioural Economics and Game Theory
- ▶ Rationality, Bounded
- ▶ Epistemic Game Theory: An Overview
- ▶ Quantal Response Equilibria

Selected Works

1978. Arbitration of two-party disputes under uncertainty. *Review of Economic Studies* 45, 595–604.

1979. Sequences of games with varying opponents. *Econometrica* 47, 1353–1366.

1980. A model in which an increase in the number of sellers leads to a higher price. *Econometrica* 48, 1575–1579.

1981. Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory* 25, 92–100.

1982. A dynamic model of duopoly with customer loyalties. *Journal of Economic Theory* 27, 69–76.

1982. (With R. Radner.) Private information and pure-strategy equilibria. *Mathematics of Operations Research* 7, 401–409.

1983. (With R. Aumann, Y. Katznelson and R. Radner.) Approximate purification of mixed strategies. *Mathematics of Operations Research* 8, 327–341.

1989. A bounded-rationality approach to the study of noncooperative games. *International Journal of Game Theory* 18, 273–292.

1990. (With J. Brown.) Testing the minimax hypothesis: a re-examination of O'Neill's game experiment. *Econometrica* 58, 1065–1081.

1990. (With R. Fernandez.) Strategic models of sovereign-debt renegotiations. *Review of Economic Studies* 57, 331–349.

1992. (With J. Glazer.) A note on Abreu–Matsushima mechanisms. *Econometrica* 60, 1435–1438.

1996. (With V. Krishna.) Simultaneous auctions with synergies. *Games and Economic Behavior* 17, 1–31.

Bibliography

Aumann, R. 1995. Backward induction and common knowledge of rationality. *Games and Economic Behavior* 8: 6–19.

Aumann, R. 1998. On the centipede game. *Games and Economic Behavior* 23: 97–105.

Binmore, K. 1987. Modeling rationality players, I. *Economics and Philosophy* 3: 9–55.

Binmore, K. 1996. A note on backward induction. *Games and Economic Behavior* 17: 135–137.

Kandori, M. 1992. Social norms and community enforcement. *Review of Economic Studies* 59: 81–92.

McKelvey, R., and T. Palfrey. 1992. An experimental study of the centipede game. *Econometrica* 60: 803–836.

McKelvey, R., and T. Palfrey. 1995. Quantal response equilibria in normal form games. *Games and Economic Behavior* 10: 6–38.

Milgrom, P., D. North, and B. Weingast. 1990. The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics and Politics* 2: 1–23.

Myerson, R. 1979. Incentive compatibility and the bargaining problem. *Econometrica* 47: 61–73.

Radner, R., and D. Ray. 2003. Robert W Rosenthal. *Journal of Economic Theory* 112: 365–368.

Reny, P. 1993. Common belief and the theory of games with perfect information. *Journal of Economic Theory* 59: 257–274.

Selten, R. 1978. The chain-store paradox. *Theory and Decision* 9: 127–159.

Rossi, Pellegrino Luigi Edoardo (1787–1848)

R. F. Hébert

Italian economist, jurist and statesman; born at Carrara in 1787, died at Rome in 1848. Rossi was a multi-national and a multi-talent. Expelled from his homeland for his zealous support of Italian unification, he emigrated to Switzerland, where he taught Roman history, championed constitutional reform, and became a naturalized citizen. After a major setback in his reform efforts he moved to France, and began lecturing on economics in 1827. In 1833 he succeeded J.B. Say in the chair of political economy at the Collège de France, winning the appointment over strong competition from Say's son-in-law, Charles Comte. The following year, Rossi became a naturalized French citizen. New honours followed quickly. In 1836 he was elected to the Académie des Sciences Morales et Politiques; he was elevated to the peerage in 1839; and in 1845, was named French Ambassador to Rome. The Revolution of 1848 cut him off from France, whereupon he became the semi-official adviser of Pope Pius IX, until an assassin's dagger took his life in the same year.

As an economist, Rossi was known for the effectiveness of his instruction and for his clearness of exposition. He made no great scientific discoveries, nor did he establish any doctrinal following. On the contrary, there is some evidence that politics diluted his economics. For example, he defended the artificial monopolies of the Paris stockbrokers, attorneys and central bankers, and he acquiesced in the sugar bounties. His *Cours d'économie politique,* mainly a pastiche of Ricardo and Say, nevertheless attained enough popularity to justify five editions over a span of almost half a century.

Schumpeter (1954, p. 382) contends that an appraisal of Rossi's performance in economics should not imply a like assessment of his person, an obvious concession to Rossi's catholic and peripatetic habits. Yet there is some substance to Schumpeter's (1954, p. 510) additional claim that the 'failures in his many political activities reveal more ability than do the successes of other people'.

Selected Works

1840. *Cours d'économie politique*, 5th ed. Paris: Joubert, 1884.
1857. *Mélanges d'économie politique, d'histoire et de philosophie,* 2 vols. Paris: Guillaumin.

Bibliography

Ledermann, L. 1929. *Pellegrino Rossi, l'homme et l'économiste, 1787–1848*. Paris: Librairie du Recueil Sirey.
Mignet, F.A.A. 1849. Notice historique sur la vie et les travaux de M. Rossi. *Institut National de France.* Paris: Firmin Didot.
Schumpeter, J.A. 1954. In *History of economic analysis*, ed. E.B. Schumpeter. New York: Oxford University Press.
Sforza, G. 1922. Fonti per la biografia di Pellegrino Rossi. *Risorgimento Italiano* 15: 1–24.

Rostas, Laslo (1909–1954)

R. F. Kahn

Rostas was born in Hungary in October 1909 and died in Cambridge, after prolonged illness, in October 1954 at the age of 45. He was educated at a grammar school, and at the University of Budapest. He was brought to England by Nicholas Kaldor in 1939. He collaborated with John and Ursula Hicks in the preparation of a book on *The Taxation of War Wealth*. He then collaborated with G. Findlay Shirras in the preparation of a book on *The Burden of British Taxation*.

His great pioneer work was in the comparison between different countries of productivity – especially the United Kingdom, United States and Germany. His results were published in a book and the articles set out in the Bibliography below.

He had become one of the country's leading authorities on industrial productivity. His important studies in this field led to his appointment at the Board of Trade, where Stafford Cripps had begun a campaign to promote higher productivity in industry. Rostas was recognized as a profound expert and his straightforward common sense made him a valuable member of a number of Committees.

In 1951 he was brought to Cambridge as a Research Officer in the Faculty of Economics. In the three years which remained before his death, he not only continued his work but exercised a marked influence on his colleagues. In his last illness his courageous refusal to let go of his intellectual interests was inspiring up to the last.

As his colleague at the Board of Trade, S.A.H. Dakin, wrote in the *Times* obituary (4 October 1954):

> The farewells to Rostas cannot be said without a word of appreciative recognition from someone who knew him in the Civil Service. My recollections are of an intense enthusiasm and belief in the importance to the country of these studies in which he was an acknowledged expert, of a mind always fertile and penetrating in analysis.

Selected Works

1940. Capital levies in central Europe. *Review of Economic Studies* 8: 20–32.
1943. Industrial production, productivity and distribution in Britain, Germany and the United States. *Economic Journal* 53: 39–54. This paper was the subject of the Presidential Address of the Royal Statistical Society in March 1944 and of the discussion which followed.
1945. Productivity of labour in the cotton industry. *Economic Journal* 55: 192–205.
1955. *The cost structure of selected British industries*. Cambridge: Cambridge University Press.

Rostow, Walt Whitman (1916–2003)

Douglas Calvin Dacy

Keywords
Development economics; Harrod–Domar growth model; International Economic Association; Rostow, W.

JEL Classifications
B31

Walt Whitman Rostow, economic historian, historian of economic thought, pioneer of modern development economics, and social scientist with interests in demography, politics, sociology and cultural aspects of development, was born in 1916. A professor of economics and history at the University of Oxford in 1946–7, Cambridge University, 1949–50, Massachusetts Institute of Technology 1950–61 and University of Texas at Austin 1968–2003, he is best known for his *Stages of Economic Growth: A Noncommunist Manifesto* (1960a) and for his service as National Security Advisor to US President Lyndon B. Johnson during the Vietnam War. He led an active intellectual life engaged in public policy issues up to his death in 2003.

Several themes developed in his first publication, ‘Investment and the Great Depression’ (1938), recur in his first book, *Essays on the British Economy in the Nineteenth Century* (1948), and his *Process of Economic Development* (1953). His book co-authored with A.D. Gayer and Anna J. Schwartz, *The Growth and Fluctuation of the British Economy* (1952), was considered a classic study, and his work co-authored with Max Millikan, *A Proposal: Key to an Effective Foreign Policy* (1957), made his reputation in the field of foreign policy. These books established Rostow as one of the world’s foremost economic historians of his age.

His *Stages of Economic Growth* was a blockbuster. It stepped on many toes, assuring his reputation as the one of the most controversial economists of the last half of the 20th century. At the time, his model clashed with that of Harrod–Domar. They modelled steady-state (equilibrium) growth, with no historical context, and focused on two variables: saving and output–capital ratios. Naturally, an economic historian would ask how an economy got there in the first place. Rostow though he saw a pattern in how countries got there. Development proceeded through five stages: traditional society, preconditions for take-off, take-off to sustained growth, drive to maturity and age of high mass consumption. His critics saw these stages as ‘empty boxes’, not empirically verifiable and devoid of predictive power. Most were especially critical of the take-off stage. He had not demonstrated empirically the necessity of a significant rise in the saving and output–capital ratios. His critics were not convinced about his dating of stages for the seven countries he studied. Besides, he had not heeded Marshall's dictum, *Natura non facit saltum* – nature does nothing in jumps. Controversy swirled over his discontinuous, disequilibrium approach to economic growth.

His work was so upsetting to many of the world's most distinguished researchers in the field of development that the International Economic Association convened a conference in Konstanz in 1960 devoted exclusively to Rostow's work. This exclusivity was a first for the Association, and is indicative of the importance placed on his work. If Rostow did not convince his critics, or they him, the conference gave him worldwide notoriety, and his ideas were embraced by many economists in developing countries. Twenty years later the controversy continued.

Seizing on earlier criticism, Rostow published a massive volume, *The World Economy: History and Prospect* (1978). It examines world economic history from 1790 to 1976 in terms of population dynamics, long-term trends, cyclical fluctuations in production, prices and international trade. It extends the work of *Stages* with later data, and expands coverage to 20 countries.

In 1982 Charles P. Kindleberger and Guido Di Tella edited a three-volume Festschrift in Rostow's honour. A reviewer, Mancur Olson (1985), noted a paradox: many of the contributors were critics, and in a Festschrift! He pondered over an interesting question: how can so many distinguished critics also be admirers? Henry Rosovsky (1965) probably had the right explanation in an earlier comment: 'I invariably learn more by disagreeing with Professor Rostow than I do by agreeing with most other writers.'

Among economists with roots in the 1960s, Rostow's visible positions in the US government made him the most influential. He helped to form the Alliance for Progress and was President John F. Kennedy's representative on it. As an architect of the Vietnam War and President Johnson's National Security Advisor, he became controversial in the political arena. He knew many of the world's leaders and was known by most of them. Through his public service, he became the Keynes of his day.

He continued to write books on important issues such as East–West relations, verification of nuclear arsenals, foreign aid and world population problems. He died in 2003 at age 87 just before his final book, *Concept and Controversy* (2003), was published. For many years, Rostow's ideas energized the field of economic development. That, even alone, is a major contribution to economics.

See Also

▶ Growth and Cycles
▶ Kondratieff Cycles

Selected Works

1938. Investment and the Great Depression. *Economic history review* 8, 136–58.

1948. *Essays on the British economy of the nineteenth century*. London: Oxford University Press.

1952. (With A. Gayer and A. Schwartz.) *The growth and fluctuation of the British economy*. Oxford: Oxford University Press.

1953. *The process of economic development*. New York: W.W. Norton.

1957. (With M. Millikan.) *A proposal: Key to an effective foreign policy*. New York: Harper.

1960a. *The stages of economic growth: A noncommunist Manifesto*. Cambridge: Cambridge University Press.

1960b. *The United States in the world arena*. New York: Harper.

1963. *The economics of take-off into sustained growth*. London: Macmillan.

1975. *How it all began: Origins of the modern economy*. New York: McGraw-Hill.

1978. *The world economy: History and prospect*. Austin: University of Texas Press.

1990. *Theorists of economic growth from David Hume to the present*. New York: Oxford University Press.

R

1998. *The great population spike and after*. New York: Oxford University Press.

2003. *Concept and controversy: Sixty years of taking ideas to market*. Austin: University of Texas Press.

Bibliography

Kindleberger, C., and G. di Tella. 1982. *Economics in the long view: Essays in honour of W.W. Rostow*. London: Macmillan.

Olson, M. 1985. Review of C. Kindleberger and G. di Tella, eds., *Economics in the long view: Essays in honor of W.W. Rostow. Journal of Economic Literature* 23, 622–625.

Rosovsky, H. 1965. The take-off into sustained controversy. *Journal of Economic History* 25: 271–275.

Rotating Saving and Credit Associations (ROSCAs)

Jean-Marie Baland

Keywords

Rotating saving and credit associations (roscas); Saving; Social sanctions

JEL Classifications

O1

Rotating saving and credit associations (roscas) are the simplest form of collective financial institution. A rosca is a group of individuals who meet at regular intervals, each of whom contributes at each meeting a pre-determined amount to a collective 'pot' which is then given to one member. The latter is then excluded from receiving the pot in future meetings, while still being obliged to contribute to the pot. The meeting process repeats itself until each member has received the pot, thereby completing a cycle. Then the rosca can start a new cycle. From this description, the main virtues of roscas are clear: they do not require storage of funds, accounting and durations of obligations are transparent, and there are no complicated interest payments or debt management. Roscas are very popular in developing countries. For instance, average membership rates in Indonesia have been estimated at 40 per cent of the population (Armendariz de Aghion and Morduch 2005), 20 per cent in Taiwan (Levenson and Besley 1996) and 40 per cent in a Kenyan slum (Anderson and Baland 2002). Although roscas do exist alongside more formal financial institutions, they are often the sole saving and credit institution in many rural areas.

Roscas vary widely in terms of the size of the contributions, the number of members and the frequency of meetings. Also, the process by which the pot is allocated can be a lottery (random roscas), or follow a fixed order imposed, for instance, by the leaders in the group (fixed roscas), or be determined by a bidding process (bidding roscas).

The literature identified four differing motives for individuals to save through roscas. First, roscas allow individuals to purchase indivisible goods earlier in expected terms than through the accumulation of individual savings (Besley et al. 1993). Roscas thus provide an implicit positive interest rate to those receiving the pot early. Second, as emphasized by Anderson and Baland (2002), roscas may be used by married women as a way to commit the household to higher saving rates than what can be done at home. Given the presence of social sanctions, husbands are then forced to comply with the saving rate imposed by the rosca. Relatedly, Gugerty (2006) argues that people facing intertemporal inconsistency join roscas to bind themselves to a particular saving pattern (see also Ardener 1964; Ambec and Treich 2007). Lastly, bidding roscas provide some insurance to their members against short-term income shocks by providing implicit short-run credit to those willing to pay the highest bid.

The most common problem of roscas has to do with enforcement. Indeed, the first members to receive the pot are de facto borrowers from the other members and, absent social sanctions, are better off not repaying their debts. Given that the size of the pot is fixed, they can always replicate (and also do better than) the best that the rosca can offer them by saving on their own. Social sanctions are then necessary to discipline members. Also, as argued in Anderson et al. (2003), the rule for allocating the pot can be chosen to partially address this issue. Selection of members is another issue faced by roscas, particularly bidding roscas, where higher bidders may be exposed to more intrinsic risks than others (see Eeckhoudt and Munshi 2005; Klonner and Rai 2006).

See Also

▶ Microcredit

Bibliography

Ambec, S., and N. Treich. 2007. Roscas as financial agreements to cope with self control problems. *Journal of Development Economics* 82: 120–137.

Anderson, S., and J.-M. Baland. 2002. The economics of ROSCAs and intrahousehold allocation. *Quarterly Journal of Economics* 117: 963–995.

Anderson, S., J.-M. Baland, and K. Moene. 2003. *Sustainability and organizational design in Roscas: Some evidence from Kenya*. Mimeo, Center for Research in the Economics of Development (CRED), University of Namur.

Ardener, S. 1964. The comparative study of rotating credit associations. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 44: 202–229.

Armendariz de Aghion, B., and J. Morduch. 2005. *The economics of microfinance*. Cambridge, MA: MIT Press.

Besley, T., S. Coate, and G. Loury. 1993. The economics of rotating savings and credit associations. *American Economic Review* 83: 792–810.

Eeckhoudt, J., and K. Munshi. 2005. *Mitigating regulatory inefficiency: The nonmarket response to financial regulation in India*. Mimeo, Brown University.

Gugerty, M. 2006. *You can't save alone: Commitment and rotating savings and credit associations in Kenya*. Working Paper No. 2006-08, Evans School of Public Affairs, University of Washington.

Klonner, S., and A. Rai. 2006. *Adverse selection in credit markets: Evidence from bidding roscas*. Mimeo, Cornell University.

Levenson, A., and T. Besley. 1996. The anatomy of an informal financial market: Rosca participation in Taiwan. *Journal of Development Economics* 51: 45–68.

Roth, Alvin (Born 1951)

Fuhito Kojima

Abstract

Roth is the major force in creating a vibrant field of matching theory and its application to market design. In doing so, he has discovered many properties of the stable matching problem (especially from the strategic viewpoint of game theory), studied real-life cases to test the relevance of the theory, conducted laboratory experiments (another field of study to which Roth made crucial contributions) and designed mechanisms in practice.

Keywords

Deferred acceptance algorithm; Game theory; kidney exchange; Labour market; Market design; Matching; NRMP; Roth; Roth–Peranson algorithm; School choice; Stability

JEL Classification

B31

Introduction

Alvin ("Al") Elliot Roth was born in 1951 in New York City, USA. He obtained his undergraduate degree at Columbia University in 1971 and a PhD at Stanford University in 1974 both in operations research. Roth wrote a dissertation in game theory under the supervision of Robert Wilson. Roth's dissertation topic was related to the von Neumann–Morgenstern stable set. Roth now recalls this study as one of the dead ends of game theory, but it trained him with the tools to study the field as well as stimulating his interest in it. Game theory was in the process of penetrating various fields, including economics, in the 1970s, and many of Wilson's students, including Paul Milgrom and Bengt Holmstrom, made important contributions in game theory and economics. The time was right for Roth to make his own significant contribution.

Upon graduation, he has taught at the University of Illinois. His position at Illinois was a joint appointment for the business school (operations research group) and economics. According to Roth's own recollection, the position was originally for operations research, but he suggested a joint appointment because he thought that economics was where the most interesting questions for game theory were to be found. The history of game theory and economics since them seems to have confirmed his foresight, although Roth was undoubtedly an important factor in making this self-fulfilling.

After Illinois, he taught at the Universities of Pittsburgh and Harvard. At the time of this writing (2013), he is the Craig and Susan McCaw Professor of Economics at Stanford University. He has received numerous prestigious awards: Roth and Sotomayor (1990), the standard reference for two-sided matching, received the Lanchester Prize awarded by Institute for Operations Research and the Management Sciences. He is a Sloan fellow, a Guggenheim fellow, a fellow of the American Academy of Arts and Sciences, and a Nobel Laureate in Economics.

Roth has made important contributions in many fields of economics, including axiomatic bargaining theory and experimental economics (see Roth (1979)) and Kagel and Roth (1995) for references). Even among his numerous achievements in various fields, his contributions in matching and market design have been especially highly regarded, as exemplified by the Nobel Prize in 2012, which was awarded to him together with Lloyd Shapley. Matching theory, initiated as an elegant but abstract mathematical theory by David Gale and Lloyd Shapley (1962) was transformed into a subject of study for economists largely by Roth's numerous contributions in the 1980s and early 1990s. Later, Roth applied matching theory to solve design problems in practice, such as labour markets, school choice, and kidney exchange. In doing so, Roth put forward the idea that an academic can usefully contribute to science and society by being an 'economist as engineer' (Roth 2003):by deploying various tools of economics – be they theory or empirics or experiments or numerical analysis or case studies – to solve an often complicated and messy problem.

Theory of Matching and Labour Markets

The Nobel Prize recognizes Roth's contribution 'for the theory of stable allocations and the practice of market design'. Roth was one of the first scholars to study economic aspects of matching theory and discovered many basic results. In the following, I will first describe the standard two-sided matching model (see Roth and Sotomayor (1990) for a more detailed survey of the theory). As we will see in this article, the model started as an abstract mathematical model, but Roth generalized and modified the theory throughout his career, as he deepened the understanding of which part of the abstract theory is applicable to real applications and to what extent. That effort paid off well, as the theory is now applied to design of real-life clearinghouses for labour matching and school choice.

In the college admission problem, there exist a set of students and a set of colleges, who are interested in being matched to one another. Students are assumed to have strict preferences over colleges and the outside option. Similarly, colleges have preferences over students and leaving a position vacant. A matching is a function that specifies which student is admitted to which college (or stays unmatched). One could also think of men and women who are seeking spouses (in fact, this model is sometimes called the 'marriage problem'.)

The central solution concept in two-sided matching is stability. This concept is composed of two requirements. First, the matching should be individually rational: that is, no student should be matched to a school that she prefers less than the outside option, and no college should be matched to a student who is less preferred than leaving the position vacant. Second, there should be no 'blocking pair': that is, a pair of student and a college who like each other better than their respective partners in the current matching. A matching is said to be stable if it is individually rational and has no blocking pair. In the context of marriage, a stable matching could be regarded as a matching such that there is no risk of divorce.

The notion of stability is arguably a reasonable one, and it is equivalent to the core, a central solution concept in cooperative game theory. Natural questions are whether a stable matching exists and, if so, whether there is a procedure to find a stable matching. These questions are answered in the affirmative by Gale and Shapley (1962), who propose the student-proposing 'deferred acceptance' algorithm (also known as the Gale–Shapley algorithm) described as follows:

Step 1: Each student applies to her most preferred college. Each college tentatively keeps its most preferred acceptable applicant and rejects all other applicants (if any).

Step $t \geq 2$: Each student who was rejected at the last step applies to her most preferred college among those that have not rejected her (the student stops applying if the next most preferred choice is the outside option). Each college considers the combined pool of both the new applicants at this step and the student tentatively kept from the previous step (if any). From this combined pool, each college tentatively keeps its most preferred acceptable student and rejects all other applicants (if any).

The algorithm terminates at a step in which no new rejection occurs, which occurs in a finite time because at least one rejection occurs at any step that does not terminate and there are only a finite number of possible rejections. At the terminal step, each tentative matching between a student and a college is finalized. Needless to say, it is possible to define a variant of this algorithm by switching the roles of colleges and students, because the model is symmetric (at least in our simple setting; generalisations for many-to-one and many-to-many matching models have also been obtained in the literature).

Once the deferred acceptance algorithm has been defined, an elementary, but innovative, proof shows that it produces a stable matching. To see this, begin by considering individual rationality. Note first that no student ever applies to a college that is unacceptable to her, so no student is matched to an unacceptable college. Colleges tentatively keep only acceptable students, so the finalised match cannot match a college with an unacceptable student either. These arguments show that the result of the algorithm is individually rational.

The less straightforward, and more interesting, is to show that the matching produced by this algorithm has no blocking pairs. To show this point, suppose that a student, say 1, prefers college A to the college (or the outside option) specified by the algorithm's outcome, say College B, while student 1 is acceptable to A (this is the only nontrivial case, because if student 1 is unacceptable to college A, then it is obvious that 1 and A cannot form a blocking pair). This means that student 1 has applied to college A and was rejected at some step of the algorithm, because the algorithm allows her to apply to B only after she gets rejected by A. At that step, 1 gets rejected only because A chose to keep a student more preferred to student 1. Then, at any of the subsequent steps, college A keeps a student who is at least as good as the one who it had at the preceding step (and thus that student is still preferred to student 1). Thus, at the end of the algorithm, the college fills its position with a student more preferred to student 1. Thus, student 1 and college A cannot be a blocking pair. So we have demonstrated that the matching produced by the deferred acceptance algorithm is stable.

Let me illustrate the theory with an example. Suppose there are students 1, 2 and 3, and colleges

A and B. The preferences of the agents are as below:

Student 1	Student 2	Student 3	College A	College B
A	A	B	3	1
B		A	2	3
			1	

This table means, for instance, Student 1 prefers A best, B second, and then the outside option; Student 2 prefers A best while B is not even acceptable, and so forth. With this input, the deferred acceptance algorithm works as follows:

1. In the first step, Students 1 and 2 apply to College A, while Student 3 applies to College B. College A keeps Student 2 and rejects Student 1, while College B keeps Student 3.
2. In the second step, Student 1 (who was rejected in the last step) applies to College B. College B, now faced with Students 1 and 3, keeps 1 and rejects 3.
3. Then, Student 3 (who was rejected in the last step) applies to College A. Now faced with Students 2 and 3, College A keeps 3 and rejects 2. Because Student 2 has been rejected by her only acceptable college, she will not make any more application, and the algorithm terminates.

Thus, in the matching produced by the deferred acceptance algorithm in this example, Student 1 and College B are matched, Ctudent 3 and College A are matched, and Student 2 is unmatched. It is easy to verify that this matching is stable (indeed, in this case this fact is obvious because both colleges A and B are matched to their respective first choice students, and hence they are not interested in forming blocking pairs).

Labour Markets for Medical Residents: Case Studies of Game Theory

While being a striking result, Gale and Shapley's work was purely mathematical. Subsequent works that followed it also focused mostly on the mathematical properties of the problem, such as algorithmic aspects of the deferred acceptance algorithm. In a series of papers in the 1980s, however, Roth recognized its economic value and changed the course of research completely. In Roth (1984), he showed that the allocation rule used in the labour market for American medical residents and hospitals since the 1950s produced a stable matching. (As Roth was one of the first to study medical matching clearinghouses, it is unclear what direct influence made him study this problem. He was, however, familiar with a similar, though different, model of the house exchange problem by Shapley and Scarf (1974). For that model, he had shown that the algorithm known as Gale's top trading cycles mechanism is strategy-proof (Roth 1982a), as well as writing another, earlier paper on that model (Roth and Postlewaite 1977). Thus it appears natural for Roth to consider strategic properties of matching algorithms.) More specifically, the National Intern Matching Program (NIMP) asked medical students and hospitals to submit their preferences to the central clearinghouse and used an algorithm which, while described differently, is equivalent to the hospital-proposing deferred acceptance algorithm. This finding suggests a far-reaching potential of game theory. The fact that a highly abstract concept of stability (core) appeared in the allocation procedure in practice suggests that the theory of the core captures an important aspect of the real world. In fact, the following anecdote appears to be a telling story. The algorithm which NIMP initially came up with was different from the deferred acceptance algorithm, and this algorithm produced unstable matchings. They quickly realised that there were problems with the algorithm and made changes to it. The final version of the algorithm, which was eventually put into use, produced a stable matching, and this is the basis of the algorithm that Roth (1984) analysed.

To inquire further into the importance of stability in practice, Roth (1991) studied the U.K. medical markets. In the 1960s, the U.K. medical labor markets were organized at a regional level, and these regional medical markets adopted central clearinghouses in a fashion similar to the NIMP. However, different regions adopted different algorithms, in which some regional algorithms produced a stable matching while others did not. Roth (1991)) found that most of the regions that had adopted stable algorithms used them successfully for an extended period of time, whereas most of the unstable algorithms were quickly abandoned or replaced by stable ones within a short period of time after experiencing problems, such as a decline in the number of participants. In other words, this work provided further confirmation of his own hypothesis (Roth 1984) that stability is important for the success of a matching clearinghouse. Kagel and Roth (2000) further confirmed this conjecture by conducting a controlled laboratory experiment.

Designing Labour Matching Clearinghouses

Roth pushed for another big change in the research direction of matching theory in the 1990s. Rather than simply studying existing labour matching clearinghouses as they were, he began to actively design them. Roth and Peranson (1999) offer a vivid description of how student movements and match organisers' efforts to address stakeholders' concerns led Roth to get involved in the re-design of the NIMP algorithm.

In order to design a real-life mechanism, one needs to take into account various aspects of the market that are often simplified or assumed away in theoretical studies. For instance, in medical matching there is a concern that students or hospitals may not report true preferences, if doing so leads to a better outcome for them, but such a strategic behaviour may result in a matching that is stable only with respect to the reported preferences and unstable with respect to the true preferences. Another complication is that a nontrivial proportion of American medical students are couples, and they usually prefer to be matched to residency programs that are close to each other. These aspects were not considered in Gale and Shapley's framework.

It was Roth who developed a pioneering theory about these questions. First, Roth (1982b) showed that the student-proposing deferred acceptance algorithm is strategy-proof for students. In other words, reporting true preferences is a weakly dominant strategy for every student (Dubins and Freedman (1981) independently showed this result). This result is quite striking, as a student can focus on figuring out what her preferences are, without using cognitive resources to figure out such things as the popularity of different hospitals, whether her favourite hospital ranks her highly, and so on. In that sense, the mechanism is 'safe' (Roth 2008a) for students.

Unfortunately, this conclusion does not extend to the hospital. That is, the student-proposing deferred acceptance mechanism is not strategy-proof for hospitals. In fact, an even more dismal conclusion was shown by Roth (1982b): There exists no mechanism that produces a stable matching with respect to stated preferences and is strategy-proof for both students and hospitals. So it is a futile quest to try to find a stable mechanism that is safe for everyone, at least if 'safe' means strategy-proof.

In a similar vein, Roth and an unpublished work by Sotomayor show that couples pose a problem. Roth (1984)) formalised the stability concept in the presence of couples and showed that stable matching does not necessarily exist when there is a couple. Therefore once again, it is futile to attempt to find a mechanism that finds a stable matching in a market with couples. Thus by the mid-1980s, Roth had discovered that certain problems in matching cannot be solved completely, no matter how clever the market design is.

However, of course, real markets need to decide an allocation, and need to do so even in the face of these complications. To cope with the demand for a mechanism in the American resident market, Roth and Peranson (1999) used an 'engineering' approach. They ran a number of alternative algorithms that modify the deferred acceptance algorithms on the data of submitted preferences in NRMP (National Resident Matching Program, renamed from NIMP and NIRMP by 1990s), and showed that they find a stable matching for each of the years for which they used data. Then they approached the strategic issues with two methods. First, assuming that the submitted data in NRMP represent the real preferences of agents, they computed the number of agents who can profitably deviate by reporting different preferences, knowing the submitted preferences by everyone else. They found that the proportion of agents who can profitably manipulate in this manner is less than one per cent. To complement this analysis, they also ran simulations based on (uniformly) randomly generated preferences. What they found is that the proportion of agents who can unilaterally manipulate stable mechanisms quickly decreases as the number of participants increases if each student find only a constant number of hospitals acceptable. In their simulation, for a market size of several thousand (the number of residents in NRMP in 1990s was more than 20,000) and the number of acceptable hospitals for each student is about 10 (similar to the average in NRMP), the proportion of agents who can profitably misreport preferences is already very small.

Based on these studies, Roth and Peranson designed an algorithm that modifies the student-proposing deferred acceptance algorithm to try to allocate couples in a stable manner. Of course, Roth's own results show that the problems of couples and incentives cannot be solved perfectly, but the point is that with a suitably designed mechanism it may be possible to keep the problems small enough so that, for practical use, the mechanism may achieve a reasonably desirable outcome most of the time. This mechanism, now called the Roth–Peranson algorithm, was adopted in NRMP in 1998 and is used there now, as well as dozens of other markets. Roth (2008b) has a list of such markets.

While studying the NRMP, Roth made a methodological point, advocating the engineering aspect of economics. Roth and Peranson (1999) use an interesting analogy:

> A rough analogy may be helpful for thinking about how the different parts of this paper hang together. Consider the design of suspension bridges. The Newtonian physics they embody is beautiful both in mathematics and in steel, and college students can be taught to derive the curves that describe the shape of the supporting cables. But no bridge could be built based only on this elegant theoretical treatment, in which the only force is gravity, and all beams are perfectly rigid. Real bridges are built of steel and rest on rock and soil and water, and so bridge design also concerns metal fatigue, soil mechanics, and the forces of waves and wind. Many design questions concerning these real-world complications cannot be answered analytically but, instead, must be explored using physical or computational models. Often these involve estimating magnitudes of phenomena missing from the simple Newtonian model, some of which are small enough to be of little consequence, while others will cause the bridge to fall down if not adequately addressed. Just as no suspension bridges could be built without an understanding of the underlying physics, neither could any be built without understanding many additional features, also physical in nature, but more varied and complex than addressed by the simple model. These additional features, and how they are related to and interact with that part of the physics captured by the simple model, are the concern of the scientific literature of engineering. Some of this is less elegant than the Newtonian model, but it is what makes bridges stand. Just as important, it allows bridges designed on the same basic Newtonian model to be built longer, stronger, and lighter over time, as the complexities and how to deal with them become better understood.

Then they continue to say that the theory of two-sided matching works as a basic guide just as Newtonian mechanics, but it is other methods such as computational explorations that can be directly used to guide design of the complex real markets. In short, Roth and Peranson declare that, in order to solve a complex problem in the real world, economists should use whatever method is useful, whether it is cooperative game theory, noncooperative game theory, case study, empirical analysis, simulation or an experiment. And these alternative approaches may be complementary. On the one hand, theory has informed what cannot be generally achieved, which motivates looking at data and simulation. On the other hand, data and simulation may point to new directions for theory to head. In fact, subsequent theoretical studies by Immorlica and Mahdian (2005) and Kojima and Pathak (2009) have offered a sense in which stable mechanisms become hard to manipulate for participants as the market size increases; similarly, subsequent studies by Kojima et al. (2009) and Ashlagi et al. (2011)

have shown a sense in which there is a probability that a stable matching can be found even in the presence of couples as the market size increases. Both of these lines of papers are highly motivated by Roth's findings that in large markets, such as the NRMP, the conclusions from his own impossibility theorems tend to show themselves very rarely.

The table below is taken from Roth (2003), which lists some matching clearinghouses in the field and reports whether they are stable or not and whether they are in use. As will be discussed below, more markets have adopted stable matching mechanisms in the past 10 years; and among them, especially notable is adoptions of stable mechanisms in school choice.

Market	Stable	Still in use (halted unraveling)
American medical markets		
NRMP	Yes	Yes (new design in '98)
Medical specialties	Yes	Yes (about 30 markets)
British Regional Medical Markets		
Edinburgh ('69)	Yes	Yes
Cardiff	Yes	Yes
Birmingham	No	No
Edinburgh ('67)	No	No
Newcastle	No	No
Sheffield	No	No
Cambridge	No	Yes
London Hospital	No	Yes
Other healthcare markets		
Dental Residencies	Yes	Yes
Osteopaths (<'94)	No	No
Osteopaths (≥'94)	Yes	Yes
Pharmacists	Yes	Yes
Other markets and matching processes		
Canadian Lawyers	Yes	Yes (except in British Columbia since 1996)
Sororities	Yes (at equilibrium)	Yes

Theory and Practice of School Choice

In the early 2000s, Roth got involved in a number of projects in school choice. School choice is a practice in which school children or their parents express preferences over schools, and the school districts take those preferences into account when deciding which schools the kids attend. School choice became popular beginning at least in the 1980s and many school districts in the USA and other countries now have some form of school choice. The problem, however, is that school seats are scarce resources, so it is not possible to assign every child to his or her preferred school. Moreover monetary transfer, which often facilitates efficient allocation, is impossible in school choice because public school seats are usually free (or fixed even when the government allows charging students). Thus a problem for the school district's policymaker is how to assign seats in different schools to different children in an efficient and fair manner, while making sure that children and their parents express their true preferences.

Abdulkadirog lu and Sönmez (2003) opened the modern market design literature on school choice. They gave an explicit interpretation of the stability concept tailored to school choice: Here, they started with the fact that often schools have priorities over students: that is, law decides which child should be placed to a school before others. Mathematically, a school priority is a strict ordering over students, just as college preference is in the standard two-sided matching model as described above. While a priority is mathematically isomorphic to preferences, the interpretation of blocking in the school choice context is different from that for the two-sided matching problem because of the differences of interpretation of college preferences and school priorities. If, for instance, Student 1 likes School A better than his allocation while Student 2, whose priority at A is lower, is matched to A, then such an outcome is unfair in the sense that Student 1 envies Student 2 and that envy is justified because the former has higher priority than the latter. In other words, stability in school choice can be interpreted as a normative fairness criterion requiring the absence of justified envy.

With this reinterpretation, much of matching theory from before can be used to guide market design in school choice. The student-proposing deferred acceptance algorithm finds a stable matching and it is strategy-proof for students. The latter result is of particular interest in school choice, because the priorities of schools are often determined by law, so there is no concern that schools may misreport priorities. In other words, the student-proposing deferred acceptance mechanism is fully strategy-proof in such an environment.

With this justification, Roth and his collaborators helped education authorities design school choice mechanisms. New York City and Boston were among the first to adopt the student-proposing deferred acceptance mechanism in the early 2000s (Abdulkadiroğlu et al. 2005a, b). Design experience in NYC prompted Abdulkadiroğlu et al. (2009) to study a prevalent feature in school choice: the fact that priorities are often given not as a linear order (as assumed by the model of Abdulkadirog lu and Sönmez (2003)) but as a weak order which allow indifferences. Their studies, as well as others, have helped spread the idea that school choice mechanisms can be designed with the help of economic analysis. Neil Dorosin, who was in charge of school choice in NYC's Department of Education when the deferred acceptance mechanism was adopted there, founded an NPO (where Roth is Chairman) that collaborates with various school districts in the USA to design school districts. As of this writing, more cities, such as New Orleans, Chicago, Denver and Washington DC, have adopted or are considering new mechanisms with input from Roth and his collaborators.

Kidney Exchange

Roth is also one of the pioneers in the economics of kidney exchange, an unlikely subject for many academic economists.

Kidney transplantation is a preferred method for curing many types of kidney disease. In 2006, for example, 10,659 kidneys were transplanted from diseased donors to patients. In addition, everyone has two kidneys and one can stay healthy with one kidney, so there are many kidney transplantations involving live donors. In 2006, for example, 6428 transplants were conducted from living donors. Despite these numbers, there is a large shortage of kidneys for transplantation. In 2006, for example, 3875 people died while on the waiting list for a kidney. One of the problems is that unless there is a tissue and blood-type match the immune system can reject the transplanted kidney. An additional problem is that trade in human organs for financial gain is banned in almost all countries. In the USA, for instance, Section 301 of the National Organ Transplant Act states that 'it shall be unlawful for any person to knowingly acquire, receive or otherwise transfer any human organ for valuable consideration for use in human transplantation'. Thus the buying and selling of organs is illegal.

In a series of papers, Roth, Sönmez and Unver (2004, 2005a, 2007) described the first game-theoretic models of kidney exchange. To get the basic idea, imagine that there are two donor–patient pairs such that (1) Pair 1 has a blood type A donor and blood type B patient, and (2) Pair 2 has a blood type B patient and a blood type A patient, so neither of the pair can perform a successful transplant on their own. Instead, the donor in Pair 1 could give her kidney to the patient in Pair 2 while the donor in Pair 2 could give his kidney to the patient in Pair 1, resulting in both patients receiving compatible kidneys (assuming, of course, there is no other incompatibility between these donors and patients).

The above is a highly stylised example, but the basic idea for the general case is similar. This is a trade, just as the trade of apple and oranges, but interestingly such a trade could be made even without monetary transfer. At least since Jevons (1876), however, trade without money can be pretty tricky because there needs to be a double coincidence of wants. Thus a problem for an academic is to analyse what kind of mechanism can facilitate the trade of kidneys (or, using a terminology that is less likely to receive hostile reaction, 'kidney paired donation'). Roth, Sönmez and Unver provided a series of possible mechanisms that achieve Pareto-efficient outcomes and are strategy-proof for alternative hypotheses. Interestingly, a desirable mechanism can be quite different depending on constraints such as how many transplants can be performed simultaneously due to hospital logistic capacities. This aspect is emblematic of Roth's thesis that academics studying market design may need to respect various constraints and provide desirable mechanisms in the face of constraints, just as he needed to resort to computational approaches when designing the NRMP mechanism in the 1990s. Roth, Sönmez and Unver (2005b) describe their experience in organizing kidney exchange mechanisms. Studies in kidney exchanges are particularly new among the fields related to Roth's contributions discussed in this article. As such, this section is necessarily preliminary as well. There are many new papers on this subject, including Roth's own contributions such as Ashlagi and Roth (2011, 2012).

Other Contributions

As mentioned at the beginning of this article, Roth has influenced many areas of economics. His first book is a monograph on axiomatic models of bargaining theory, to which he made contributions early in this career. The mathematical methods typical in such models – propose desirable properties, and look for solutions satisfying those properties – may have helped develop approaches in some of Roth's works in matching (such as Roth 1982a, b; 1984) as well as many papers in the field. However, in his autobiography prepared for the Nobel Prize Roth writes 'Axiomatic theories… were beautiful, and I enjoyed pushing the theory forward, but their failure to account for the kinds of behavior we observed so clearly in experiments convinced me that these… were a dead end for economics', and he shifted his research focus away from axiomatic bargaining later in his career.

Consistently with his view, Roth has extensively studied laboratory experiments. Important methodologies to control lab experiments were introduced by Roth and Malouf (1979), which introduced the experimental design using binary lotteries, and by Roth and Murnighan (1978), which used probabilistic termination of repeated games. The behaviour of real people, rather than that posited by axiomatic theories, was studied in many of his papers, including a series of papers on learning in games by Roth and Erev (1995) and Erev and Roth (1998). In a similar spirit of studying the behaviour of real people, Roth et al. (1991) studied subjects' behaviour in bargaining and market experiments in four different countries. In an experiment conducted in Slovakia, Slonin and Roth (1998) varied the financial stakes for subjects by a factor of 25 and tested how the magnitude of financial stakes changes people's behaviour. The fact that Roth has made major contributions in experimental economics as well as market design seems to be a natural, if not necessary, consequence of how research in market design has been conducted. As mentioned above, theory provides analysts with testable predictions, but some theories have better predictive power than others, which is the kind of fact that experiments can help understand. Moreover, theory abstracts away many aspects of the real world, as we have seen in the labour market example above, so using other methods, such as laboratory experiments, is a natural complement to theory. The paper by Kagel and Roth (2000) mentioned above is an example of the complementary nature of laboratory experiment and theory in the research in market design. More generally, Roth's approach is to utilise any method as long as it is useful for understanding the problem that needs to be solved.

See Also

▶ Matching and Market Design

Acknowledgments I am grateful to Al Roth and Fanqi Shi for helpful comments.

Bibliography

Abdulkadiroğlu, A., and T. Sönmez. 2003. School choice: A mechanism design approach. *American Economic Review* 93: 729–747.

Abdulkadiroğlu, A., P. Pathak, and A. Roth. 2005a. The New York City high school match. *American Economic Review: Papers and Proceedings* 95: 364–367.

Abdulkadiroğlu, A., P. Pathak, A. Roth, and T. Sönmez. 2005b. The Boston public school match. *American Economic Review: Papers and Proceedings* 95: 368–371.

Abdulkadirog lu, A., P. Pathak, and A. Roth. 2009. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review* 99: 1954–1978.

Ashlagi, I., and A. Roth. 2011. New challenges in multi-hospital kidney exchange. *American Economic Review: Papers and Proceedings* 102: 354–359.

Ashlagi, I., and A. Roth. 2012. *Individual rationality and participation in large scale, multi-hospital kidney exchange*. Cambridge, MA: Mimeo.

Ashlagi, I., M. Braverman, and A. Hassidim. 2011. *Stability in large matching markets with complementarities*. Mimeo.

Dubins, L.E., and D.A. Freedman. 1981. Machiavelli and the Gale–Shapley algorithm. *American Mathematical Monthly* 88: 484–494.

Erev, I., and A. Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88: 848–881.

Gale, D., and L. Shapley. 1962. College admissions and the stability of marriage. *American Mathematical Monthly* 69: 9–15.

Immorlica, N., and M. Mahdian. 2005. Marriage, honesty, and stability. In *Proceedings of the sixteenth annual ACM-SIAM Symposium on Discrete Algorithms*, 53–62. New York: Association for Computing Machinery Philadelphia.

Jevons, W. 1876. *Money and the mechanism of exchange*. New York: D. Appleton and Company.

Kagel, J., and A. Roth (eds.). 1995. *The handbook of experimental economics*. Princeton: Princeton University Press.

Kagel, J. and A. Roth. 2000. The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics* 115: 201–235.

Kojima, F., and P. Pathak. 2009. Incentives and stability in large two-sided matching markets. *American Economic Review* 99: 608–627.

Kojima, F., P. Pathak, and A. Roth. 2009. *Matching with couples*. Cambridge, MA: Mimeo.

Roth, A. 1979. *Axiomatic models of bargaining*, Lecture notes in economics and mathematical systems #170. New York: Springer.

Roth, A. 1982a. Incentive compatibility in a market with indivisible goods. *Economics Letters* 9: 12–132.

Roth, A. 1982b. The economics of matching: Stability and incentives. *Mathematics of Operations Research* 7: 617–628.

Roth, A. 1984. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy* 92: 991–1016.

Roth, A. 1991. A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *American Economic Review* 81: 415–440.

Roth, A. 2003. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70: 1341–1378.

Roth, A. 2008a. What have we learned from market design? Hahn Lecture. *Economic Journal*, 118: 285–310.

Roth, A. 2008b. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory* (Special Issue in Honor of David Gale on his 85th birthday), 36: 537–569.

Roth, A. 2013. *Intellectual autobiography.* Prepared by the Laureate to be posted at Nobel Organization's website: http://www.nobelprize.org/

Roth, A. and I. Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8: 164–212.

Roth, A. and M.K. Malouf. 1979. Game theoretic models and the role of information in bargaining. *Psychological Review* 86: 574–594.

Roth, A., and J.K. Murnighan. 1978. Equilibrium behavior and repeated play of the prisoners' dilemma. *Journal of Mathematical Psychology* 17: 189–198.

Roth, A., and E. Peranson. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review* 89: 748–780.

Roth, A., and A. Postlewaite. 1977. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics* 4: 131–137.

Roth, A. and M. Sotomayor. 1990. *Two-sided matching: A study in game-theoretic modeling and analysis*, Econometric society monographs No. 18. Cambridge, MA: Cambridge University Press.

Roth, A., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review* 81: 1068–1095.

Roth, A., T. Sönmez, and U. Ünver. 2004. Kidney exchange. *Quarterly Journal of Economics* 119: 457–488.

Roth, A., T. Sönmez, and U. Ünver. 2005a. Pairwise kidney exchange. *Journal of Economic Theory* 125: 151–188.

Roth, A., T. Sönmez, and U. Ünver. 2005b. Kidney exchange clearinghouse in New England. *American Economic Review: Papers and Proceedings* 95: 376–380.

Roth, A., T. Sönmez, and U. Ünver. 2007. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review* 97: 828–851.
Shapley, L., and H. Scarf. 1974. On cores and indivisibility. *Journal of Mathematical Economics* 1: 23–37.
Slonin, R. and A. Roth. 1998. Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica* 66: 569–596.

Major Publications

Roth, A., and A. Postlewaite. 1977. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics* 4: 131–137.
Roth, A. and J.K. Murnighan. 1978. Equilibrium behavior and repeated play of the prisoners' dilemma. *Journal of Mathematical Psychology* 17: 189–198.
Roth, A. 1979. *Axiomatic models of bargaining*, Lecture notes in economics and mathematical systems #170. New York: Springer.
Roth, A. and M.K. Malouf. 1979. Game theoretic models and the role of information in bargaining. *Psychological Review* 86: 574–594.
Roth, A. 1982. Incentive compatibility in a market with indivisible goods. *Economics Letters* 9: 12–132.
Roth, A. 1982. The economics of matching: Stability and incentives. *Mathematics of Operations Research* 7: 617–628.
Roth, A. 1984. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy* 92: 991–1016.
Roth, A. 1985. The college admission problem is not equivalent to the marriage problem. *Journal of Economic Theory* 36: 277–288.
Roth, A. and M. Sotomayor. 1990. *Two-sided matching: A study in game-theoretic modeling and analysis*, Econometric society monographs No. 18. Cambridge, MA: Cambridge University Press.
Roth, A. 1991. A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *American Economic Review* 81: 415–440.
Roth, A., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir. 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *American Economic Review* 81: 1068–1095.
Kagel, J., and A. Roth (eds.). 1995. *The handbook of experimental economics*. Princeton: Princeton University Press.
Roth, A., and Erev, I. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8: 164–212.
Erev,I. and A. Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88: 848–881.
Slonin, R., and A. Roth. 1998. Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica* 66: 569–596.
Roth, A. and E. Peranson. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review* 89: 748–780.
Kagel, J. and A. Roth. 2000. The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics* 115: 201–235.
Roth, A., and A. Ockenfels. 2002. Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet. *American Economic Review* 92: 1093–1103.
Roth, A. 2003. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70: 1341–1378.
Roth, A., T. Sönmez, and U. Ünver. 2004. Kidney exchange. *Quarterly Journal of Economics* 119: 457–488.
Abdulkadiroğlu, A., P. Pathak, and A. Roth. 2009. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review* 99: 1954–1978.

Rothbard, Murray N. (1926–1995)

Benjamin Powell and Edward Stringham

Keywords

Anarchism; Antitrust enforcement; Apriorism; Austrian economics; Cartels; Economic history; Economics, definition of; Great depression; Hayek, F. A. von; Hazlitt, H; Interpersonal utility comparisons; Libertarianism; Market failure; Mises, L. E. von; Monetary policy; Monopolies; Perfect competition; Private enforcement of law; Public goods; Rand, A; Rothbard, M. N; Welfare economics

JEL Classifications

B31

Murray Rothbard was influential in continuing the tradition of the Austrian school of economics in America. In more than two dozen books and hundreds of articles, his work spanned economics, history, philosophy and political science. He earned his Ph.D. from Columbia University, but was influenced mainly by Ludwig Von Mises' seminar at New York University. Rothbard was a strong believer in *apriorism*, the idea that economic laws could be discovered using logical reasoning (as opposed to empirical testing), and he attempted to build on and extend the economic logic of Mises and others in that tradition. Rothbard's treatise, *Man, Economy, and State* (1962), analysed the economics of market exchange, while his follow-up volume, *Power and Market* (1970), analysed the economics of government intervention. An underlying theme of his work is that the market is the realm of mutually beneficial exchange, whereas the government is the realm of coercion where some gain at the expense of others.

Rothbard considered economics to be a value-free science, but he believed economic reasoning can be used to determine whether normative views are internally consistent. He was strongly critical of government intervention in the economy, arguing against those who believe that government policies can be Pareto-superior and make all people better off. For example, Rothbard was one of the only economists writing in the 1950s and 1960s to argue against all antitrust laws. He thought that perfect competition was an unattainable ideal, and he said that monopolies or cartels do not pose problems on the free market. He believed that the only monopolies that warrant concern are those sanctioned by government. Rothbard was critical of arguments about market failure in general, insisting that mainstream notions of economic efficiency were unrealistic. He criticized the welfare economics of his day on the grounds that it rests on unscientific interpersonal comparisons of utility.

In addition, Rothbard wrote a great deal on economic history, often documenting government getting in the way of markets. For example, his 1963 book *America's Great Depression* argued that government caused and lengthened the Great Depression through distortionary monetary and regulatory policies. Rothbard also devoted much of his writing to political philosophy, and here too he was unabashedly libertarian. Rothbard's contribution is particularly noteworthy because he was one of the first economists to argue that markets do not depend on the existence of government. Before him, even the most free-market theorists, such as Ludwig von Mises, Henry Hazlitt, Ayn Rand, and Friedrich Hayek, had simply assumed that services like law enforcement must be provided collectively by the state. But in *Power and Market* and *For a New Liberty* (1973) Rothbard maintained that public goods such as law enforcement must be analysed in terms of marginal units and, as with other goods, those marginal units can be provided privately. He pointed to historical examples of private law enforcement and speculated how a purely private system might function. Rothbard's ideas advancing private property anarchism were radical, but they influenced many economists who now write about alternatives to government law (Stringham 2006). Rothbard's thorough libertarian views pushed free-market thinking to become more free market.

See Also

- Austrian Economics: Recent Work
- Hayek, Friedrich August von (1899–1992)
- Libertarianism
- Menger, Carl (1840–1921)
- Mises, Ludwig Edler von (1881–1973)

Selected Works

1962. *Man, economy, and state: A treatise on economic principles*. Princeton: Van Nostrand.

1963. *America's great depression*. Princeton: Van Nostrand.

1970. *Power and market*. Kansas City: Sheed Andrews and McMeel.

1973. *For a new liberty: The libertarian manifesto*. New York: Macmillan.

Many of Rothbard's writings are freely available online at http://www.rothbard.org. Accessed 26 Sept 2006.

Bibliography

Raimondo, J. 2000. *An enemy of the state: The life of Murray N. Rothbard*. Amherst: Prometheus Books.
Stringham, E. (ed.). 2006. *Anarchy and the law: The political economy of choice*. Somerset: Transaction Publishers.

Rothbarth, Erwin (1913–1944)

J. R. N. Stone

Rothbarth was born on 16 December 1913 in Frankfurt am Main, and died on 25 November 1944 at Venraij. He emigrated to England in 1933, where he attained his BA(Econ.) at the London School of Economics in 1936. He became an assistant in Statistical Research in the Faculty of Economics in Cambridge from 1938 to 1940. From May to August 1940 he was interned as an enemy alien. After his release he returned to Cambridge to teach economic statistics. He then volunteered for active service in the British Army in 1944, and was killed in action in Holland.

In his short career Rothbarth made several interesting contributions (Kalecki 1944–5). He updated for Keynes Colin Clark's national income estimates and may have been partly responsible for their presentation in accounting format (Keynes 1940; Cuyvers 1983). He applied index-number theory to the measurement of real income under rationing and to price grouping in demand analysis (Rothbarth 1941, 1944). He contributed two valuable appendices on family income and saving to Madge (1943). His last paper, published posthumously, compares productivity in the US and the UK (Rothbarth 1946).

See Also

▶ Rationing

Selected Works

1940. Statistical contribution to J.M. Keynes, *How to pay for the war.* London: Macmillan.
1941. The measurement of changes in real income under conditions of rationing. *Review of Economic Studies* 8: 100–107.
1943. Appendices IV and V to C. Madge, *Wartime patterns of saving and spending.* National Institute of Economic and Social Research Occasional Paper IV. Cambridge: Cambridge University Press.
1944. A note on an index number problem. *Review of Economic Studies* 11(2): 91–98.
1946. Causes of the superior efficiency of U.S.A. Industry as compared with British industry. *Economic Journal* 56: 383–390.

Bibliography

Cuyvers, L. 1983. Keynes's collaboration with Erwin Rothbarth. *Economic Journal* 93(371): 629–636.
Kalecki, M. 1944–5. The work of Erwin Rothbarth. *Review of Economic Studies* 12(2): 121–122.

Rothschild, Kurt Wilhelm (born 1914)

K. Laski

Rothschild was born in Vienna on 20 October 1914. He studied law in Vienna and, after emigration to England, economics in Glasgow. From 1940 to 1947 he was a lecturer in economics at the University of Glasgow. He returned to Vienna in 1947 and worked until 1966 in the Austrian Institute for Economic Research (AIER). From 1966 until his retirement in 1985 he was professor of economics at the University of Linz.

Rothschild has been both an empirical researcher and a gifted theoretician. As a leading member of the AIER he concentrated his research on labour market analysis and foreign trade. His field of interest in theory was wide. Best known

R

are his book (1954) on the theory of wages, his paper (1947) on price theory and oligopoly and the readings on 'Power in Economics' edited by and provided with an introduction of Rothschild (1971). Other publications have dealt with economic growth in Austria, unemployment, income distribution, disarmament, disequilibrium theory, forecasting and methodological problems of economics.

Rothschild always tried to link pure theory with relevant practical investigations. Economic institutions, social classes and their political and economic power thus played an important role, especially in his analysis of income distribution and of prices. Another important feature of his theoretical work was a pragmatic and commonsense approach. He can hardly be classified as belonging to one school in economics. Influenced by the tradition of Marshallian microeconomics he was a leading representative of Keynesian macroeconomics in Austria, shared interest for neoclassical equilibrium with that for disequilibrium theory and, last but not least, combined a rather radical left 'Weltanschauung' and political activity with a luke-warm attitude towards the Marxian labour value theory and radical economics.

Selected Works

1947. Prices theory and oligopoly. *Economic Journal* 57: 299–320.

1954. *The theory of wages.* Oxford: Basil Blackwell.

1957. Approaches to the theory of bargaining. In *The theory of wage determination,* ed. J. Dunlop. London: Macmillan.

1971. (ed.) *Power in economics,* Penguin modern economics readings. Harmondsworth: Penguin Books.

1973a. Military expenditure, exports and growth. *Kyklos* 26(4): 804–14.

1973b. Distributive aspects of the Austrian theory. In *Carl Menger and the Austrian school of economics,* ed. J.R. Hicks and W. Weber. London/New York: Oxford University Press.

1981. *Einführung in die Ungleichgewichtstheorie.* Berlin/Heidelber/New York: Springer-Verlag.

1985. Some notes on Weintraub's eclectic theory of income shares. *Journal of Post-Keynesian Economics* 7(4):575–593.

Rotten Kid Theorem

Theodore C. Bergstrom

Abstract

The rotten kid theorem states that, if a household head is sufficiently rich and benevolent towards other household members, then it is in the self-interest of other household members to take those actions that maximize the total income of the household, even at a cost to their own private income. This theorem holds under certain restrictive assumptions, but the assumptions needed for it to be true are not satisfied in many common family decision-making environments.

Keywords

Becker, G.; Conditional transferable utility; Dawkins, R.; Family economics; Kin selection; Leisure; Normal goods; Rotten kid theorem

JEL Classifications

D11

$u_i(x_i) = x_i$, and the utility function of the household head, $u(x_0, \ldots, x_n)$, is strictly increasing in all the x_i's. Every household member earns some personal income, the amount of which depends on her own actions a_i, but possibly also on the actions of other household members. Let a be the vector of actions chosen by household members, let $m_i(a)$ be i's personal income, and let $m(a) = \sum m_i(a)$. Feasible allocations must satisfy the household budget constraint, $\sum_i x_i = m(a)$. For any income y, define $(x_0(y), \ldots x_n(y))$ as the allocation that maximizes $u(x_0, \ldots x_n)$ subject to $\sum_i x_i = y$. Assume that consumption for each i is a *normal good* so that $x_i(y)$ is a strictly increasing function of y. Finally, assume that the household head has personal income large enough so that in equilibrium he chooses to donate money to all other persons in the household. This means that, for all feasible a and for each kid, i, $x_i(m(a)) > m_i(a)$. Consider the following two-stage game. In the first stage, household members choose their actions and thus determine total family income $m(a)$. In the second stage the household head finds the allocation $x(m(a))$ that maximizes $u(x_1, \ldots x_n)$ subject to $\sum_i x_i = m(a)$ and donates $x_i(m(a)) - m_i(a)$ to kid i. In the first stage of the game, each kid realizes that, after the head has redistributed income, her own consumption will be $x_i(m(a))$. The normal goods assumption implies that $x_i(m(a))$ is an increasing function of $m(a)$. Therefore, the self-interest of each kid coincides with maximizing total family income, $m(a)$. (To ensure that a maximum exists, assume that each m_i is continuous and that each a_i must be chosen from a closed bounded set.)

The trouble with the rotten kid theorem is that it fails to hold in models that make slight concessions toward realism. Bergstrom (1989) shows that, in general, the rotten kid theorem fails if kids care about their activities as well as about consumption. For example, if leisure is a complement to consumption, a child can manipulate the parents' transfer in his or her favour by taking too much leisure. Lindbeck and Weibull (1988) and Bruce and Waldman (1990) show that the rotten kid theorem fails when individuals can choose between current and future consumption. Lundberg and Pollak (2003) show a dramatic failure of the rotten kid theorem when families choose between discrete options like whether to move house or whether to have a child.

Bergstrom (1989) explored the most general conditions under which a rotten kid theorem can be proved. He showed that, in general, a necessary and sufficient condition for the conclusion of the rotten kid theorem to be satisfied is that there is 'conditional transferable utility'. This means that the utility possibility sets corresponding to all possible activity choices are nested and are bounded above by parallel straight line segments. For example, there is conditional transferable utility if kids care only about their consumption, so that $u_i(x_i, a) = x_i$, and if total family income is $m(a)$. Then the utility possibility frontier conditional on a is the simple $\{(u_1, \ldots, u_n) | \sum_1^n u_i \le m(a) \text{ and } u_i \ge 0 \text{ for all } i\}$. In general, however, if the kids' utilities depend on their actions, kids will be able to influence the 'slope' of the utility possibility frontier by their choice of actions, a. For example, a selfish kid may benefit by choosing an action that reduces family income but makes it 'cheaper' for the parent to invest in her utility rather than that of her sibling. Bergstrom shows that the most general class of environments for which there is conditional transferable utility requires that each kid i has a utility function of the form $u(x_i, a) = A(a)x_i + B_i(a)$ where x_i is i's expenditure on consumer goods and a is the vector of family members' activities. This allows the possibility that activities a_i generate externalities in consumption as well as in income-earning. (Bergstrom and Cornes 1983, show that in a public goods economy the efficient quantity of public goods is independent of income distribution if and only if preferences can be represented in this form, which is dual to the Gorman polar form for public goods.) Then, for any a, the upper boundary of the utility possibility set is $\{u | \sum u_i = A(a)m(a) + \sum B_i(a)\}$. If utilities of kids are normal goods for the head, then each kid will maximize her utility by maximizing $F(a) = A(a)m(a) + \sum B_i(a)$. Thus selfish kids would act in the family interest, as the rotten kid theorem asserts.

An interesting debate in evolutionary biology parallels the economists' rotten kid theorem. Alexander (1974) maintained that natural selection favours genetic lines in which offspring act so as to maximize family reproductive success. Dawkins (1976) disputed Alexander's argument, citing Hamilton's theory of kin selection (1964), which implies that in sexual diploid species offspring value the reproductive success of their siblings at only half of their own. Alexander

R

(1979) conceded Dawkins's point, but offered an additional reason that offspring would act in the interest of their parents, namely, that 'the parent is bigger and stronger than the offspring, hence in a better position to pose its will'. Bergstrom and Bergstrom (1999) propose an evolutionary model that could support the Becker–Alexander conclusion that children will act in the family interest. They construct a two-locus genetic model, where a gene at one locus controls an animal's behaviour when the animal is a juvenile and a gene at the other controls its behaviour when it is a parent. Then the frequency of recombination between genes at these two loci determines the evolutionary outcome of parent–offspring conflict. If recombination between these genes is rare, offspring will tend to act in the genetic interest of their parent. If recombination is frequent, there can be an equilibrium where some offspring successfully 'blackmail' their parents into giving them more resources than is optimal for the family's reproduction.

See Also

► Becker, Gary S. (Born 1930)
► Family Economics

Bibliography

Alexander, R. 1979. *Darwinism and human affairs*. Seattle: University of Washington Press.
Alexander, R. 1974. The evolution of social behavior. *Annual Review of Ecological Systems* 5: 25–83.
Becker, G. 1974. A theory of social interactions. *Journal of Political Economy* 82: 1063–1093.
Becker, G. 1981. *A treatise on the family*. Cambridge, MA: Harvard University Press.
Bergstrom, T. 1989. A fresh look at the rotten kid theorem. *Journal of Political Economy* 97: 1138–1159.
Bergstrom, C., and T. Bergstrom. 1999. Does Mother Nature punish rotten kids? *Journal of Bioeconomics* 1: 47–72.
Bergstrom, T., and R. Cornes. 1983. Independence of allocative efficiency from distribution in the theory of public goods. *Econometrica* 51: 1753–1765.
Bruce, N., and M. Waldman. 1990. The rotten-kid theorem meets the Samaritan's Dilemma. *Quarterly Journal of Economics* 105: 155–165.
Dawkins, R. 1976. *The selfish gene*. Oxford: Oxford University Press.
Hamilton, W. 1964. The genetical evolution of social behavior, I and II. *Journal of Theoretical Biology* 7: 1–52.
Lindbeck, A., and J. Weibull. 1988. Altruism and efficiency: The economics of fait accompli. *Journal of Political Economy* 96: 1165–1182.
Lundberg, S., and R. Pollak. 2003. Efficiency in marriage. *Review of the Economics of the Household* 1: 153–167.

Roundabout Methods of Production

K. H. Hennings

Methods of production are roundabout if they use produced means of production or the services of capital goods as well as those of land and labour, the latter being considered original or primary factors of production.

The concept of roundaboutness of production methods thus draws a distinction between 'rude' and more advanced methods of production, the latter capital good using, the former not. This, however, is not a distinction of much use as even in very primitive economies man-made tools are used. But the concept has been associated with the proposition that more roundabout methods of production yield more output per unit of input, but require more time because the capital goods they use have to be produced, too. This proposition in turn has been associated with the idea that the roundaboutness of production methods depends on the division of labour as well as on the fact that production takes time. The concept therefore plays an important role in those variants of capital theory like the 'Austrian' theory of Böhm-Bawerk (1889) which emphasize the time-consuming nature of production in an economy characterized by division of labour.

The proposition that production processes take time, and that their implementation therefore requires 'advances' in the form of wage goods as well as durable capital goods was fundamental to physiocratic theory as expounded by Quesnay

(1759) and Turgot (1770). It also appears in such classical writers as Ricardo (1817, ch. 1, sect. iv, v). But from Adam Smith (1776) onwards, most classical economists followed Josiah Tucker (1774, p. 24) and linked the use of capital goods to the division of labour rather than the time-consuming nature of production. This can be seen as the consequence of a shift of emphasis from agricultural to industrial production processes. Time requirements in agricultural production are given by nature, and cannot be overcome by an appropriate organization of production processes. Industrial production processes, by contrast, can be, and often are, staggered in such a way that outputs are obtained continuously and the temporal structure of production processes does not matter.

Longfield (1834) combined both aspects by arguing that production takes time on account of the division of labour. Because an increasing division of labour requires more and more different capital goods, the production processes in which they are produced lengthen the overall or composite production process which links the original factors of production to the output obtained with their help as well as the help of intermediate produced means of production. Likewise, Rae (1834) argued that increased division of labour goes with increased durability of capital and hence longer periods of time required for production. Rae also emphasized that, at any time, entrepreneurs have a choice between industrial production processes of different degrees of roundaboutness, i.e. different productivity as well as associated time requirements. These ideas, however, remained outside the classical tradition, which remained wedded to the idea that the division of labour was the main reason for the use of roundabout methods of production. The time element in production was submerged in the conception of capital goods as 'stored-up' or 'congealed' labour (and land), and thus as the result of previous production processes, which had been advanced by Ricardo (1817, ch. 1, sect. iii) and James Mill (1821). On the basis of this conception Senior (1836, 57–8) introduced the distinction between land and labour as 'primary' and capital goods as 'secondary' requisites of production; this distinction was later converted into one between 'original' and 'derived' factors of production. Most authors, however, used these notions in a rather vague way in order to characterize the nature of capital goods, and to assert the advantages of their accumulation.

When formulating his temporal ('Austrian') theory of capital Böhm-Bawerk (1889) built upon this tradition. He went beyond it, however, when he posited the existence of a production function in which the level of output obtained per unit of input was made a function of the degree of roundaboutness of the production method employed which, he argued, was positive but subject to diminishing marginal returns. This required him to define formally the degree of roundaboutness. Considering only one original factor of production, say x, which is used in different stages x_t of the production process, and hence remains 'invested' in it for varying lengths of time s_t, Böhm-Bawerk defined the degree of roundaboutness as the average period of production S as

$$S = \sum_t x_t s_t \Big/ \sum_t x_t$$

Here $\Sigma_t x_t s_t$ is what Jevons (1871, ch. vii) had called the 'amount of investment of capital', while $\Sigma_t x_t$ is Jevons's 'amount of capital invested'.

Böhm-Bawerk used this definition in his proposition referred to above, and erected on it a theory of the role of capital goods in production which issued in a theory of distribution. This 'Austrian' theory of capital was elaborated by Wicksell (1893) and, in modern form, by Faber (1979). The gist of the argument is that more roundabout methods of production are both more capital intensive and more productive, such that the relative availability of capital determines the method of production used, the amount of output obtained per unit of original input, and, via marginal productivity conditions, factor prices.

This theory gave rise to a lengthy debate which ran for almost half a century; for partial summaries, see Kaldor (1937), Haavelmo (1960), and Reetz (1971). Much of it centred on Böhm-Bawerk's definition of the degree of roundaboutness.

When Böhm-Bawerk turned the rather vague notion of roundaboutness into the more precise concept of an average period of production, he tied it to a linear, unidirectional view of production in which original factors of production are turned into raw materials, these with the help of further original factors into capital goods, which in turn help to produce consumer goods with the help of still further original factors. However, the attempt to 'dissolve' in the classical manner all capital goods into various amounts of original factors of production that had helped to produce them was soon seen to involve an infinite historical regress (Steindl 1937). Similarly, if the problem is conceived, as Rae (1834) had done, as a planning problem (i.e. in a forward looking rather than a backward looking manner), taking into account all future effects on output of durable capital goods produced in the present leads to an infinite historical progress. In both cases, therefore, the time span involved is infinite. The answer to these conundrums, which stem from the fact that production processes which use capital goods are characterized by circularity as well as uni-directional linearity, was formulated as early as 1904 by Dmitriev. He demonstrated that in a well-integrated system of production processes the amount of original factors of production used directly and indirectly in the production of final outputs (i.e. consumer goods) could be calculated without resorting to the fiction of going backward or forward in time, and hence without any infinite regress or progress. Dmitriev also showed that the existence of more than one original factor of production does not pose a problem in this calculation. Yet his contribution remained unnoticed until Sraffa (1960) again drew attention to the issue of circularity and linearity in production processes.

Another major difficulty raised by the way in which Böhm-Bawerk concretized the notion of roundaboutness was that it seems superfluous for the analysis of an ongoing production process. As for example Clark (1899) argued, if production processes are appropriately staggered all one needs to observe, once the processes are in operation, are inflows of original inputs and outflows of outputs, without attempting to trace which outputs are due to which inputs, and thus without paying attention to the temporal structure of production processes. It can indeed be shown that in such circumstances Böhm-Bawerk's average period of production is equivalent to the capital–output ratio as used in modern growth theory (Dorfman 1959). However, while correct, this argument applies to steady states only. In dynamic analyses, and particularly in transitions from one steady state to another (Hicks 1973), or when starting more roundabout production processes, the temporal structure of production does play a role, and in these contexts the concept of a period of production may prove useful as a measure of the roundaboutness of production processes.

Yet another difficulty was pointed out by Samuelson (1966) and Steedman (1972): because Böhm-Bawerk's measure of roundaboutness has the nature of an average, there are necessarily various time profiles of original factors of production which give the same average period of production. If in addition a rate of interest is used when calculating the amount of investment of capital (as they argue one should do) before dividing by the amount of capital invested to obtain the average period of production, the latter turns into a function of the rate of interest which is not unique in the sense that it may exhibit reswitching.

Thus, if the concept of differing degrees of roundaboutness of production methods is to be given a precise meaning, it will have to be defined not as an average, but as an absolute measure relating to technical characteristics of the production processes involved. Moreover, it will have to be shown that more roundabout methods of production necessarily yield higher levels of output per unit of original input. These are the crucial assumptions. Without their validity being demonstrated, the notion of roundabout methods of production remains intuitively appealing but fruitless from an analytical point of view.

See Also

▶ Böhm-Bawerk, Eugen von (1851–1914)
▶ Fisher, Irving (1867–1947)
▶ Period of production
▶ Rae, John (1845–1915)

Bibliography

Clark, J.B. 1899. *The distribution of wealth*. New York: Macmillan.

Dmitriev, V.K. 1904. *Ekonomicheskie Ocherki.* Moscow: Richter. Trans. as *Economic essays on value, competition and utility,* Cambridge: Cambridge University Press, 1974.

Dorfman, R. 1959. Waiting and the period of production. *Quarterly Journal of Economics* 73: 351–372.

Faber, M. 1979. *Introduction to modern Austrian capital theory.* Berlin: Springer.

Haavelmo, T. 1960. *A study in the theory of investment*. Chicago: Chicago University Press.

Hicks, J.R. 1973. *Capital and time*. Oxford: Clarendon Press.

Jevons, W.S. 1871. *The theory of political economy.* London: Macmillan.

Kaldor, N. 1937. Annual survey of economic theory: The recent controversy on the theory of capital. *Econometrica* 5: 201–233.

Longfield, M. 1834. *Lectures on political economy.* Dublin: Milliken.

Mill, J. 1821. *Elements of political economy.* London: Baldwin, Craddock & Joy.

Quesnay, F. 1759. *Tableau économique*. Paris: Imprimerie Royale.

Rae, J. 1834. *Statement of some new principles on the subject of political economy.* Boston: Hilliard Gray & Co.

Reetz, N. 1971. *Produktionsfunktion und Produktionsperiode*. Göttingen: Schwartz.

Ricardo, D. 1817. *On the principles of political economy and taxation*. London: Murray.

Samuelson, P.A. 1966. A summing up. *Quarterly Journal of Economics* 80(November): 568–583.

Senior, N.W. 1836. *Political economy.* London: Griffin.

Smith, A. 1776. *An inquiry into the nature and causes of the wealth of nations*. London: Strahan & Cadell.

Sraffa, P. 1960. *Production of commodities by means of commodities*. Cambridge: Cambridge University Press.

Steedman, I. 1972. Jevons's theory of capital and interest. *Manchester School of Economic and Social Studies* 40(March): 31–52.

Steindl, J. 1937. Der historische Regress in der Theorie der Produktionsumwege. *Jahrbücher für Nationalökonomie und Statistik* 145: 143–157.

Tucker, J. 1774. *Four tracts together with two sermons on political and commercial subjects* (written 1758). Gloucester/London: Raikes & Rivington.

Turgot, A.R.J. 1770. Réflexions sur la formation et la distribution des richesses. *Ephémérides du Citoyen.* Trans. as 'Reflections on the formation and distribution of wealth (1766)'. In *Turgot on progress, sociology, and economics,* ed. R.L. Meek, Cambridge: Cambridge University Press, 1973.

von Böhm-Bawerk, E. 1889. *Positive Theorie des Kapitales*. Innsbruck: Wagner. Trans. as *The positive theory of capital,* London: Macmillan, 1891.

Wicksell, K. 1893. *Über Wert, Kapital und Rente.* Jena: Fischer. Trans. as *Value, capital and rent,* London: George Allen & Unwin, 1954.

Rousseau, Jean Jacques (1712–1778)

S. C. Stimson

Political philosopher, moral reformer, citizen of Geneva. Rousseau's economic thought cannot readily be placed within the mainstream of the schools of 18th-century economic discourse. The entire thrust of his work, comprising a sustained argument against the luxury and conspicuous consumption of the rising European bourgeosie of new commerce, implied a sharp rejection of the practices as well as principles of the mercantilist. Rousseau's most explicit contribution to economic thought, a contribution to the *Encyclopédie* entitled 'Economie politique' (vol. v, 1755), significantly preceded publication of the earliest published statement of the Physiocrats, Quesnay's *Maximes générales du gouvernement économique d'un royaume agricole* (1758), and their positions on important issues of property and to a lesser extent taxation bear comparison but are by no means identical. For his position on the right of the State to tax its citizens and the inseparable relationship between justice and the sacred rights of property (see *Political Writings,* I, 234), 'Rousseau appears to have appealed to and hardly superseded Locke. Nonetheless, Rousseau influenced both contemporary and later proponents of economie as well as political reform through his single-minded opposition to economic inequality, his disbelief in the benign effects of unregulated laissez faire, and his attack on what he considered the trivialized conception of liberal public life which accepted the interactions of the market as an adequate substitute for a theory of social relations.

From the earliest writing to bring him public notoriety, a discourse on the question of whether the progress of the Arts and Sciences had tended to the purification or the corruption of morality

(1751), Rousseau was certainly more than an economic opponent of nascent capitalism. His adversary was nothing less than the 'progress' of modern society, and the injury he believed men suffered when labour was socially divided and property distributed under conditions of radical inequality. Writing initially as a moral reformer, Rousseau's work attacked the social institutions and entrenched inequalities of a feudal society in transition which combined remnants of feudal personal dependence with a set of new bourgeois commercial values and individual self-serving relationships later characterized as 'the get ahead spirit' (Tocqueville, *Democracy in America*). Of the principal distinctions of inequality articulated by Rousseau – general wealth, nobility or rank, power and personal merit – he argued consistently that 'wealth is the last to which they are reduced in the end because, being the most immediately useful to well-being and the easiest to communicate, it is easily used to buy all the rest' (*l'Inégalité,* in *Political Writings,* I, 192). Thus the early aesthetic and moral critique of the *Discours sur les Sciences et les Arts,* that science and civilization served directly and indirectly to oppress rather than uplift man's well-being, was given an explicitly social and economic point in a second work, *Discours sur l'origine et les fondements de l'inégalité parmis les hommes* (1754).

> From the extreme inequality of conditions and fortunes, from the diversity of passions and talents, from useless arts, from pernicious arts, from frivolous sciences would come scores of prejudices equally contrary to reason, happiness, and virtue (*Political Writings* I, 193).

In the *Economie politique,* Rousseau does not confine himself to purely economic matters but, consonant with the intellectual style of the period, often blurs what are now distinctive disciplines to offer his tentative reflections on the proper construction of political society. The *Economie* discusses the relations between the family and the State and those differences necessarily separating familial regulation and political authority, the relationship of the individual to the State, and the power and importance of civic education and political law itself to create equal public citizens out of unequal private men. The discussions of both the family and the law would later reappear almost verbatim in Rousseau's more mature political statement, the *Contrat social* (1762). More importantly, the *Economie* also reproduces the essence of his earlier Lockean considerations on the right of property and a lengthy discussion, comprising more than a third of the entire essay, of the problem of taxation. It is these discussions of taxes and property rights that prove problematic for the coherence of Rousseau's economic thought.

With regard to taxes, Rousseau appears in agreement with the adage that in all but the ideal world, taxation, like death, is inevitable. Yet he offers no consistent position on the best type or method of taxation. In the *Economie,* he rejects on principle both a land and a corn tax in favour of a conditional capitation tax and heavy luxury taxes. In his last work, *Le Gouvernement de Pologne* (1772), his proposals are almost exactly the reverse (*Political Writings* I, 269–71; II, 482–4). By this work Rousseau had come to believe that education rather than sumptuary laws was more efficacious in directing the opinions of the citizens toward economic reform.

In regard to private property, it has been commonly noted that Rousseau's suggestion in the *Economie* of social ties (however rudimentary) and property relations predating the State and providing 'le vrai fondement de la société civile, et le vrai garant des engagements des citoyens' (*Political Writings,* I, 259), represents a lapse into an individualism sharply at odds with the organic and collectivist position of his most famous political concept, the General Will. On Rousseau's account, submission to the General Will means submission to a will which is so inflexible that no individual can escape and dominate others, and in which 'the [property] right exercized by each individual over his own particular share is always subordinate (*toujours subordoné*) to the right (*droit*) of the Community over everything (*à sur tous*)' (*Political Writings,* II, 39). Liberty is thus understood not as the removal of all 'chains' of obligation, but rather as the substitution of legitimate bonds of sovereign law for the personal dependence fostered by inequality of wealth.

The theory of government developed in the *Economie* and the later *Contrat social* expresses the sovereign law of the General Will, and is therefore legitimate, because its public policies eliminate the manipulation of political power by the wealthy. However, Rousseau did not expect government to eliminate all inequality. Although the author of *Contrat social* was interpreted by the leaders of the French Revolution, and most notably, Robespierre, as a proponent of a radical political egalitarianism, such a view sits uneasily within a general understanding of Rousseau's work. The fit is particularly poor with his final writing on Poland. There Rousseau is explicit in his claim that civil society is dependent upon rulers and that so long as rights and duties under the rule of law are respected and private citizens are unable to direct public affairs in the service of their private interests, inequality of authority is accepted. In this sense, Rousseau's own thought cannot be termed 'socialism', either economically or politically, though its overarching concern with the human and moral effects of extreme economic inequality has often supplied theoretical foundations to which later socialists have turned for support.

Where applicable, my translations have been matched to the standard translation of Roger D. Masters and Judith R. Masters, *The First and Second Discourses,* New York: St Martins, 1964.

Selected Works

1915. *The political writings of Jean Jacques Rousseau*, 2 vols. ed. C.E. Vaughan, Cambridge: Cambridge University Press. Reprinted, Oxford: Basil Blackwell, 1962.
1964. *The first and second discourses*. Trans. R.D. and J.R. Masters. New York: St Martin's.
1964–80. *Oeuvres complètes*, 4 vols. Published under the supervision of Bernard Gangnebin and Marcel Raymond, Paris: Gallimard, Bibliothèque de la Pléiade.

Bibliography

Sklar, J.N. 1969. *Men and citizens: A study of Rousseau's social theory*. London: Cambridge University Press.

Roy Model

James J. Heckman and Christopher Taber

Abstract

The Roy (1951) model of self-selection on outcomes is one of the most important models in economics. It is a framework for analysing comparative advantage. The original model analysed occupational choice with heterogeneous skill levels and has subsequently been applied in many other contexts. This article presents the model, discusses its identification, and describes some empirical applications based on the model.

Keywords

Comparative advantage; Discrete choice; Estimation; Identification; Occupational choice; Roy model; Selection bias; Self-selection; Separability

JEL Classifications

D3; J3

The Roy (1951) model of self-selection on outcomes is one of the most important models in economics. It is a framework for analysing comparative advantage. The original model analysed occupational choice with heterogeneous skill levels and has subsequently been applied in many other contexts. We first discuss the model. We then summarize what is known about identification of the model. We end by describing some applications based on the model and its extensions.

R

Basic Models

In the original Roy (1951) model, agents can pursue one of two possible occupations: hunting and fishing. They cannot pursue both at the same time. There is no interaction among agents so the choice of one agent does not affect the choice of another agent either through prices or through external effects. Let π_f and π_r be the price of fish and rabbits respectively in the village. Let F_i denote the number of fish that individual i would catch if he chooses to fish. Similarly let R_i denote the number rabbits he would catch. Then individual i's wage is

$$w_{fi} = \pi_f F_i$$

if he fishes and

$$w_{ri} = \pi_r R_i$$

if he hunts. The income that worker i receives for working in sector j is thus proportional to π_j (where $j \in \{r, f\}$). If workers are pure income maximizers, they will choose the occupation with higher income. Thus a worker chooses to fish if $w_{fi} > w_{ri}$. If F_i and R_i are continuous random variables, $\Pr(\pi_r R_i = \pi_f F_i) = 0$, so the indifference set is negligible. A fundamental aspect of the Roy model is that it allows for heterogeneity in (F_i, R_i). This heterogeneity can arise from inherent ability differences or human capital investment.

An important issue is self-selection. Under what conditions will the best workers self-select into an occupation? Will people who self-select be above average? For example, for fishing, under what conditions is the average productivity of people working in the fishing sector above the population mean productivity:

$$E\left[\log(F_i) \mid \pi_f F_i \geq \pi_r R_i\right] > E[\log(F_i)]?$$

Assume, as did Roy (1951), that log skills are jointly normally distributed

$$\begin{bmatrix} \log(F_i) \\ \log(R_i) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_f \\ \mu_r \end{bmatrix}, \begin{bmatrix} \sigma_{ff} & \sigma_{fr} \\ \sigma_{fr} & \sigma_{rr} \end{bmatrix}\right).$$

Then it is straightforward to show that

$$E\left[\log(F_i) \mid \pi_f F_i \geq \pi_r R_i\right] = \mu_f + \underbrace{\frac{(\sigma_{ff} - \sigma_{fr})}{\sigma} \lambda\left(\frac{\log(\pi_f) - \log(\pi_r) + \mu_f - \mu_r}{\sigma}\right)}_{selection effect}$$

where σ^2 is the variance of $\log(F_i/R_i)$ and $\lambda(\cdot)$ is the inverse Mills ratio. (See selection bias and self-selection.)

The function λ is positive but decreasing in its arguments with $\lim_{\pi f \to \infty} \lambda(\cdot) = 0$. The selection effect is the second term on the right-hand side of this expression. There is a parallel expression for $E(\log R_i | \pi_F F_i \leq \pi_R R_i)$ with the subscripts f and r interchanged.

Recall that $E[\log(F)] = \mu_f$ and that λ and σ must both be positive. Therefore, the question of whether there is positive selection into fishing depends only upon the sign of σ_{ff}-σ_{fr}. It does not depend on skill prices. Moreover, since

$$\sigma^2 = (\sigma_{ff} - \sigma_{fr})\sigma_{rr} - \sigma_{fr} > 0,$$

at least one of $\sigma_{ff} - \sigma_{fr}$ and $\sigma_{rr} - \sigma_{fr}$ must be positive. Thus, there must be positive selection into one of the occupations, and there can be positive selection into both.

If, however, there is positive selection into only one occupation, the question arises as to which occupation is most likely to have positive selection. Roy argues that relatively simple tasks (setting traps for rabbits in his case) can be described by a small standard deviation of skill. For more difficult skills (fishing in his example) the standard deviation will be relatively higher as there is a bigger difference between the most skilled and the least skilled. Thus, if fishing is the more difficult task, $\sigma_{ff} > \sigma_{rr,}$ there must be positive selection into fishing (that is, $E(\log(F_i)|\pi_f F_i \geq \pi_r R_i) > E(\log(F_i))$).

Whether there is positive selection into hunting depends on the value of σ_{fr} relative to σ_{rr}. When $\sigma_{fr} < 0$, we will see positive selection into hunting. At the other extreme, if hunting and fishing are perfectly correlated, then σ_{fr} must be larger than σ_{rr}, and there is negative selection into hunting. Intuitively, since F and R are perfectly positively correlated, and F is more dispersed, persons with low values of F can avoid low incomes by using their value of R. Persons with high values of F (and should fish because the upper tail of F is more dispersed. For cases in between, either positive or negative selection is possible depending on the sign of σ_{rr} -σ_{fr}. Heckman and Honoré (1990) generalize this result to a broader class of distribution functions.

This model has been generalized in a number of ways. There can be more than two occupational choices. Following Heckman and Sedlacek (1985), one can assume that individuals possess a vector of skills S_i and that different tasks use the different skills according to the function $T_j(S_i)$. We still let π_j denote task prices so that we can write an individual's wage at task j as

$$w_{ji} = \pi_j T_j(S_i).$$

Another extension of the model allows individuals to care about aspects of the job other than just their wages (see Heckman and Sedlacek 1985). Let $U_{ji}(w)$ be the utility that individual i would receive from performing task j under wage level w. This allows for some tasks (such as playing basketball) to be generally preferred to more unpleasant tasks (such as cleaning bathrooms). Individuals then choose the occupation that yields the highest level of utility for them $U_{ji}(w_{ij})$. This is the generalized Roy model in which the generalization comes in the agent decision rules.

The generalized Roy model can be trivially extended to a model of labour force participation by allowing non-market work to be one of the tasks. To see this, let $j = 0$ denote the home sector as in Gronau (1974) and Heckman (1974). Of course, in general, there will not be a market price for home-produced goods, but one can interpret $T_0(S_i)$ as the value of goods produced at home. One could also assume that staying at home is pure leisure in which $T_0(S_i)=$ 0, but people enjoy staying at home $U_{0i}(0) > U_{ji}(0)$ for $j > 0$. The Roy model has been generalized to allow for uncertainty in agent decision making in Cunha, Heckman and Navarro (2005). See the reviews in Heckman, Lochner and Todd (2006) and Cunha and Heckman (2007).

Identification

The economics of these models is simple, but identification and estimation are considerably more difficult. Heckman and Honoré (1990) consider identification of the basic Roy model with two occupations and income maximization. They consider two different cases: (a) the standard Roy model, in which the two occupations represent two different sectors of the economy and the econometrician has data on wages in both sectors; and (b) a case motivated by labour supply in which the econometrician has wage data from one sector (the market sector) but not from the other (the home sector). It is important to keep in mind that the comparative advantage decision at the heart of the Roy Model is just one factor that can lead to selection bias. selection bias and self-selection discusses the more general framework for thinking about sample selection and also discusses in some detail how the Roy model fits into this framework.

Heckman and Honoré (1990) consider identification from a single cross section. When one can observe wages in both sectors, under log normality, the Roy model is identified even without any regressors in the model. However, when one relaxes the log normality assumption, without regressors in the outcome equation the model is no longer identified. This is true despite the strong assumption of agent income maximization.

Heckman and Honoré (1990) provide conditions under which one can identify these models using variation across markets, or by using variation in observables within a market. To see the intuition behind the latter case, consider the model in which

$$\log(F_i) = g_f(Z_{fi}, X_i) + \varepsilon_{fi} \log(R_i)$$
$$= g_r(Z_{ri}, X_i))\varepsilon_{ri},$$

and prices are normalized to 1. In this context, it is helpful for identification to have an exclusion restriction – that is, a variable Z_{fi} that varies separately from (X_i, Z_{ri}) and a variable Z_{ri} that varies separately from (X_i, Z_{fi}). As long as there is sufficient variation in the excluded variables, Heckman and Honoré (1990) show that with a location normalization the full model is identified provided that (ε_{fi}, ε_{ri}) are independent of (Z_{fi}, Z_{ri}, X_i), that is, they identify g_f, g_r and the joint distribution of (ε_{fi}, ε_{ri}). (They also establish identification when only one sector's output is observed.)

To see the intuition for why the model is identified, consider an 'identification at infinity' argument. For convenience, take the location normalization to be

$$E(\varepsilon_{fi}) = E(\varepsilon_{ri}) = 0.$$

Suppose that g_r is such that for any x, say x_0,

$$\lim_{zr \to -\infty} g_r(z_r, x_0) = -\infty.$$

Let $J_i \in \{f, r\}$ be an indicator of the occupation that was chosen by individual i. Then

$$\lim_{z_r \to -\infty} E[\log(F_i) | J_i = f, X_i = x, Z_{fi} = z_f, Z_{ri} = z_r]$$
$$= g_f(z_f, x) + \lim_{z_r \to -\infty} E\Big(\varepsilon_{fi} | g_f(z_f, x) + \varepsilon_{fi} > g_r(z_r, x)$$
$$+ \varepsilon_{ri}, X_i = x, Z_{fi} = z_f, Z_{ri} = z_r\Big) = g_f(z_f, x).$$

By varying (z_f, x) one can trace out g_f. This occurs because conditioning on the event

$$g_f(z_f, x) + \varepsilon_f > g_r(z_r, x) + \varepsilon_r$$

becomes irrelevant as z_r becomes arbitrarily small. Identification of g_r is analogous using variation in z_f.

To identify the joint distribution of (ε_{fi}, ε_{ri}) note that from the data one can observe

$$\Pr(J_i = f, \log(F_i) < s | X_i = x, Z_{fi} = z_f, Z_{ri} = z_r)$$
$$= \Pr\Big(g_f(z_f, x) + \varepsilon_{fi} > g_r(z_r, x) + \varepsilon_{ri}, g_f(z_f, x) + \varepsilon_{fi}$$
$$< s | X_i = x, Z_{fi} = z_f, Z_{ri} = z_r) = \Pr\Big(\varepsilon_{fi} - \varepsilon_{ri} < g_f$$
$$(z_f, x) - g_r(z_r, x), \varepsilon_{fi} < s - g_f(z_f, x) s | X_i = x,$$
$$Z_{fi} = z_f, Z_{ri} = z_r)$$

which is the cumulative distribution function of (ε_{fi} −ε_{ri}, ε_{fi}) evaluated at the point (g_f (z_f, x)-g_r(z_r, x), s -g_f (z_f, x)). By varying the point of evaluation one can identify the joint distribution of (ε_{fi} −ε_{ri}, ε_{fi}) from which one can derive the joint distribution of (ε_{fi}, ε_{ri}). Thus the model is identified. Heckman and Honoré (1989) also present conditions for identification of a competing risk version of a Roy model when there are no exclusion restrictions ($Z_r = X = Z_f$) but g_f and g_r can be independently varied. Buera (2006) makes stronger differentiability assumptions and relaxes the separability assumption in the choice equation. He also identifies a Roy model without exclusion restrictions.

Identifying the more general model where individuals choose fishing when

$$U_{fi}(w_{fi}) > U_{ri}(w_{ri})$$

is possible under a variety of assumptions. Consider the separable case in which

$$U_{fi}(w_{fi}) - U_{ri}(w_{ri}) = h(Q_i, Z_{fi}, Z_{ri}, X_i) + v_i$$

where Q_i is an additional variable that might affect the relative utilities of the two options. The function h is identified up to a normalization (see, for example, Matzkin 1992).

Identification of parts of the model follows from the preceding reasoning. If there is a variable that affects sectoral choice, but not wages as a fisherman, we can identify g_r. Note that this exclusion restriction could be in the form of either Q_i or Z_{ri}. We can then identify the joint distribution of (v_i,ε_{ri}) using an argument analogous to the above. Using the same argument we can identify g_f and the joint distribution of (v_i, ε_{fi}). A formalization of this argument can be found in Heckman (1990) for the case in which h is linear and is extended in

Heckman and Smith (1998), Carneiro, Hansen and Heckman (2003) and Heckman and Navarro (2007). One cannot, without further assumptions, identify the joint distribution of (v_i, ε_{ri}, ε_{fi}). (Abbring and Heckman 2007, present conditions for identification of the joint distribution by restricting dependence relations. See also Aakvik et al. 2005.)

If one is interested in evaluating policies in which wages can change, this reduced form model is not enough since there is no separation of wage effects from non-wage effects in the choice model. Assume further that we can write

$$\begin{aligned} h(Q_i, Z_{fi}, Z_{ri}, X_i) + v_i &= \alpha_1 F_i - \alpha_2 R_i) h^*(Q_i, Z_{fi}, Z_{ri}, X_i) + v_i^* \\ &= \alpha_1 g_f(Z_{fi}, X_i) - \alpha_2 g_r(Z_{ri}, X_i) + h^*(Q_i, Z_{fi}, Z_{ri}, X_i) + \alpha_1 \varepsilon_{fi} \\ &\quad - \alpha_2 \varepsilon_{fi} + v_i^* \end{aligned}$$

Identification of this model is possible if there are exclusion restrictions in Z_{fi} and Z_{ri}, that is, if there are components of Z_{fi} and Z_{ri} that do not affect h^*. Under sufficient variation of these variables and imposing a normalization, the model is identified. An interesting special case of the model is when $\alpha_1 = \alpha_2$. In this case one needs a somewhat weaker exclusion restriction in that one could use variation in X_i. That is, we could use a variable that affects labour market outcomes, but not sectoral choice directly.

Empirical Models

There are many examples that build on the Roy model, but in labour economics three stand out. The earliest empirical application of this model is to the labour supply decision (Heckman 1974; Gronau 1974). We refer interested readers to labour supply rather than discuss these models explicitly. The second application is to occupational choice, which is most closely linked to the original Roy model. The third, and perhaps most well known, application is to education.

We start by describing the empirical applications of the model to education. Willis and Rosen (1979) consider a model in which students decide whether to attend college. Students may have a comparative advantage in either the college sector or the high school sector. Their model assumes that decisions about schooling are made in an environment of perfect certainty on the principle of income maximization. They assume access to outcome measures in two periods. The decision to attend college depends on interest rates which are not observable to the econometrician. (One could reinterpret their model as a generalized Roy model if one interprets the interest rate as representing utility differences rather than interest rates.) Semiparametric identification requires two types of exclusion restrictions: a variable that influences the decision to attend college but not directly wages, and a variable that influences wages but not the decision to attend college directly. For the former type of exclusion restriction, Willis and Rosen (1979) use family background variables, arguing that they will be correlated with interest rates but uncorrelated with wages. For the latter type they use test scores, arguing that they are related to skill as in the Roy model, but unrelated to the interest rate.

Although they discuss comparative advantage in the labour market, as did Roy, they do not present direct empirical evidence on this question because they cannot estimate the joint distribution of schooling outcomes across both choices. They present some indirect evidence on the importance of comparative advantage in the labour market because they can identify the counterfactual means of what college students would earn if they had been high school students and what high schools students would earn had they been college students.

There are many extensions of this model, including Taber (2000), Cameron and Taber (2004) and Heckman, Lochner and Todd (2006). Cunha, Heckman and Navarro (2005) and Cunha and Heckman (2007) extend the model to allow for uncertainty, to identify agent information and to directly test for comparative advantage in the labour market by identifying the joint distribution of outcomes for the two counterfactual states (college and high school).

In a series of papers, Heckman, Lochner and Taber (1998a, b, c) estimate a general equilibrium version of this model. That is, they allow the skill prices π_j to be endogenous. They show that accounting for equilibrium effects is essential for estimating the impact of policy on earnings inequality. In particular Heckman, Lochner and Taber (1998b) show that ignoring equilibrium effects overstates the impact of a tuition subsidy on college enrolment by an order of magnitude. They also decompose the policy effect on earnings inequality into its various components.

Other papers estimate a Roy model of occupational choice. Most notably, Heckman and Sedlacek (1985, 1990) estimate models in which workers choose between industrial sectors. In some cases they allow for non-market work. They show how to estimate the model, but reject a pure Roy model. They show instead that a more general model with utility maximization and non-participation can fit the data well. Gould (2002) extends this framework to address the changing wage structure. He shows that workers choose sectors to maximize their comparative advantage and that this activity tends to decrease earnings inequality. However, he shows that the importance of this effect decreases over time as sectors increasingly value more similar skill sets.

Keane and Wolpin (1997) and Eckstein and Wolpin (1999) estimate dynamic discrete choice models of occupational and educational choice that extends the Roy model to a dynamic setting with uncertainty with serially independent shocks. Agents in their model make labour supply, education and occupational choice simultaneously. Heckman and Navarro (2007) present a nonparametric identification analysis of a dynamic discrete choice model with serially correlated shocks. Abbring and Heckman (2007) survey the dynamic discrete choice literature, including these papers.

See Also

▶ Labour Economics
▶ Labour Supply
▶ Returns to Schooling
▶ Selection Bias and Self-Selection

Bibliography

Aakvik, A., J.J. Heckman, and E.J. Vytlacil. 2005. Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs. *Journal of Econometrics* 125: 15–51.

Abbring, J.H., and J.J. Heckman. 2007. Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In *Handbook of econometrics*, ed. J. Heckman and E. Leamer, Vol. 6. Amsterdam: North-Holland.

Buera, F.J. 2006. Non-parametric identification and testable implications of the Roy model. Unpublished manuscript, Department of Economics, Northwestern University.

Cameron, S.V., and C. Taber. 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112: 132–182.

Carneiro, P., K. Hansen, and J.J. Heckman. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. 2001 Lawrence R. Klein Lecture. *International Economic Review* 44: 361–422.

Cunha, F., and J.J. Heckman. 2007. Identifying and estimating the distributions of ex post and ex ante returns to schooling: A survey of recent developments. *Labour Economics* 14: 870–893.

Cunha, F., J.J. Heckman, and S. Navarro. 2005. Separating uncertainty from heterogeneity in life cycle earnings. The 2004 Hicks Lecture. *Oxford Economic Papers* 57: 191–261.

Eckstein, Z., and K.I. Wolpin. 1999. Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica* 67: 1295–1339.

Gould, E.D. 2002. Rising wage inequality, comparative advantage, and the growing importance of general skills in the United States. *Journal of Labor Economics* 20: 105–147.

Gronau, R. 1974. Wage comparisons – a selectivity bias. *Journal of Political Economy* 82: 1119–1143.

Heckman, J.J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42: 679–694.

Heckman, J.J. 1990. Varieties of selection bias. *American Economic Review* 80: 313–318.

Heckman, J.J., and B.E. Honoré. 1989. The identifiability of the competing risks model. *Biometrika* 76: 325–330.

Heckman, J.J., and B.E. Honoré. 1990. The empirical content of the Roy model. *Econometrica* 58: 1121–1149.

Heckman, J.J., L.J. Lochner, and C. Taber. 1998a. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics* 1: 1–58.

Heckman, J.J., L.J. Lochner, and C. Taber. 1998b. General-equilibrium treatment effects: A study of tuition policy. *American Economic Review* 88: 381–386.

Heckman, J.J., L.J. Lochner, and C. Taber. 1998c. Tax policy and human-capital formation. *American Economic Review* 88: 293–297.
Heckman, J.J., L.J. Lochner, and P.E. Todd. 2006. Earnings equations and rates of return: The Mincer equation and beyond. In *Handbook of the economics of education*, ed. E.A. Hanushek and F. Welch. Amsterdam: North-Holland.
Heckman, J.J., and S. Navarro. 2007. Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics* 136: 341–396.
Heckman, J.J., and G.L. Sedlacek. 1985. Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market. *Journal of Political Economy* 93: 1077–1125.
Heckman, J.J., and G.L. Sedlacek. 1990. Self-selection and the distribution of hourly wages. *Journal of Labor Economics* 8(1, Part 2): S329–S363.
Heckman, J.J., and J.A. Smith. 1998. Evaluating the welfare state. In *Econometrics and economic theory in the twentieth century: The Ragnar Frisch Centennial Symposium*, ed. S. Strom. New York: Cambridge University Press.
Keane, M.P., and K.I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105: 473–522.
Matzkin, R.L. 1992. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica* 60: 239–270.
Roy, A. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146.
Taber, C.R. 2000. Semiparametric identification and heterogeneity in discrete choice dynamic programming models. *Journal of Econometrics* 96: 201–229.
Willis, R.J., and S. Rosen. 1979. Education and self-selection. *Journal of Political Economy* 87(5, Part 2): S7–S36.

Roy, René François Joseph (1894–1977)

Maurice Allais

Keywords

Demand functions; Index numbers; Roy, R. F. J.

JEL Classifications

B31

René Roy was born in Paris on 21 May 1894. He entered the Ecole Polytechnique in 1914, and joined the army on 15 August 1914. He was seriously wounded on 14 April 1917 at the Chemin des Dames, as a result of which he was blinded at the early age of 23. This tragedy, which meant the collapse of all his youthful hope and dreams, brought him to the slough of despond, and exceptional spiritual strength alone enabled him eventually to accept the unacceptable with serenity and to undertake a double career as an engineer and economist that was to last 60 years.

He studied at the Ecole Polytechnique (from 1918 to 1920) graduating first in his year, and then at the Ecole Nationale des Ponts et Chaussées (1920–2). He entered the Ministry of Public Works and Transport as a state engineer in 1922, specializing in problems of local railway networks and urban transport until his retirement in 1964. He died in Paris in 1977.

In parallel with this activity, he became Professor of General Political Economy and Social Economy at the Ecole des Ponts et Chaussées in 1929, and Professor of Econometrics at the Statistical Institute of the University of Paris in 1931. In 1949 he taught econometrics at the Ecole d'Application de l'Institut National de la Statistique et des Etudes Economiques (School of Instruction of the National Institute of Statistics and Economic Studies). From 1947 he was in charge of an Econometrics Seminar at the National Centre of Scientific Research.

He was elected President of the Paris Statistical Society in 1949 and of the International Econometrics Society in 1953. He was also a fellow of the International Statistical Society (1949), a member of the Academy of Moral and Political Science (1951), and an honorary fellow of the Royal Statistical Society (1957). He received the degree of Doctor Honoris Causa from the University of Geneva in 1964.

René Roy's research was focused mainly on transport, demand functions, economic indices, fields of choice and their respective relationships. His main published works are *Le régime économique des voies ferrées d'intérêt local* (1925),

R

his doctoral thesis; *La demande de biens de consommation directe* (1935); *De l'utilité – contribution à la théorie des choix* (1942); 'Les nombres indices' (*Journal de la Société de Statistique de Paris*, 1949); and *Eléments d'économétrie* (1970). In addition, in collaboration with François Divisia and Jean Dupin, he published in 1953–4 *A la recherche du franc perdu*, whose three volumes cover the movement of prices, production and wealth respectively in France from 1914 to 1950. Roy's analysis of the basic relationships of demand functions and price and quantity index numbers are contained in his 1949 publication.

René Roy's ability to analyse very difficult questions and constantly stay abreast of the main publications of his era was a truly remarkable achievement for a totally sightless person. He showed that accomplishment is possible in the face of an irremediable adversity by dint of unremitting energy associated with remarkable intelligence. His book *Vers la lumière* (1930) gives us his message as a blind man.

See Also

▶ Indirect Utility Function

Selected Works

1925. *Le régime économique des voies ferrées d'intérêt local*. Paris.

1930. *Vers la lumière*. Paris: Bibliotheque-Charpentier.

1935. *Le demande de biens de consommation directe*. Paris: Hermann.

1942. *De l'utilité – contribution à la théorie des choix*. Paris: Hermann.

1949. Les nombres indices. *Journal de la Société de Statistique de Paris* 90(1–2): 15–34.

1953–4. (With F. Divisia and J. Dupin.) *A la recherche du franc perdu*. Paris: Société d'Editions Hommes et Mondes.

1970. *Eléments d'économétrie*. Paris: Presses Universitaires de France.

Royal Economic Society

Roger Middleton

Keywords

American Economic Association; British Association for the Advancement of Science; British Economic Association; Edgeworth, F. Y.; Harrod, R. F.; Keynes. J. M.; Marshall, A.; Royal Economic Society; Royal Statistical Society

JEL Classifications

B1

Originally founded in 1890 as the British Economic Association (BEA), the Royal Economic Society (RES) assumed its current title in 1902 when it obtained a Privy Council charter and royal patronage. The RES is now unquestionably the leading organization of professional economists in Britain, with its flagship publication the *Economic Journal* (*EJ*), a world-class general journal for theoretical and applied research (having in 2004 an International Statistical Institute, ISI, journal citation ranking of 15/172 and 1.723 impact factor). Such dominance, however, has not always been the case and was not easily achieved, with the RES's fortunes, like those of many other long-established economics associations, subject to a changing complex of pressures, including at times competitors.

The RES was the eventual institutional result of the long process whereby political economy was transformed into economics as the 'sciences of the social' were dissolved and reconstructed into the modern social sciences in the second half of the 19th century. The establishment of the BEA followed a long period of consultations over whether a new learned society was required to propagate what has come to be known in the modern literature as Marshall's mission for the professionalization of economics (Middleton 1998), or whether the Royal Statistical Society (established 1834) and/or Section F (Economics and Statistics, established 1835) of the British Association for the Advancement of Science, were sufficient vehicles.

The year 1890 witnessed also the publication of Marshall's *Principles of Economics* and the completion of the first instalment of the first Palgrave dictionary, which was published the following year, as was the first issue of the *EJ*. For Keynes (1940, p. 409), these happy concurrences made this the beginnings of the 'modern age of British economics'. This now looks somewhat questionable: the BEA and *EJ* were part of a pre-professionalization trend towards 'clearer demarcation and definition of the field of scholarly endeavour' (Kadish and Freeman 1990, p. 23) and the achievements of the professionalization agenda were as yet limited. That this was the case is apparent from what was one of the central issues of the debate preceding the BEA's formation: would it be a closed society of professional economists (at this time, it was still not true that a majority of these consisted of academics) or would it be open to all, however imperfect their claim to be called an economist? Mindful of the contemporaneous example of the American Economic Association (AEA), where many of its leading figures had been deeply involved in methodological and policy disputations which had been unhelpful to the professionalization agenda, the BEA's founders resolved to follow a more cautious and restricted policy than its transatlantic counterpart: membership was not open but dependent upon a candidate's approval by Council (with the chosen designated as Fellows from 1902 to 1964 under the Royal Charter); the BEA was to refrain from organizing discussions and conferences, partly for fear of exposing the substantial differences of opinion within their ranks; and, in its early years, the Council routinely chose a prominent public figure rather than an academic as its president, beginning with the then Chancellor of the Exchequer, G. J. Goschen.

While the early years of the BEA were often difficult, the *EJ* was an undoubted success, and this notwithstanding the rival Oxford publication, the *Economic Review,* which had been launched a few months earlier (and survived until 1914). The *EJ* was not conceived as a specialist publication for an exclusively academic or quasiacademic economics audience, but instead as a 'means of disseminating economic truth amongst readers from all walks of life, while setting new standards of economic investigation' (Kadish and Freeman 1990, pp. 36–7). This general informational role was to endure strongly until the 1970s and was still be present in the 1990s when, as the journal became increasingly internationalized, it almost exclusively focused on scholarly papers and its policy forum, with less emphasis on book reviews and with reports on its learned society activities hived off to the website (http://www.res.org.uk).

The *EJ's* initial editor was F. Y. Edgeworth, followed by Keynes (singularly and jointly, 1912–45) and then R. F. Harrod (1945–61), during which time the *EJ* consolidated its status as the leading British economic journal, despite the appearance of several rivals. Cambridge, Oxford and London economists initially dominated, but from the 1970s onwards, as the balance of professional influence and authority shifted away from the older centres, the Council had to respond to pressures to make the RES a more democratic organization. Concurrently, the *EJ* editors found that the ever-increasing scale, scope and technical nature of the discipline necessitated increased personnel to provide expert opinion on journal submissions. An editorial board was established in 1971 and has since evolved; most importantly, it now has more foreign (largely, but not exclusively, US) than British economists.

RES membership is now over 3,300 individuals, of whom 60 per cent are not British residents, and there are a further 2,400 institutional subscribers to the *EJ* and the *Econometrics Journal* (established 1998). Notwithstanding the internationalization of the *EJ*, the RES mission statement remains essentially British: to be the 'professional association which promotes the encouragement of the study of economic science in academic life, government service, banking, industry and public affairs'. Increasingly an umbrella organization for a large number of activities, mainly but not exclusively to do with university economics, the RES maintains also its activities as a major publisher of scholarly editions, of which the 30-volume collected writings of Keynes is one of its major achievements, even if it did nearly bankrupt the Society. It has also finally resolved one of the issues of disagreement at its creation: since 1990 the RES has operated an annual conference, selected papers from which appear in a special issue of the *EJ*, itself now enlarged for its second century from a quarterly to a bimonthly publication.

See Also

- ► Keynes, John Maynard (1883–1946)
- ► Marshall, Alfred (1842–1924)

Bibliography

Coats, A.W. 1968. The origins and early development of the Royal Economic Society. *Economic Journal* 78: 349–371.

Kadish, A., and R.D. Freeman. 1990. Foundation and early years. In *A century of economics: 100 years of the Royal Economic Society and the Economic Journal*, ed. J.D. Hey and D. Winch. Oxford: Basil Blackwell.

Keynes, J.M. 1940. The Society's jubilee, 1890–1940. *Economic Journal* 50: 401–409.

Middleton, R. 1998. *Charlatans or Saviours? Economists and the British economy from Marshall to Meade*. Cheltenham: Edward Elgar.

Rubin Causal Model

Guido W. Imbens and Donald B. Rubin

Abstract

The Rubin Causal Model (RCM), a framework for causal inference, has three distinctive features. First, it uses 'potential outcomes' to define causal effects at the unit level, first introduced by Neyman in the context of randomized experiments and randomization-based inference, but not used formally in non-randomized studies or with other modes of inference until Rubin (1974, 1975). Second is its formal use of a probabilistic assignment mechanism, which mathematically describes how treatments are given to units, with possible dependence on background variables and the potential outcomes themselves. Third is an optional probability distribution on all variables, including the potential outcomes, which thereby unifies frequentist and model-based forms of statistical inference for causal effects within one framework.

Keywords

Assignment mechanism; Assignment-based inference; Bayesian inference; Causal inference; Fisher, R. A.; Haavelmo, T.; Hurwicz, L.; Instrumental variables; Interval estimates; Markov chain Monte Carlo methods; Matching; Multiple imputation; Neyman, J.; Posterior predictive distribution; Potential outcomes; Principal stratification; Probability; Program evaluation; Propensity scores; Randomization-based inference; Randomized experiment; Regression coefficients; Roy model; Rubin causal model; Simultaneous equations models; Tinbergen, J.; treatments; Units

JEL classifications

C14

The Rubin Causal Model (RCM) is a formal mathematical framework for causal inference, first given that name by Holland (1986) for a series of previous articles developing the perspective (Rubin 1974, 1975, 1976, 1977, 1978, 1979, 1980). There are two essential parts to the RCM, and a third optional one. The first part is the use of 'potential outcomes' to define causal effects in all situations – this part defines 'the science', which is the object of inference, and it requires the explicit consideration of the manipulations that define the treatments whose causal effects we wish to estimate. The second part is an explicit probabilistic model for the assignment of 'treatments' to 'units' as a function of all quantities that could be observed, including all potential outcomes; this model is called the 'assignment mechanism', and defines the structure of experiments designed to learn about the science from observed data or the acts of nature that lead to the observed data. The third possible part of the RCM framework is an optional distribution on the quantities being conditioned on in the assignment mechanism, including the potential outcomes, thereby allowing model-based Bayesian 'posterior predictive' (causal) inference. This part of the RCM focuses on the model-based analysis of observed data to draw inferences for causal effects, where the observed data are revealed by applying the assignment mechanism to the science. A full-length text that discusses estimation and inference for causal effects from this perspective is Imbens and Rubin (2006).

Implications of the RCM for Research Design

Before defining each of these three parts of the RCM, it is helpful to consider the implications of this structure for applied research about causal effects. The first part implies that we should always start by carefully defining all causal estimands (quantities to be estimated) in terms of potential outcomes, which are all values that could be observed in some real or hypothetical experiment that compares the results under an active treatment with the results under a control treatment. That is, causal effects are defined by a comparison of (*a*) the values that would be observed if the active treatment were applied and (*b*) the values that would be observed if, instead, the control treatment were applied. This step contrasts with the common practice of defining causal effects in terms of parameters in some model, where the manipulations defining the active versus control treatments are often left implicit and ill-defined, with the resulting causal inferences correspondingly weak and ill-defined. This first part can be completely abstract and can take place before any data are observed or even collected. In the RCM, however, there is 'no causation without manipulation' (Rubin 1975, p. 238), where the manipulation (that is, the treatment) could be real or hypothetical. The collection of potential outcomes with and without this manipulation defines the scientific objective of causal inference in all studies, whether randomized, observational or entirely hypothetical.

The second part of the RCM, the assignment mechanism, implies that, given the defined science, we should continue by explicating the design of the real or hypothetical study being used to estimate that science. The assignment mechanism describes why some study units will be (or were) exposed to the active treatment and why other study units will be (or were) exposed to the control treatment, and the reasons are formalized by the mathematical statement of the assignment mechanism. When the study is a true experiment, the assignment mechanism may involve the consideration of background (that is, pretreatment assignment) variables for the purpose of creating strata of similar units to be randomized into treatment and control, thereby improving the balance of treatment and control groups with respect to these background variables (that is, covariates). A true experiment automatically cannot use any outcome (post-treatment) variables to influence design because they are not yet observed. If the observed data were not generated by a true experiment, but rather by non-randomized observational data, there still should be an explicit design phase. That is, in an observational study, the same guidelines as in an experiment should be followed.

More explicitly, the design step in the analysis of an observational data set for causal inference should structure the data to approximate (or reconstruct or replicate) a true randomized experiment as closely as possible. In this design step, the researcher never uses or even examines any outcome data but rather identifies subsets of units such that the treatments can be thought of as being randomly assigned within the subsets. This assumed randomness of treatment assignment is assessed by examining, within these subsets of units, the similarity of the distributions of the covariates in the treatment group and in the control group. Because this design step is focused on creating these subsets of units with balanced distributions of covariates between treatment and control groups, and never uses outcome data, the researcher cannot select a design to produce a desired answer, even unconsciously.

The third part of the RCM is optional; it derives inferences for causal effects from the observed data by conceptualizing the problem as one of imputing the missing potential outcomes. That is, once outcome data are available (that is, observations of the potential outcomes corresponding to the treatments actually received by the various units), then the modelling of the outcome data given the covariates should be structured to derive predictions of those potential outcomes that would have been observed if the treatment assignments had been different. This modelling will generate stochastic predictions (that is, imputations) for all missing potential outcomes in the study, which, when combined with the actually observed potential outcomes, will allow the calculation of any causal-effect estimand. Because the imputations of the missing potential outcomes are stochastic, repeating the process results in different values for the causal-effect estimand. This variation across the multiple imputations (Rubin 1987, 2004a) generates interval estimates and tests for the causal estimands. Typically, in practice this third part is implemented using simulation-based methods, such as Markov chain Monte Carlo computation applied to Bayesian models.

The conceptual clarity in the first two steps of the RCM often allows previously difficult causal inference situations to be easily formulated and handled. The optional third part often extends this success by relying on modern computational power to handle analytically intractable problems. With this overview in place, we consider features of the RCM in more detail.

Potential Outcomes and Causal Effects

For defining causal effects, there are three basic primitives – concepts that are fundamental and on which we must build: units, treatments and potential outcomes. A unit is a physical object, for example a person, at a particular point in time. A treatment is an action that can be applied or withheld from a unit. We focus on the case of two treatments, although the extension to more than two treatments is simple in principle although not necessarily so with real data. Associated with each unit are two potential outcomes: the value of an outcome variable Y at a future point in time if the active treatment is applied, and the value of Y at the *same* future point in time if instead the control treatment is applied. The objective is to learn about the causal effect of the application of the active treatment relative to the control on Y, where, by definition, the causal effect is a comparison of the two potential outcomes. For example, the unit could be a person 'now' without a job, the active treatment could be participating in a job training programme, and the control could be not participating. The outcome Y could be the total income over the next three years, with the two potential outcomes being the total income with and without job training; the causal effect of being trained versus not being trained is the comparison of the person's three-year total income with and without the training.

Notationally, let W indicate which treatment the unit receives: $W = 1$ the active treatment, and $W = 0$ the control treatment. Also let $Y(1)$ be the value of the potential outcome if the unit received the active version, and $Y(0)$ the value if the unit received the control version. The causal effect of the active treatment relative to the control is the comparison of $Y(1)$ and $Y(0)$ – typically the difference, $Y(1) - Y(0)$, or perhaps the difference in logs, $\log[Y(1)] - \log[Y(0)]$, or some other comparison, possibly the ratio. We can observe only one or the other of $Y(1)$ and $Y(0)$ as indicated by $W: Y_{\text{obs}} = WY(1) + (1 - W)Y(0)$. The 'fundamental problem facing inference for causal effects' (Rubin 1978, p. 38) is that, for any individual unit, we observe the value of the potential outcome for this unit under only one of the possible treatments, namely, the treatment actually assigned, and the potential outcome under the other treatment is missing. Thus, inference for causal effects is a missing-data problem – the 'other' value is missing, so the nature of causal inference is that at least 50 per cent of the values of the potential outcomes are missing. Covariates have values that are unaffected by the treatments, such as age or sex of the unit in the job training example, and are denoted by X. Even when X represents a lagged Y, such as total income last year, $Y(1) - X$ is not the causal effect of training unless $Y(0) = X$, but rather a change of income across time.

To clarify the RCM set-up with potential outcomes, consider a specific difficult case: what is the causal effect of race on hiring practices? To consider this explicitly causal question in the RCM, we must consider the manipulations that define the active and control treatments. Literally changing one's race is presumably impossible given current medical technology, but one can conceptualize experiments that can plausibly capture what researchers want to know, that is, how employers react to race when all else is constant. For example, suppose that résumés are submitted by mail to groups of employers, where the treatment to be applied to each résumé (that is, each unit) is the name attached to it (see, for example, Bertrand and Mullainathan 2004). Here, the active treatment is the use of a distinctive African-American name on the résumé, and the control treatment is the use of a traditional name. In this case, the explication of what is meant by 'the causal effect race' is through the description of the manipulations, and the causal effect to be estimated is thereby well-defined: the causal effect of having a résumé with an African-American name compared with a traditional name on the resultant hiring outcome. Whether that effect corresponds to what the investigator wants to estimate or to what others believe is relevant to policy is another issue, but the causal nature of the comparison is clear. If it is not the desired quantity estimand or is deemed not relevant, then other more appropriate manipulations should be described.

Suppose, now that there are N units rather than only one. To make the representation with only two potential outcomes for each unit adequate, must accept an assumption, the stable unit treatment value assumption (SUTVA; Rubin 1980), which rules out interference between units (Cox 1958) and rules out different versions of the treatments for the units (for example, no 'technical errors'; Neyman 1935; Rubin 1990b). SUTVA can be weakened, but still some such assumption regarding the full set of potential outcomes is required. Often, in practice, SUTVA is made more plausible by aggregating the units. For example, training some of the unemployed in a local labour market may affect job opportunities for others in that local market. Therefore changing the unit of analysis to be the local labour market in a study with many geographically separated local labour markets may make it more plausible that there is no effect of the exposure of one unit to the treatment on other units.

Under SUTVA, all causal estimands (quantities to be estimated) can be defined from the matrix of values with ith row: $(X_i, Y_i(0), Y_i(1))$, $i = 1, \ldots, N$. A causal estimand involves a comparison of $Y_i(0)$ and $Y_i(1)$ on all N units, or on a common subset of units; for example, the average causal effect across all units that are female as indicated by their X_i, or the median $Yi(1)$ minus the median $Yi(0)$ for the set of units with X_i indicating male and $Y_i(0)$ indicating no income. By definition, all relevant scientific information that is recorded is encoded in this matrix, and so the labelling of its rows is a random permutation of $1, \ldots, N$; that is, the N-row matrix $\{X, Y(0), Y(1)\}$ is row exchangeable. For convenience, we refer to this array of values as the 'science', functions of which we wish to estimate.

Brief History of Potential Outcomes to Define Causal Effects

The basic idea that causal effects are the comparisons of potential outcomes on a common set of units seems so direct that it must have ancient roots, and we can find elements of this definition of causal effects among both philosophers (for example, Mill 1843, p. 327) and experimenters (for example, Fisher 1918, p. 214). But apparently there was no formal notation for potential outcomes until Neyman (1923), which appears to have been the first place where a mathematical analysis is written for a randomized experiment. This notation became standard for work in randomized experiments with randomization-based inference, and was a major advance. Independently and nearly simultaneously, Fisher (1925) recommended physically randomizing treatments to units in experiments, as well as a different, but compatible, method of randomization-based inference, although Fisher apparently never used the potential outcomes notation. But despite the almost immediate acceptance in the late 1920s of Fisher's proposal for randomized experiments, and of Neyman's notation for potential outcomes in randomized experiments, and of both men's proposals for randomization-based inference, this potential outcome notation was not used for causal inference more generally for a half century thereafter, apparently not until introduced by Rubin (1974). As a result, the insights into causal inference that accompanied the use of the potential outcomes notation were entirely limited to the relatively simple setting of randomization-based inference in randomized experiments.

The approach used in nonrandomized settings, during the half-century following the introduction of Neyman's seminal notation for randomized experiments, was based on mathematical models (for example, regression models) relating the observed value of the outcome variable $Y_{\text{obs},i}$ to X_i and W_i, and then defining causal effects as parameters (for example, regression coefficients) of these models. This was the standard approach in medical and social science, including economics, and led to substantial confusion – the role of randomization cannot even be directly stated mathematically using the observed outcome notation. Of course, there were seeds of this first part of the RCM in social science before 1974, in particular in economics, in Tinbergen (1930), Haavelmo (1944) and Hurwicz (1962), but we can find no previous use of explicit notation like Neyman's to define causal effects. The use of the idea of potential outcomes certainly did appear in discussions in economic theory, for example, in the context of supply and demand functions (for example, Haavelmo 1944) or the Roy (1951) model, but these discussions did not lead to inference in terms of potential outcomes. Instead, inference took place in terms of the specification of simultaneous equations using observed quantities and distributional properties of error terms (for example, Heckman and Robb 1984, in the context of program evaluation models).

Nevertheless, the potential outcome part of the RCM framework for defining causal effects, namely, a generalization of Neyman's notation to allow non-randomized data, seems to have been basically accepted and adopted by most researchers by the end of the 20th century; compare, for example, Imbens and Angrist (1994) and Heckman et al. (1998) with the earlier formulation in Heckman and Robb (1984). An article exploring whether the full potential outcomes framework can be avoided when conducting causal inference is Dawid (2000), which included discussion by others that was largely supportive of the propriety of potential outcomes for causal inference.

The Assignment Mechanism and Assignment-Based Causal Inference

The second part of the RCM framework is the specification of an 'assignment mechanism': a probabilistic model for how some units received the active treatment and how other units received the control – how we conceptualize the design for how some potential outcomes were revealed and others remained hidden (that is, missing). The assignment mechanism is fundamental to causal inference. It specifies the conditional probability of each vector of assignments $W = (W_1, \ldots, W_N)^T$ given the matrix of all covariates and potential outcomes:

$$\Pr\left(W \middle| X, Y(0), Y(1)\right). \tag{1}$$

It appears that Rubin (1975) was the first place that expressed the possible dependence of the assignment vector on the potential outcomes in this direct way, which allows the statement of what makes randomized experiments special, and more generally, generates a classification of assignment mechanisms. Again, economic theory sometimes implied a specific assignment mechanism, but this theory was never explicitly stated as in the general form of (1). For example, individuals may choose the occupation that maximizes their earnings, as in the Roy model, which would lead to $W_i = \text{argmax}_w(Y_i(W))$, or more generally individuals may optimize an objective function that involves expectations over the unknown components of the potential outcomes. Imbens and Rubin (2006) provide details of such examples.

Randomized experiments are special in that they have 'unconfounded' and 'probabilistic' assignment mechanisms. Unconfounded assignment mechanisms (Rubin 1990a) are free of dependence on either $Y(0)$ or $Y(1)$:

$$\Pr\left(W \middle| X, Y(0), Y(1)\right) = \Pr\left(W \middle| X\right). \quad (2)$$

Assignment mechanisms are 'probabilistic' (or 'probability' as in Rubin 1990a) if each unit has a positive probability of receiving either treatment:

$$0 < \Pr(W_i = 1 | X, Y(0), Y(1)) < 1. \quad (3)$$

'Strongly ignorable' assignment mechanisms (Rosenbaum and Rubin 1983a) satisfy (2) and (3), and thus have unit level probabilities, or 'propensity scores', $\Pr(W_i = 1|X_i)$, that are strictly between 0 and 1, and are free of all potential outcomes.

Ignorable assignment mechanisms (Rubin 1978), are free from dependence on missing potential outcomes but may depend on observed potential outcomes $Y_{\text{obs}} = \left\{Y_{\text{obs}\ ,\ i}\right\}$

$$\Pr\left(W \middle| X, Y(0), Y(1)\right) = \Pr(W|X, Y_{\text{obs}}). \quad (4)$$

Ignorable but confounded assignment mechanisms arise in practice, especially in sequential experiments. All strongly ignorable assignment mechanisms are unconfounded, and all unconfounded assignment mechanisms are ignorable, but not the other way. Strongly ignorable assignment mechanisms allow particularly straightforward estimation of causal effects, and are the basic template for the analysis of observational studies. More generally, observational studies have possibly confounded, non-ignorable, assignment mechanisms. A confounded assignment mechanism is one that depends on the potential outcomes, and so does not satisfy (2); a non-ignorable assignment mechanism does not even satisfy (4), and thus allows treatment assignment (or, to use common economics terminology, 'selection') to depend on unobserved values, that is, the missing potential outcomes, $Y_{\text{mis}} = \left\{Y_{\text{mis}\ ,\ i}\right\}$, $Y = \left\{Y_{\text{obs}}, Y_{\text{mis}}\right\}$.

When the assignment is strongly ignorable, it can generally be represented as a 'regular' assignment mechanism, which is proportional to the product of the propensity scores:

$$\Pr(W|X, Y(0), Y(1)) \propto \prod_{i=1}^{N} \Pr\left(W_i = 1 \middle| X_i\right) \quad (5)$$

Regular assignment mechanisms are the basic template in the RCM for the analysis of observational data, because two units with the same propensity score but different treatments are essentially randomized into the two treatment conditions. Therefore, with regular assignment mechanisms, matching on the propensity score (for example, as in Rosenbaum and Rubin 1984), or subclassifying on it (for example, as in Rosenbaum and Rubin 1985), restores the assumed underlying experimental design, and inference is straightforward based only on the assignment mechanism. These assignment-based methods of inference are due to Neyman (1923) and Fisher (1925), and they involve the calculation of large-sample confidence intervals and exact significance tests for null hypotheses, respectively; both are discussed in Rubin (1990a, 1990b, 1991). For the validity of either Fisher's or Neyman's approach, the analysis must formally be defined a priori, as part of the design. But the existence of these assignment-based modes of inference helps justify the view in the RCM that the model for the assignment mechanism is more fundamental for causal inference than the model for the science, which is not needed for randomization-based inference.

Thus, in the RCM an observational study should be designed as if its data arose from a 'broken' randomized experiment, where the unknown propensity scores must be reconstructed on the basis of the covariates X prior to ever observing any potential outcomes. In such settings, it is often quite advantageous to use estimated propensity scores (for example, as in

Rosenbaum and Rubin 1984; Rubin and Thomas 1992a, 1992b, 1996, 2000; Hirano et al. 2003). When estimated propensity scores for some units are so low that they have essentially no chance of being treated, then those units should be discarded from further consideration when estimating the treatment effect in the treated (see, for example, Peters 1941; Belson 1956; Cochran and Rubin 1973; Rubin 1973a, 1973b, 1977; Rosenbaum and Rubin 1985; Dehijia and Wahba 1999; Crump et al. 2005). The result of the design phase should be treatment and control groups with very similar distributions of observed Xs, because of either matching or subclassification. If a data-set does not permit similar X distributions to be constructed in treatment and control groups, it cannot be used to support causal inferences without extraneous assumptions justifying extrapolations. Rubin (2002) offers an example of such matching and subclassification in the context of the US tobacco litigation, and Rubin (2006a) is a book devoted to matched sampling.

A striking example of the applied success of the above approach to inference in observational studies is Dehijia and Wahba (1999), which reanalysed the classic Lalonde (1986) data on job-training experiments but using the assignment-based approach of the RCM. In contrast to the wild variety of contradictory, but highly significant, answers found by the traditional econometric methods, Dehijia and Wahba used matching on the propensity score to arrive at inferences that tracked those from the underlying randomized experiment in the overall sample and in a variety of subsamples (see also Abadie and Imbens 2006).

Posterior Predictive, or Model-Based, Causal Inference

The third part of the RCM involves an optional distribution on the N-row array of science, $\Pr(X, Y(0), Y(1))$, thereby allowing Bayesian, or model-based, inference as well as assignment-based inference. An important virtue of the RCM framework is that it distinctly separates the science – its definition in the first part (and a possible model for it in the third part) from the design of what is revealed about the science – the assignment mechanism in the second part, which can also involve some scientific insights as when it is assumed to be generated by equilibrium conditions, as in supply and demand models, or by optimizing behaviour, and so on.

Bayesian inference for causal effects directly and explicitly confronts the missing potential outcomes, Y_{mis}, by using the specification for the assignment mechanism and the specification for the underlying data to derive the posterior predictive distribution of Y_{mis}, that is, the distribution of Y_{mis} given all observed values:

$$\Pr\left(Y_{\text{mis}} \middle| X, Y_{\text{obs}}, W\right).$$

This distribution is posterior because it conditions on all observed values (X, Y_{obs}, W) and is predictive because it predicts (stochastically) the missing potential outcomes. From this distribution and all of the observed values (the observed potential outcomes, Y_{obs}; the observed assignments, W; and observed covariates, X), the posterior distribution of any causal effect can, in principle, be calculated. This conclusion is immediate if we view the posterior predictive distribution as specifying how to take a random draw of Y_{mis}. Once a value of Y_{mis} is drawn, any causal effect can be directly calculated from the drawn value of Y_{mis} and the observed values of X and Y_{obs}, for example, the median causal effect for males: $\text{med}\{Y_{\text{i}}(1) - Y_{\text{i}}(0) | X_{\text{i}} \text{ indicate males}\}$. Repeatedly drawing values of Y_{mis} and calculating the causal effect for each draw generates the posterior distribution of the desired causal effect. Thus, we can view causal inference entirely as a missing data problem, where we multiplyimpute (Rubin 1987, 2004a) the missing potential outcomes to generate a posterior distribution for the causal effects.

For example, the treated units have $Y_i(1)$ observed and $Y_i(0)$ missing. Under ignorability, the regression of $Y_i(0)$ on X_i among treated units, for which there is no direct evidence, can be shown to be the same as the regression of $Y_i(0)$ on X_i among controls, for which we have data. Thus, this third part of the RCM tells us to build a realistic model of $Y_i(0)$ given X_i among control subjects, and use it to impute the missing $Y_i(0)$ among the treated from their X_i values, while being wary of issues of extrapolation beyond the observed range of X_i control values. Analogously, build a model of $Y_i(1)$ given X_i among the treated, and use it to impute the missing $Y_i(1)$ among controls. The general structure is outlined in Rubin (1978), and is developed in detail in Imbens and Rubin (2006); a chapter-length summary appears in Rubin (2007).

Advantages of the RCM

Because of the flexibility in the RCM for (*a*) formulating causal estimands, and (*b*) positing assignment mechanisms, it can handle difficult cases in principled ways. With observational studies, estimated propensity scores play a key role, because the initial analysis proceeds as if the assignment mechanism were unconfounded. To assess the consequences of this assumption, sensitivity analyses can be conducted under various hypothetical situations, typically with fully missing covariates, U, such that treatment assignment is unconfounded given U but not given the observed data. Assumed relationships (given X) between U and W, and between U, $Y(0)$ and $Y(1)$, are then varied, for example, as in Rosenbaum and Rubin (1983b), utilizing the third part of the RCM. Ideally, this speculation occurs at the design stage. Extreme versions of sensitivity analyses lead to large-sample bounds (for example, Manski 2003).

A complication, common when the units are people, is non-compliance with assigned treatment. Early work related to this issue can be found in economics using the terminology of instrumental variables, and the bridge from this terminology to the basic RCM is developed in Imbens and Angrist (1994) and in Angrist et al. (1996), and the connection to the full RCM approach is presented in Imbens and Rubin (1997) and in Hirano et al. (2000). Another complication is censoring due to death, where units may 'die' before the final outcome can be measured. This problem is formulated from the RCM perspective in Rubin (2006b), with bounds given in Zhang and Rubin (2003); see Zhang et al. (2007) for application to the evaluation of job-training programmes. This topic is also related to 'direct' and 'indirect' causal effects (Rubin 2004b, 2005). Combinations of such complications are considered in Barnard et al. (2003) in the context of a school choice example, as well as in Mealli and Rubin (2002, 2003), Jin and Rubin (2007) and Frangakis and Rubin (1999, 2001) in other contexts. The above examples can all be viewed as special cases of 'principal stratification' (Frangakis and Rubin 2002).

The references in the preceding paragraph are clearly idiosyncratic in the sense of their being specific applications of the RCM in which the authors of this article have been participants, and are not representative, but we hope they provide indications of the breadth of recent applications of the RCM.

See Also

- ▶ Bayesian Econometrics
- ▶ Bayesian Statistics
- ▶ Econometrics
- ▶ Matching
- ▶ Matching Estimators
- ▶ Treatment Effect

R

Bibliography

Abadie, A., and G.W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74: 235–267.

Angrist, J.D., Imbens, G.W. and Rubin, D.B. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–72 (an applications invited discussion article with discussion and rejoinder).

Barnard, J., J. Hill, C. Frangakis, and D. Rubin. 2003. A principal stratification approach to broken randomized experiments: a case study of vouchers in New York City. *Journal of the American Statistical Association* 98: 299–323 (with discussion and rejoinder).

Belson, W.A. 1956. A technique for studying the effect of a television broadcast. *Applied Statistics* 5: 195–202.

Bertrand, M., and S. Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94: 991–1013.

Cochran, W.G., and D.B. Rubin. 1973. Controlling bias in observational studies: A review. *Sankhya* 35: 417–446.

Cox, D.R. 1958. *The planning of experiments*. New York: Wiley.

Crump, R., Hotz, J., Imbens, G. and Mitnik, O. 2005. Moving the goalposts: addressing limited overlap in estimation of average treatment effects by changing the estimand. Unpublished manuscript, Department of Economics, University of California, Berkeley.

Dawid, A.P. 2000. Causal inference without counterfactuals. *Journal of the American Statistical Association* 95: 407–424 (with discussion).

Dehijia, R., and S. Wahba. 1999. Causal effects in non-experimental studies: reevaluating the evaluations of training programs. *Journal of the American Statistical Association* 94: 1053–1062.

Fisher, R.A. 1918. The causes of human variability. *Eugenics Review* 10: 213–220.

Fisher, R.A. 1925. *Statistical methods for research workers*. 1st ed. Edinburgh: Oliver and Boyd.

Frangakis, C., and D.B. Rubin. 1999. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 86: 366–379.

Frangakis, C.E., and D.B. Rubin. 2001. Addressing an idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics* 57: 333–342 (with discussion and rejoinder, 343–53).

Frangakis, C.E., and D.B. Rubin. 2002. Principal stratification in causal inference. *Biometrics* 58: 21–29.

Haavelmo, T. 1944. The probability approach in econometrics. *Econometrica* 15: 413–419.

Heckman, J., and R. Robb. 1984. Alternative methods for evaluating the impact of interventions. In *Longitudinal analysis of labor market data*, ed. J. Heckman and B. Singer. Cambridge: Cambridge University Press.

Heckman, J., H. Ichimura, and P. Todd. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* 65: 261–294.

Hirano, K., G. Imbens, D.B. Rubin, and X. Zhou. 2000. Estimating the effect of an influenza vaccine in an encouragement design. *Biostatistics* 1: 69–88.

Hirano, K., G. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189.

Holland, P.W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81: 945–970.

Hurwicz, L. 1962. On the structural form of interdependent systems. In *Logic, methodology, and philosophy of science, proceedings of the 1960 international congress*, ed. E. Nagel, P. Suppes, and A. Tarski. Stanford, CA: Stanford University Press.

Imbens, G.W., and J. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–476.

Imbens, G.W., and D.B. Rubin. 1997. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* 25: 305–327.

Imbens, G.W., and D.B. Rubin. 2006. *Causal inference in statistics and the medical and social sciences*. Cambridge: Cambridge University Press.

Jin, H., and D.B. Rubin. 2007. Principal stratification for causal inference with extended partial compliance: application to Efron–Feldman data. *Journal of the American Statistical Association* 103(481): 101–111.

Lalonde, R. 1986. Evaluating the econometric evaluations of training programs. *American Economic Review* 76: 604–620.

Manski, C.F. 2003. *Partial identification of probability distributions*. New York: Springer-Verlag.

Mealli, F., and D.B. Rubin. 2002. Assumptions when analyzing randomized experiments with noncompliance and missing outcomes. *Health Services Outcome Research Methodology* 3: 225–232.

Mealli, F., and D.B. Rubin. 2003. Assumptions allowing the estimation of direct causal effects: discussion of ‘Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status’ by Adams et al. *Journal of Econometrics* 112: 79–87.

Mill, J.S. 1843. A system of logic. In *Collected works of John Stuart Mill*, ed. J.M. Robson, Vol. 7. Toronto: University of Toronto Press 1973.

Neyman, J. 1923. On the application of probability theory to agricultural experiments: Essay on principles, section 9. *Statistical Science* 5(1990): 465–480 . Translated

Neyman, J. 1935. Statistical problems in agricultural experimentation. *Journal of the Royal Statistical Society* B2: 107–108 (Supplement) (with discussion). (With cooperation of K. Kwaskiewicz and St. Kolodziejczyk.).

Peters, C.C. 1941. A method of matching groups for experiments with no loss of population. *Journal of Educational Research* 34: 606–612.

Rosenbaum, P.R., and D.B. Rubin. 1983a. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

Rosenbaum, P.R., and D.B. Rubin. 1983b. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, B* 45: 212–218.

Rosenbaum, P.R., and D.B. Rubin. 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79: 516–524.

Rosenbaum, P.R., and D.B. Rubin. 1985. Constructing a control group using multivariate matched sampling incorporating the propensity score. *American Statistician* 39: 33–38.

Roy, A.D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146.

Rubin, D.B. 1973a. Matching to remove bias in observational studies. *Biometrics* 29: 159–183 .Correction note: 1974. *Biometrics* 30, 728
Rubin, D.B. 1973b. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* 29: 185–203.
Rubin, D.B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.
Rubin, D.B. 1975. Bayesian inference for causality: The importance of randomization. *Proceedings of the social statistics section of the American Statistical Association*, 233–239.
Rubin, D.B. 1976. Inference and missing data. *Biometrika* 63: 581–592.
Rubin, D.B. 1977. Assignment of treatment group on the basis of a covariate. *Journal of Educational Statistics* 2: 1–26.
Rubin, D.B. 1978. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* 6: 34–58.
Rubin, D.B. 1979. Discussion of 'Conditional independence in statistical theory' by A.P. Dawid. *Journal of the Royal Statistical Society Series B* 41: 27–28.
Rubin, D.B. 1980. Discussion of 'Randomization analysis of experimental data in the Fisher randomization test' by Basu. *Journal of the American Statistical Association* 75: 591–593.
Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. New York: Wiley.
Rubin, D.B. 1990a. Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference* 25: 279–292.
Rubin, D.B. 1990b. Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* 5: 472–480.
Rubin, D.B. 1991. Practical implications of modes of statistical inference for causal effects. *Biometrics* 47: 1213–1234.
Rubin, D.B. 2000. The utility of counterfactuals for causal inference. Comment on A.P. Dawid, 'Causal inference without counterfactuals'. *Journal of the American Statistical Association* 95: 435–438.
Rubin, D.B. 2002. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology* 2: 169–188.
Rubin, D.B. 2004a. *Multiple imputation for nonresponse in surveys*. New York: Wiley Reprinted with new appendices as a Wiley Classic.
Rubin, D.B. 2004b. Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* 31: 161–170 (with discussion and reply, 196–8).
Rubin, D.B. 2005. Causal inference using potential outcomes: Design, modeling, decisions, 2004 Fisher Lecture. *Journal of the American Statistical Association* 100: 322–331.
Rubin, D.B. 2006a. *Matched sampling for causal effects*. Cambridge: Cambridge University Press.
Rubin, D.B. 2006b. Causal inference through potential outcomes and principal stratification: applications to studies with 'censoring' due to death. *Statistical Science* 21: 299–321.
Rubin, D.B. 2007. Statistical inference for causal effects, with emphasis on applications in epidemiology and medical statistics. In *Handbook of statistics: Epidemiology and medical statistics*, ed. C.R. Rao, J.P. Miller, and D.C. Rao. Amsterdam: North-Holland.
Rubin, D.B., and N. Thomas. 1992a. Affinely invariant matching methods with ellipsoidal distributions. *Annals of Statistics* 20: 1079–1093.
Rubin, D.B., and N. Thomas. 1992b. Characterizing the effect of matching using linear propensity score methods with normal covariates. *Biometrika* 79: 797–809.
Rubin, D.B., and N. Thomas. 1996. Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52: 249–264.
Rubin, D.B., and N. Thomas. 2000. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 95: 573–585.
Tinbergen, J. 1930. Determination and interpretation of supply curves: an example. *Zeitschrift fur Nationalokonomie*. Reprinted in *The Foundations of Econometric Analysis*, ed. D.F. Hendry and M.S. Morgan. Cambridge: Cambridge University Press, 1997.
Zhang, J., and D.B. Rubin. 2003. Estimation of causal effects via principal stratification when some outcomes are truncated by 'death'. *Journal of Educational and Behavioral Statistics* 28: 353–368.
Zhang, J., D. Rubin, and F. Mealli. 2007. Evaluating the effects of job training programs on wages through principal stratification. *Advances in Econometrics* 21.

Rueff, Jacques (1896–1978)

Roger Dehem

R

Born in Paris, Rueff graduated from Ecole Polytechnique (1921), where he had been a pupil of Clément Colson. He lectured at the Institut de Statistique (1923–31) and held a chair at the Ecole libre des sciences politiques (1930–50). Rueff owes his reputation to his exceptional career in public administration and his persuasive talent. In Poincaré's monetary reform (1926–8), he was called to determine the new value of the franc. In 1930 he was posted to London, as financial attaché at the French Embassy. In 1934 he entered the Ministry of Finance, where as Director of Treasury (1936–9) he had to cope with the acute financial difficulties of the governments of the time. As Vice-Governor of the Bank of France (1939–40), he was in charge of exchange controls. He headed the Inter-Allied Agency for Reparations (1946–52). From 1952 to 1962 he was a Magistrate first at the Court of Justice of the ECSC and from 1958 on, at the Court of Justice of the European Communities. In 1958 Rueff played a leading role in the monetary reform that led to the convertibility and the stabilization of the franc. This was followed by his masterly contribution to the Armand–Rueff report on *The Obstacles to Economic Expansion* (1960).

As he argued in his first essay, *From the Physical to the Social Sciences* (1922), Rueff believed that the methodological principles of the natural sciences should also apply to the human sciences. This explains his imperturbable faith in the process of economic equilibrium. In the reparations controversy of the 1920s, Rueff saw the core of the problem in the budgetary difficulty of levying the reparations. In opposition to Keynes, Rueff (1929) held that the trade balance would adjust quickly and adequately. The persistent payments imbalances in the post-World War II period were also seen as the consequence of the reluctance to reduce internal demand in the deficit countries.

In the 1960s, Rueff became a vocal detractor of the gold exchange standard. As the adjustment mechanism in such a system can be seen as biased to the advantage of the keycurrency country, Rueff's thesis obtained the official backing of President de Gaulle. The restoration of a symmetrical gold standard thus became the French alternative to proposals to extend the IMF prerogatives.

Besides his strong Ricardian-like monetary beliefs, Rueff had wider human concerns. Deeply impressed by the political consequences of the financial disorders and the gross interferences in the price mechanism since the Twenties, he heralded the dangers ahead in *L'ordre social* (1945), the French counterpart of Hayek's *Road to Serfdom*. Rueff's distinction between true and false claims (*vrais et faux droits*), that is, between those backed by real assets and those that are not, gives the clue to his distinction between civilizations based on the rule of law within a free market system, and those requiring compulsion to settle disorders resulting from defective markets. Government distribution and monetization of false claims, by leading to inflation, macro-imbalances and controls, are seen as the main threat to individual freedom.

Selected Works

1922. *Des sciences physiques aux sciences morales. Introduction á l'étude de la morale et de l'économie politique rationelles*. Paris: Alcan. Trans. by H. Green as From the *Physical to the social sciences. Introduction to a study of economic and ethical theory.* Baltimore/London: Johns Hopkins Press/H. Milford and Oxford University Press, 1929.

1925. *Sur une théorie de l'inflation*. Nancy/Paris: Berger-Levrault.

1927. *Théorie des phénomènes monétaires; statique*. Paris: Payot.

1929. *Une erreur économique: l'organisation des transferts*. Doin.

1935. *La crise du capitalisme*. Paris: Editions de la Revue Bleue.

1945. *L'ordre social*. Paris: Recueil Sirey.

1949. *Epître aux dirigistes*. Paris: Gallimard.

1953. *La régulation monétaire et le problème institutionnel de la monnaie*. Paris: Recueil Sirey.

1961. *Discours sur le crédit*. Paris: Editions du Collège Libre des Sciences Sociales et Économiques.

1964. *The age of inflation*. Chicago: H. Regnery.

1967. *Balance of payments: Proposals for the most pressing world economic problem of our time*. New York: Macmillan.

1971. *Le péché monétaire de l'Occident*. Paris: Plon.

1972. *Combats pour l'ordre financier; Mémoires et documents pour servir à l'histoire du dernier demisiècle*. Paris: Plon.

1977. *Les fondements philosophiques des systèmes économiques. Textes de Jacques Rueff et essais rédigés en son honneur*. Paris: Payot.

Ruggles, Richard (1916–2001)

Edward N. Wolff

Keywords

Household surveys; Kuznets, S; Life-cycle saving model; National economic accounting; National income accounting; Ruggles, R; Saving and investment; Stone, J. R. N.

JEL Classifications

B31

Richard Ruggles and his wife, Nancy Ruggles, with whom he co-authored almost all of his work, did pioneering work in the field of national economic accounting. Mr Ruggles attended Harvard for both undergraduate and graduate study, earning his BA in 1939, an MA in 1941 and his Ph.D. in 1942. After earning his doctorate, Richard Ruggles joined the Office of Strategic Services as an economist during the Second World War. He worked for the office in London, where he estimated the production rates of tanks at German factories using photographs of the serial numbers from captured or destroyed tanks. In 1945–6 he was with the US Strategic Bombing Survey in Tokyo and Washington. Mr Ruggles returned briefly to Harvard as an instructor in 1946 before joining the Yale faculty a year later as an assistant professor of economics. He was named an associate professor in 1949 and a full professor in 1954. He was appointed the Stanley Resor Professor of Economics in 1954. He chaired the department of economics from 1959 to 1962. He also conducted research for numerous government agencies and bodies, including the United Nations, the Organization of American States, the Federal Reserve Board, the Bureau of the Census and the National Bureau of Economic Research, as well as the Ford Foundation.

Three principal themes emerge in the work of the Ruggles. The first is the reconciliation of macrodata with microdata. National accounts were developed during the 1930s and 1940s by Simon Kuznets, Richard Stone and Richard Ruggles, among others. The 1950s saw the development of microdata such as the Current Population Survey (CPS) in the United States. In principle, the data contained in microdata sources such as income should be consistent with the corresponding entries in the national accounts. In practice, however, this was seldom the case. Several requirements are put forward by Nancy and Richard Ruggles to fully integrate the two sources. First, the definition of sectors should be the same. For example, while household microdata include only households, the macro 'household' accounts often include non-profits and group quarter residents, such as those living in college dormitories, nursing homes and prisons. Second, definitions and imputations, such as the treatment of pensions and imputed rent, should be consistent between the two sources. Third, alignment of macro and microdata should not rely exclusively on macro totals. For example, in national accounts personal interest is computed as a residual whereas in microdata the household provides a direct estimate.

The second theme is the synthesis of microdata from several sources. Since household surveys can ask only a limited number of questions, different surveys concentrate on different characteristics. The CPS focuses on demographics and income, while the Consumer Expenditure Survey is very strong on expenditures and the Survey of Consumer Finances concentrates on assets and liabilities. Another problem is that different microdata focus on different parts of the income distribution. The CPS focuses mainly on the middle classes but its income data are weak for the lower and upper tails while the Internal Revenue Service Tax Model, a sample of tax returns, contains detailed income data on the upper tail but limited information on the bottom tail since these families do not file tax returns.

The solution proposed by the Ruggles is a statistical match of microdata. The idea is to merge microdata files which are complementary in terms of the variables they contain or the parts of the distribution that they cover. One such successful match described in their work was between the 1970 Census of Population and the 1969 Tax Model.

A third theme is the importance of institutional sectoring for the analysis of economic behaviour. In several papers, the Ruggles focus on the measurement of savings. Though most theories of savings, such as the life-cycle model, implicitly assume that all savings is done by households, Nancy and Richard Ruggles argue that savings is done by different institutions. In their accounting scheme, they develop separate current and capital accounts for the household, enterprise, and government sectors. They find that the household and the enterprise sectors are each self-financing. On net, the household sector channels almost no financial savings to the enterprise sector, and almost all investment done by enterprises is financed through enterprise savings. These results have wide-ranging implications for theories of savings and investment.

Selected Works

1947. *An introduction to national income and income analysis*. New York: McGraw-Hill.

1993. Accounting for saving and capital formation in the United States, 1947–1991. *Journal of Economic Perspectives* 7(2): 3–17.

1970. (With N.D. Ruggles.) *The design of economic accounts*. New York: NBER.

1975. (With N.D. Ruggles.) The role of microdata in the national economic accounts. *Review of Income and Wealth* 21: 203–216.

1986. (With N.D. Ruggles.) The integration of macro- and micro data for the household sector. *Review of Income and Wealth* 32: 245–276.

1992. (With N.D. Ruggles.) Household and enterprise saving and capital formation in the United States, 1947–1991: Market transactions view. *Review of Income and Wealth* 38: 119–164.

1977. (With N.D. Ruggles and E.N. Wolff.) Merging microdata: Rationale, practice and testing. *Annals of Economic and Social Measurement* 6: 407–428.

Run on Northern Rock

C. A. E. Goodhart

Abstract

On September 14, 2007, Northern Rock, a medium-sized bank specialising in residential mortgages, suffered the first substantial run of retail depositors in the UK since the 19th century. It had previously adopted a policy of fast growth, largely financing itself by borrowing in wholesale markets and by securitisation. When the financial crisis struck in August, the wholesale funding markets closed to it and it could not get a further securitisation financed; so it became massively illiquid. The Financial Services Authority (FSA) had been focussing primarily on NR's compliance with the Basel II capital adequacy requirement, and had been remiss in assessing the risks inherent in its overall business plan. After an unsuccessful review of alternative rescue policies, the authorities felt that a massive emergency loan from the Bank of England was the least bad alternative; closure would have probably led to contagion to other, similarly placed, banks. Unfortunately the news of the emergency lending leaked prematurely, and its interpretation in the media helped to trigger the run. The run was not stopped until the following Monday when the Chancellor of the Exchequer stepped in to guarantee all NR deposits, and then to provide 100% deposit insurance to all other banks as well.

Keywords

Bank run; Emergency lending; Financial supervision; Deposit insurance; Wholesale funding; Securitisation

JEL Classifications

E58; G21; G33

Introduction

On the morning of Friday 14 September 2007, queues of depositors began to form inside, and then outside, the (relatively) few branches of Northern Rock (only nine in the London area, for example). This was the first substantial run in the UK by retail depositors since the 19th century. Northern Rock had been a building society until 1997, with a large local presence in the north-east (headquartered in Gosforth, Newcastle upon Tyne), but otherwise then not widely known and subject to relatively strict Building Society requirements. In that year it demutualized, became a bank and later embarked on a massive program of expansion, under its incoming Chief Executive, Adam Applegarth.

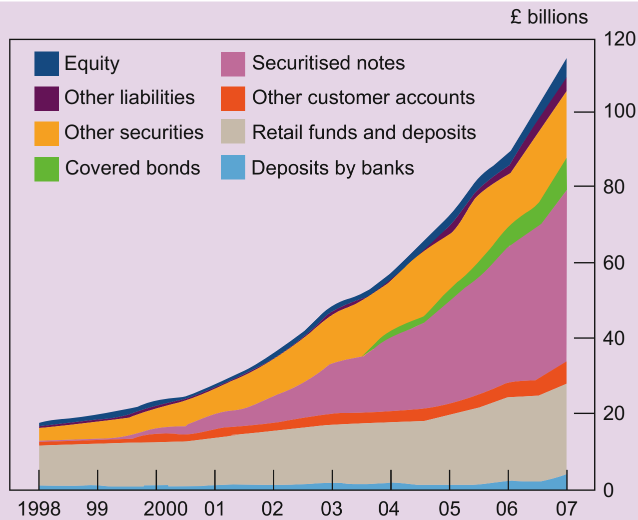
Northern Rock, however, stuck to its traditional area of expertise, financing household mortgages, after demutualization; but it improved its terms to borrowers, on relative spreads and down-payment, in order to take an ever larger share of the domestic mortgage market. The rate of expansion of its lending then vastly exceeded the rate of growth of its retail deposits, leading to a sharp rise in its loan to deposit ratio. It financed this gap in three main ways. First, having originated such loans, it pooled them together and transferred these to a securitization vehicle, a special-purpose entity termed Granite. Second, it issued covered bonds, secured against both the mortgages and the bank. Finally, it borrowed in wholesale markets (see Fig. 1).

'Mr Applegarth outlined the overall funding of Northern Rock:

> 50% was securitisation, which had an average life of three and a half years; 10% was covered bonds, which had an average life of about seven years; and of our wholesale borrowings, which is 25%, half of that had a duration longer than one year and the other half was less than one year's duration.' House of Commons Treasury Committee Report on 'The Run on the Rock', p. 13.

So Northern Rock was largely reliant on continued access to (short-term) wholesale markets for funding, especially during the intervals in which it was originating and warehousing mortgages before transferring them in a pool as an additional tranche to its associated SPV, Granite. It was due to launch an exceptionally big securitization in August/September, and therefore was unusually heavily reliant on short-term money-market funding when the crisis broke. When the wholesale funding markets, especially but not only in Europe, began to break down on 9 August 2007, Northern Rock soon realized that it was in trouble, thus:

Run on Northern Rock, Fig. 1 Northern Rock: balance sheet growth and liability structure – June 1998–June 2007



Sources: Northern Rock Interim Annual Reports; Bank of England *Financial Stability Report*, October, 2007, p. 10.

> 'The then Chairman and Chief Executive of Northern Rock first discussed these problems with each other on Friday 10 August'. (Treasury Select Committee, *ibid.*, p. 35.)

For a number of reasons, detailed in the next section, none of the proposed remedies for Northern Rock's illiquidity proved feasible. One factor in August was the uncertainty over how long the securitization markets would be closed. It was hoped (not just by Northern Rock) that this would be temporary, and Northern Rock and its advisers were still hoping to launch an issue in early September. Its illiquidity worsened rapidly from chronic to critical, especially since the transfer of a further tranche of mortgages to Granite could not possibly succeed in those circumstances. So the stark alternatives soon became to allow Northern Rock to go bankrupt or to provide it with a massive loan from the Bank of England.

For a variety of reasons, detailed in the next section, it was felt that the existence, occasion and extent of such a loan would have to be publicly revealed. The loan itself was finally settled on the evening of Thursday 13 September. At a conference dinner at the Bank that evening, senior Bank officials kept on mysteriously getting up and leaving and then returning; so it was obvious something was occurring, but none of the others present knew what.

The idea was to announce the loan early on Monday 17 September. The announcement of the loan could then have had two effects on the retail depositors (and general public). The first could have been, 'The Bank of England is now in support; we can stop worrying', but the second was 'We did not realize how bad the situation was. It must be bad if the Bank has to help so much'.

The premise on which the deposit insurance scheme in the UK had been established was that a retail deposit run was unthinkable in the UK. So the insurance was aimed to limit moral hazard, by being a full 100% only up to d2000, partial from there on (90%) up to a cap of d35,000, and none thereafter. The aim of the scheme was to make it (politically) easier to allow bank failures (by bailing out 'widows and orphans'), *not* to stop runs. So, naturally, in so far as retail depositors realized that they had any insurance at all (doubtful), it was often still worth their while to run, so long as there was any significant probability of Northern Rock failing. Moreover the time lag before *any* such insurance payment might be received could be weeks, if not months.

If the announcement of Bank of England support for Northern Rock had been made as planned on Monday morning, it would have been accompanied by a battery of anodyne Press comment, from the Bank and Treasury, about 'no need to worry', 'everything in hand', etc. But, even before the final terms of the loan were settled on the Thursday evening, Robert Peston of the BBC reported at 8:30 p.m. that Northern Rock had asked for and would receive emergency financial support from the Bank of England. Naturally, given the juicy nature of this leak, it was not in the interest of the media to play this down, either immediately or subsequently in the next few days. We shall probably never know how and why this leak occurred, though there are a range of (unsubstantiated) rumours about this.

Be that as it may, nothing had been made ready for the announcement. Not only were Press departments unprepared, but Northern Rock was unable to draft in additional bank tellers or expand its website to meet the expected flood of withdrawals and urgent enquiries by worried depositors. So the website crashed, and Internet depositors may have feared that such a shutdown was deliberate, to block withdrawals, and many then went to do so at the branches. The limited branch staff, in their often small branch offices, could not process withdrawals quickly enough to prevent queues snaking out of the door and around the block. These were, of course, captured on television, which served to feed the run.

In the face of this unforeseen event, the authorities seemed paralyzed on Friday and over the weekend, with no effective response. It was not until Monday that the Chancellor of the Exchequer, Alistair Darling, appointed to the position as recently as late June, announced that the government would provide a full guarantee for all Northern Rock deposits. This soon led to queries as to whether the government could discriminate in this way on behalf of just one bank, and it soon became accepted that the government would soon have to legislate, *de jure*, to provide 100% deposit insurance (at least up to some upper limit) for all bank deposits, and that it was already in that position, *de facto*. These measures did stop the run, and prevented immediate contagion, for example to Alliance and Leicester, and Bradford & Bingley.

Clearly, the Northern Rock episode forced a change in the UK's deposit insurance regime. But the effects went much further. The event underlined how limited were the options that the authorities had to hand for dealing with a failing bank, and underscored the need for a new Special Resolution Law for financial institutions, later passed in 2009. The need for such a Special Resolution Regime had been perceived beforehand, but not pursued with sufficient vigour. If such an SRR had been in place, the episode could have been handled much more smoothly. While the main blame for the collapse fell on the Board and management of Northern Rock for having embarked on such a risky business plan, questions were asked as to how the supervisor, the Financial Services Authority (FSA), had allowed it to get that way. The resulting study, by the internal audit division of the FSA, 'The supervision of Northern Rock: a lessons learned review', (March 2008), was open and candid, and exposed a litany of shortcomings. The reputation not only of the FSA, but also of its vaunted, 'Principles based' and 'Light touch' approach, was damaged. Meanwhile, the failure of the authorities, working together in the Tripartite Committee of FSA, Bank and Treasury, to come up with a better solution before the run occurred, led to queries about 'Who was in charge?', and whether the whole UK regulatory and supervisory structure needed reconsideration (as did occur later when the Conservative/Liberal coalition came into power in May 2010).

The fallout from the Northern Rock run extended yet further into the political orbit. It formed part of the background to the decision by Gordon Brown, who had become Prime Minister in June, following the resignation of Tony Blair in the aftermath of the Iraq affair, *not* to hold a general election in October 2007 to reinforce his political position.

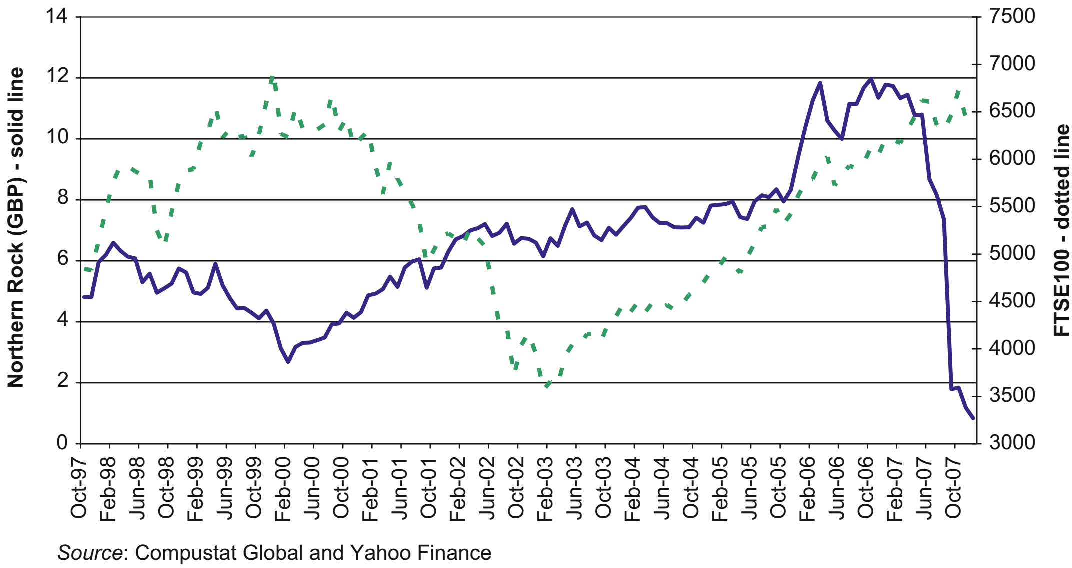
So the run on Northern Rock had important ramifications. In the next section we will discuss in more detail what led up to it, and particularly why it was not prevented. Then we cover what has happened since, and we finish with some brief conclusions.

The Antecedents of the Run

With the inestimable benefit of hindsight, there is no question that Northern Rock had embarked on an extremely risky business strategy. The Treasury Committee Report described that strategy as being 'high-risk, reckless'. So why did no one stop them? Let us consider the following potential gatekeepers: (1) the Board, (2) shareholders, and (3) the supervisor (FSA).

The reason why the Board was content with its strategy is set out in the Treasury Committee Report, pp. 15/16:

> 'Two aspects of this worldwide liquidity squeeze appeared to surprise Northern Rock, and overcome the attempts highlighted above to combat the tightening in credit markets. One was the absence of a so-called "flight to quality". Dr. Ridley told us that:
>
> What we did not expect was that there would be no flight to quality in that process [of a tightening in credit markets]. In other words, we expected that as markets became tighter and as pricing for risk changed that low-risk prime UK mortgages (and we have below half the industry average of arrears on our mortgage book) and such a low-risk book would remain easier to fund than sub-prime mortgages elsewhere. That is why we were very determined to keep the credit quality of our book high, in order to be able to attract funding.
>
> 'Mr. Applegarth told us that Northern Rock had wrongly "believed that high-quality assets and transparency [were] the way to maintain liquidity". Sir Derek Wanless [Chair of the Risk Committee of the Board] told us that Northern Rock's "first line of defence [was] good credit quality". 'Secondly, Northern Rock had not foreseen all its funding markets closing simultaneously, as happened after 9 August. Dr. Ridley explained:
>
> We deliberately diversified our funding platform so that we would have . . . three different types of funding and indeed a diversified programme within the wholesale funding, and geographically we had programmes in the United States, Europe, the Far East, Canada and Australia. That was deliberately so that if one market closed we would still have access to others. The idea that all markets would close simultaneously was unforeseen by any major authority.



Run on Northern Rock, Fig. 2 Northern Rock share price and FTSE index, 1997–2007 (monthly)

> 'The idea of all markets closing to Northern Rock was repeatedly characterised to us by Northern Rock officials as "unforeseeable".'

Thus there was a failure to realize that, in really stressed conditions, correlations go to unity and previously acceptable diversification ceases to be effective.

It would generally be an error to expect much risk control from shareholders. Given their limited liability, they often favour and reward risk-taking. Indeed, Northern Rock, with its expansionary policy and risky strategy, had been a favourite of the London Stock Exchange. It was not until relatively late in the day that its share value declined sharply (Fig. 2).

So, finally, we come to the supervisor (FSA). Here there is considerable information from the FSA's published internal audit on 'lessons learned' (March 2008). First, there was a particular and unusual degree of maladministration in the case of Northern Rock, as noted in Fig. 3 taken from page 9 of the FSA's audit.

Even more important was that the FSA was excessively focused on the implementation of the Basel II recommendations at this juncture, so much so that other aspects and dimensions of risk were largely ignored, such as excessive growth, very high leverage and poor liquidity. Largely because mortgage lending had a low Basel II risk weight, and the securitization of mortgages into Granite lowered the risk weighting even further, Northern Rock passed its Basel II test with flying colours in June 2007, at a time when, by US leveraged standards, it would have been considered 'critically undercapitalized', with a core equity to total assets ratio of under 2% (over 50 to 1!). This is a leading example of the folly of putting all one's supervisory concerns into the single Basel II basket.

As the Treasury Committee Report notes (p. 25),

> The adoption of an advanced approach requires a waiver from the Financial Services Authority. On 29 June 2007, Northern Rock was told by the FSA that its application for a Basel II waiver had been approved.
>
> Due to this approval, Northern Rock felt able to announce on 25 July 2007 an increase in its interim dividend of 30.3%. This was because the waiver and other asset realizations meant that Northern Rock had an "anticipated regulatory capital surplus over the next 3–4 years". Mr Applegarth explained how Northern Rock had achieved this waiver. The company had come to the end of a two and a half year process, during which period Northern Rock had undergone several stress tests, a matter we consider further later in this chapter. As well as

Northern Rock is in the highlighted box in each population

Supervisory Period – data taken from IRM

Firms with regulatory periods of:		
18–24 months	24	63%
25–30	10	26%
36 months	4	11%

(As at 1 August 2007, high impact firms only: MRGD-26 and WIBD-12)

Turnover of HoDs experienced by MRGD firms

Number of HoDs (1 Jan 05 – 9 Aug 07)	1 HoD	2 HoDs	3 HoDs
Number of high impact firms	18	6	2
	69%	23%	8%

C&C meetings – estimates made by staff based on various sources for high impact firms

	1 Jan – 9 Aug 2007	2006	2005	Period Total
Average for WIBD firms	13	24	18	55
Average for MRGD firms	22	29	23	74
Average for MRGD excl 5 largest banks	17	22	19	58
Average for 5 largest retail banks	43	59	41	143
Northern Rock	7*	1	0	8

* Out of which five meetings were held on one day and two were by telephone

Risk mitigation programmes (RMP) – data taken from IRM

Number of firms with RMP	37
Number of firms without RMP	1

Source: The Financial Services Authority (FSA) Internal Audit Division. The supervision of Northern Rock: a lessons learned review, March 2008, p. 9.

Run on Northern Rock, Fig. 3 The supervision of Northern Rock: lessons learned

this, in order to obtain a Basel II waiver Northern Rock had to “show that [Northern Rock could] dynamically manage scorecards from new lending all the way through to arrears and possessions and put that information back into [Northern Rock’s] front end score cards”. Mr Applegarth explained that the waiver had led to a dividend increase because:

when you get your Basel II approval, the relative risk weighting of certain assets in your balance sheet changes. So what we had, because of the quality of the loan book, was you saw our risk weighting for residential mortgages come down from 50% to 15%. That clearly required less capital behind it, so that links to why we were able to increase the dividend.

Hardly more than a month afterwards, on 9 August 2007, the financial crisis began, and wholesale markets began to dry up. A generalized shortage of liquidity ensued, with Northern Rock particularly at risk. In previous decades, at least until the 1970s, banks might have weathered this by selling, or borrowing against the collateral of, their government debt holdings. But banks, including Northern Rock, had shifted from asset liquidity (government debt) to cheaper funding liquidity from wholesale markets. They had no government debt holdings left to fall back upon. Under these circumstances,

> In August 2007, the Bank of England was approached by banks arguing that the Bank of England should provide additional liquidity, at no penalty rate. The FSA had transmitted the banks' request to the Bank of England, but refused to state to us whether it had supported the banks in requesting this additional liquidity, on the grounds that conversations between Tripartite members ought to remain private. On 12 September 2007, in advance of his oral evidence on 20 September, the Governor of the Bank of England wrote a letter to the Chairman of this Committee. In that letter, the Governor pointed out that he did not agree with the suggestions for additional measures that others believed the Bank of England should undertake: lending at longer maturities, removing the penalty rate or increasing the range of collateral against which the Bank would be prepared to lend. In the letter, he gave three reasons for his position. First, he stated that "the banking system as a whole is strong enough to withstand the impact of taking onto the balance sheet the assets of conduits and other vehicles". Second, "the private sector will gradually re-establish valuations of most asset backed securities, thus allowing liquidity in those markets to build up". Third, there would be a risk of "moral hazard". In essence, this "moral hazard" argument is that, should the central bank act, and effectively provide extra liquidity at different maturities against weaker collateral, markets would, especially if the liquidity were provided at little or no penalty, take it as a signal that the central bank would always rescue them should they take excessive risk and get into difficulties. Such a signal would lead to ever more risk taking, and the next crisis would consequently be greater than it would otherwise have been. (Treasury Committee Report, pp. 38/39.)

The Governor's position was clearly contentious. Moreover, the ECB, though for historical reasons rather than from taking a different stance on moral hazard, was prepared to lend against a wider range of assets. It may be that, had Northern Rock been within the euro zone, it could have survived this first storm. But now, with further securitization into Granite now impossible, and liquidity running out, Northern Rock had either to sell itself to another (bigger) bank, or apply for a large, formal support facility from the Bank of England.

One major high street retail bank, widely believed to be Lloyds, showed considerable interest, but the bank, having discovered the Rock's liquidity problems, wanted a loan from the Bank. 'The Chancellor of the Exchequer stated clearly that the financial support requested was in the form of a loan, which "could have been as much as d30 billion. . . to be given at commercial rates by the Bank of England"' (Treasury Committee Report, p. 51). There were two main problems with this. First, if such favourable terms were going to be offered to one bank, they would have to be offered to all banks, which would have had to be done publicly and would have taken a long time. Second, would it be consistent with the European Union prohibition on state aid to commercial banks? So such a liquidity facility was denied, and that ended any prior hope of merging Northern Rock with a stronger partner. In any case, such a merger would have had to be agreed by the shareholders, and that would have meant exposing their liquidity problems.

It is not clear whether liquidation was ever seriously considered. If Northern Rock had been allowed to fail, there would almost certainly have been immediate contagion to the other weaker mortgage lenders in the UK, notably Bradford & Bingley. In any case, 'It appears that a decision in principle that Northern Rock would be granted a support facility should neither securitisation or a takeover prove possible was taken at a meeting between the Chancellor of the Exchequer, the Chairman of the FSA and the Governor of Bank of England on Monday 3 September. The final decision was that of the Chancellor of the Exchequer, but his decision was taken on the basis of a joint recommendation of the Governor of the Bank of England and the Chairman of the FSA' (Treasury Committee Report, p. 54).

There was then much – quite confused – discussion whether such emergency lending to Northern Rock would have to be announced or could be kept covert. The problem was that this support clearly involved a material change in the financial condition of Northern Rock, and the question was whether European and/or British rules on market disclosure required that such a change be publicly announced. The Board of Northern Rock were legally advised that such an announcement was needed. So it was arranged for the early morning of Monday 17 September, but, as already described, this was overtaken by the leak to the BBC on the evening of Thursday 13 September.

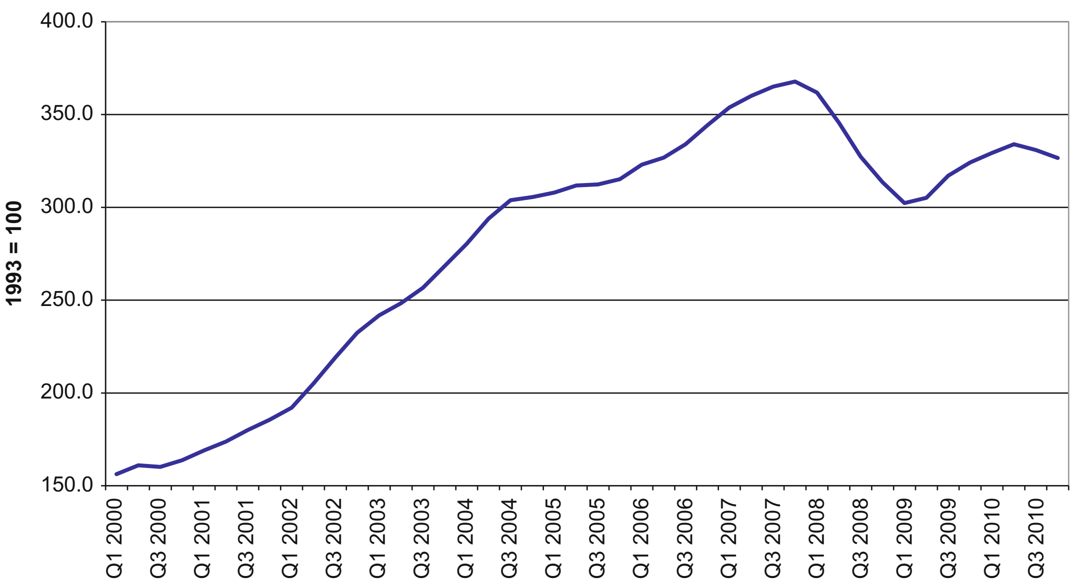
The Subsequent History

Throughout this whole episode, the management of Northern Rock, and all the regulatory authorities (the Bank of England, FSA and Treasury), were adamant in their claims that the Rock's mortgage assets were of 'high quality'. They no doubt were, in the sense that all the proper procedures and paperwork had been gone through in the correct manner, with none of the fiddles and deficiencies that had besmirched the US sub-prime market. But Northern Rock had concentrated on UK mortgages and was expanding very rapidly (averaging 20% p.a.) at the top of a huge housing boom. Moreover, as noted in the FSA's internal report on 'Lessons Learned', p. 38,

> A key element of Northern Rock's product range was the "Together" product, which represented 31% of gross mortgage lending by the firm in 2006 and 26% in the first six months of 2007. The key features of Together were that it offered a secured loan of up to 95% loan to value (LTV) coupled with an unsecured loan of up to 30% LTV all at a single rate and serviced by one monthly payment. Together was essentially targeted at first-time buyers and the unsecured element was designed to finance associated house-purchase or home-making costs. The scope for a loan of up to 125% LTV meant the product was regarded by many as particularly high-risk.

It was inevitable that in a sharp bust in the housing market (one such was clearly on its way by the latter half of 2007 – Fig. 4) a sizeable proportion of the Rock's mortgage borrowers would move into negative equity. Although the legal arrangements (mortgages were recourse, not non-recourse loans) and UK social culture meant that arrears of payment, default and foreclosure would still be far less than in the USA, the forward-looking expectation by end-2007 must have been that not only was Northern Rock illiquid, but that it would very likely also become insolvent; even though in September 2007, on a backwards-looking basis, it was solvent (with relatively few arrears comparatively – not such a difficult achievement when house prices had been rising so rapidly. Also Northern Rock was, apparently unusually quick to foreclose, which thereby kept arrears low.)

Be that as it may, the announcement by the Chancellor, Alistair Darling, on Monday 17 September that all existing deposits at Northern Rock would be guaranteed (new deposits there also became fully guaranteed on 9 October) stopped



Run on Northern Rock, Fig. 4 UK house prices, seasonally adjusted, 2000–2010 (*Source*: The Nationwide House Price Index (http://www.nationwide.co.uk/hpi/historical.htm)) the run; but this was not until some d4–5 billion more had been withdrawn, raising the necessary liquidity assistance from the Bank of England to over d25 bn. Meanwhile, the government guarantee covered an additional d30 bn. So, having rescued Northern Rock depositors at taxpayers' expense, the main concern now was to limit that bill.

Besides its mortgage assets, whose potential future value depended on an increasingly fragile housing market, Northern Rock had an existing infrastructure of branches, staff and systems that could give a new entrant access to the (oligopolistic) UK retail banking market. In pursuit of such an entry a wide group of institutions expressed some interest in acquiring Northern Rock, putative names apparently including Virgin Group, Oliphant, Cerberus, J.C. Flowers, Lloyds (again), Lehman Bros, Bradford & Bingley and Tesco. By the time the deadline for such bids occurred (4 February 2008), only two remained on the table, from Virgin and an in-house bid. After a study of the options, the Treasury, assisted by Goldman Sachs, decided that (temporary) public ownership would be cheaper for the taxpayer than either bid, so on 17 February 2008, the Chancellor announced the nationalization of Northern Rock.

In the meantime, not surprisingly, most of the prior top management and Board had resigned and been replaced. When Northern Rock was nationalized on 20 February, the question arose as to what compensation, if any, should be paid to the shareholders, an issue driven forward by a couple of hedge funds who had bet on a recovery. The authorities decreed that the appropriate amount for compensation should be the likely valuation of Northern Rock *absent any official support*, which was, of course, as an independent valuer determined, precisely zero, which was accepted in all the law courts who have so far tried the case.

Since nationalization, Northern Rock has split into two parts at the outset of 2010: the good part, Northern Rock plc, and the bad part, Northern Rock (Asset Management) plc, whose role is to manage the run-down of the worst assets (it was later merged with Bradford & Bingley plc into a single holding company, UK Asset Resolution UK). The good part, Northern Rock plc, continues, with conflicting pressures to expand its loans (at a time when other banks are not doing so) on the one hand and to reduce its book to allow the government to exit nationalized banking on the other hand. In February 2010 the government removed the 100% guarantee for Northern Rock depositors, leaving such depositors in the same position as depositors in other UK banks, with the first d50,000 fully guaranteed. Meanwhile some of the Bank of England's support loan, transferred to HMT in the summer of 2008, has been repaid, (the June 2010 half-yearly balance sheet shows some d22.5 billion of such debt to HMT still outstanding), and there are intermittent rumours of private sector interest in buying Northern Rock plc back from the government. On 18 January 2011 it was announced in the Press that the public sector owner UK Financial Investments (UKFI) was asking investment banks and other possible advisers to apply to assist with funding a private sector buyer.

Conclusions

The story of the run on Northern Rock is an unhappy one. The management of Northern Rock adopted an excessively expansionary and risky business plan, and neither the Board nor their supervisor, the FSA, checked them. This was partly due to illusion and delusion, on the part of management, that they could always fund in wholesale markets and that the UK housing market would not collapse; and on the part of the FSA that full implementation of Basel II was a sufficient guarantee that all would be well. Its downfall damaged a lot of reputations, not least that of 'principles based', 'light touch' UK regulation.

See Also

▶ Bank of England
▶ Credit Crunch Chronology: April 2007–September 2009
▶ Euro Zone Crisis 2010
▶ Foreclosure, Economics of

Bibliography

For a more general study of the circumstances and implications of the event, see Bruni, F. and Llewellyn, D. (eds.) 2009. The Failure of Northern Rock: A Multi-dimensional Case Study. SUERF – The European Money and Finance Forum, Vienna.

For further details of events since the run, this article has relied on Wikipedia, notably *Nationalisation of Northern Rock 2020* (http://en.wikipedia.org/wiki/Nationalisation_of_Northern_Rock), as modified on 2 January 2011, and other Press sources, notably *Timeline: the Northern Rock Crisis* by the *Guardian* (http://www.guardian.co.uk/business/2008/mar/26/northernrock

The FSA Internal Audit Division then examined *The Supervision of Northern Rock: A Lessons Learned Review* in their Report of March 2008.

The most comprehensive academic account of the event is Milne, A. and Wood, G. 2009. Shattered on the Rock? British financial stability from 1866 to 2007. *Journal of Financial Regulation*, 10(2), 89–127. Geoffrey Wood both gave evidence to, and was the adviser of, the Treasury Committee for this Report.

There are three main public sector reports. The best is that by The Treasury Committee of the House of Commons: *The Run on the Rock*, Fifth Report of Session 2007–8, Vol. 1, HC56–1 (26 January 2008).

These reports only covered events leading up to and surrounding the run itself. For an account of the aftermath, the National Audit Office (NAO) report on *HM Treasury: The Nationalisation of Northern Rock*, 20 March 2009 (www.nao.org.uk/publications/0809/northern.rock.aspx), should be consulted.

Ruskin, John (1819–1900)

Lawrence Goldman

Ruskin was born in London in 1819 and died at Brantwood, his house by Coniston Water in Cumberland, in 1900. In the 1840s and 1850s he rose to eminence in Victorian Britain as a critic of painting and architecture. A developing concern with the social relations of art and the influence of Thomas Carlyle led him to outspoken social criticism and from the 1860s he assumed a position as one of the most virulent opponents of 19th-century industrial capitalism. His lectures in Manchester in 1857 on *The Political Economy of Art* were followed by a series of works that castigated Victorian society in general and political economy in particular for sanctioning commercial immorality. Of these, the most important were *Unto This Last* (1862) – so controversial that its original publication as essays in the *Cornhill Magazine* in 1860 was stopped by the proprietor; a series of lectures and letters published in the mid-1860s including *Sesame and Lilies* (1865), *Ethics of the Dust* (1866), *Crown of Wild Olive* (1866) and *Time and Tide* (1867); *Munera Pulveris* (1872), which first appeared as four articles in *Fraser's Magazine* in 1862–3 and was to have formed the preface to a larger treatise on political economy that was never written; and the enigmatic and highly individual monthly letters 'to the workmen and labourers of Great Britain' issued by Ruskin between 1871–1884 as *Fors Clavigera.*

Much of Ruskin's writing on political economy originated in the attempt to refute specific economic doctrines as formulated by the discipline's more illustrious propagandists. He was not a systematic thinker but the most cogent exposition of his own economic ideas can be found in *Unto This Last* which was written to provide, as Ruskin explained in the preface, 'an accurate and stable definition of wealth' and to show that its acquisition was possible 'only under certain moral conditions of society'. These objects were at the heart of Ruskin's endeavours. He sought to redefine all the basic categories of political economy – not only wealth but value, labour and capital as well – as a prelude to the construction of harmonious social relations in an ideal, moral society to be characterized by cooperation, justice and hierarchic order rather than competition, avarice and flux. He launched an assault on the discipline of economics as an abstraction which deliberately excluded all questions of moral action, which caricatured human nature, and which spread doctrines that were inconsistent with Christianity. Political economy sanctioned speculation, the taking of interest on invested capital and the policy of laissez-faire: in practice it led to the intensification of social divisions and the dehumanization of work by divorcing mental from manual labour. To all of this Ruskin was implacably opposed.

His critique was by no means novel: Tory paternalists earlier in the century and the Christian Socialists who Ruskin met at the Working Men's College in London where he taught art from 1854 to 1860 had used similar arguments before him. But no other Victorian approached Ruskin's style of invective, his fluent fury and vehemence. Middle-class opinion ignored his 'windy hysterics' – *Unto This Last* sold only 900 copies in eleven years – and his challenge to orthodox economics came from too far outside accepted economic discourse to have troubled its practitioners. But Ruskin was an important influence on celebrated individuals including William Morris, Tolstoy, Gandhi and Proust, who translated *Sesame and Lilies* and *The Bible of Amiens* (1880–5) into French. And though personally hostile to socialism, Ruskin was read and revered throughout the emerging British labour movement at the turn of the century.

Selected Works

1857. *The political economy of art.* Later published as *A joy for ever.* London: Smith, Elder & Co., 1880. In *Works,* vol. 16, 1904.

1862. *Unto this last. Four essays on the first principles of political economy.* London: Smith, Elder & Co. In *Works,* vol. 17, 1905. Also available in a good modern edition, ed. P.M. Yarker. London: Collins, 1970.

1871–84. *Fors Clavigera. Letters to the workmen and labourers of Great Britain*. London. In *Works*, vols 27–9, 1907.

1872. *Munera Pulveris.* London. In *Works,* vol. 17, 1905.

1893. *A complete bibliography of the writings in prose and verse of John Ruskin, LL.D.* 2 vols, ed. T.J. Wise. London. Reprinted. London: Dawsons of Pall Mall, 1964.

1903–12. *The works of John Ruskin.* 39 vols, ed. E.T. Cook and A. Wedderburn. London: George Allen. (This is still the standard collection of Ruskin's works.)

References

J.A. Hobson's sympathetic biography (Hobson, 1898), which examines Ruskin's social and economic ideas in detail, is testimony to his influence over advanced liberals and socialists at the turn of the century. Other biographical studies include Leon (1949), Hunt (1982), and Hilton (1985). For modern studies of Ruskin's social and economic thought see Sherburne (1972) and Anthony (1983).

Anthony, P.D. 1983. *John Ruskin's labour. A study of Ruskin's social theory.* Cambridge: Cambridge University Press.

Hilton, T. 1985. *John Ruskin: The early years*. New Haven/London: Yale University Press.

Hobson, J.A. 1898. *John Ruskin, social reformer.* London: J. Nisbet & Co.

Hunt, J.D. 1982. *The wider sea: A life of John Ruskin*. London: Dent.

Leon, D. 1949. *Ruskin, the great Victorian*. London: Routledge & Kegan Paul.

Sherburne, J.S. 1972. *John Ruskin or the ambiguities of abundance*. Cambridge, MA: Harvard University Press.

Russian Economy

Michael Alexeev and Shlomo Weber

Although the Russian economy performed poorly in the first 7 years following the collapse of the USSR, the reforms of the 1990s laid the foundation for strong growth in 1998–2008. The early 2000s also saw successful reforms, such as the flat income tax and streamlined regulations. Serious challenges remain, however, including high dependence on natural resources, pervasive corruption, and rapid aging of the population and its distorted geographical distribution. The ability to rid itself of the remaining Soviet legacies in the economic structure, institutions and people's mindset will determine whether Russia can break its historical pattern of growth spurts followed by stagnation.

Introduction

The Russian economy has undergone a dramatic transition since the collapse of the Soviet Union in

Russian Economy, Table 1 Main indicators of Russia's economic development

Indicator/year	1992	1998	2000	2008	2011
GDP per capita, constant 2000 US$	2106	1511	1775	3044	3055
GDP per capita, PPP, constant 2005 US$	10219	7329	8613	14767	14821
GDP per capita annual growth %	−14.6	−5.0	10.0	5.4	4.3
Inflation (CPI; annual %)	1354	27.7	20.8	14.1	8.4
Unemployment (% of labour force)	5.2	13.3	10.6	6.3	7.5[b]
Budget balance (all government levels; % of GDP)	−8.0	−3.6	2.8	4.9	1.6
Oil output (million tones)	398.8	304.3	323.3	488.5	511.4
Oil rents (% of GDP)	8.32	5.78	19.00	16.80	14.23[b]
Natural gas output (billion cubic meters)	582.8	532.7	528.5	601.7	607.0
Total reserves (including gold, million current US$)	9817.7[a]	12,043.0	27,656.3	426,278.8	497,410.2
Population (thousands)	148,689	146,899	146,303	141,950	141,930

[a]1993; [b]2010

Sources: WDI (2013), except for 1992 inflation (IMF 1993), budget balance (IMF 1993; IET 2006; IEP 2012; Goskomstat 2003) and oil and gas output (BP 2012)

late 1991. While the first 7 years of transition were characterised by poor economic performance, its main economic indicators (see Table 1) between 1998 and 2008 reveal remarkably strong growth. Most notably, Russia's per capita GDP in PPP terms has increased more than twofold since 1998. In terms of total PPP GDP, Russia is now sixth in the world. It is the top producer of oil and lead exporter of natural gas, second in production of natural gas, and fourth in the amount of foreign exchange reserves and in electricity generation (CIA 2013).

While the economic achievements of the last 13 years look impressive, Russia's economy may be in danger of repeating the pattern of fast growth followed by slowdowns that has characterised its long-running attempts to catch up with the developed western economies. Since Peter the Great, each significant growth spurt has been interrupted either by a war or because the institutions necessary for sustained growth failed to take root or were not developed at all (Mau and Drobyshevskaya 2013). Currently, Russia's inadequate institutions again seem to be impeding sustained economic development. One major reason for this is the Soviet legacy, which continues to exert its influence despite the apparent significant restructuring of Russia's economy and institutions since the break-up of the USSR. Some authors even argue that the impact of the Soviet economic, social and political system on contemporary Russia outweighs the influence of Russia's culture, geography and natural endowments (Ericson 2013). While the 'physical' Soviet legacies and structure of the economy have been fading, the Soviet institutional and behavioural heritage is harder to overcome. Moreover, even though markets provide microeconomic coordination and incentives to economic agents, politicians largely control the key sectors of the economy, including energy, natural resources, the financial sector and railroads. The tension between markets and political control, combined with the institutional and behavioural legacies of the Soviet period, arguably represent the defining characteristics of the modern Russian economy and are crucial for an understanding of its functioning. In the following, we describe both the momentous changes that have taken place in Russia's economy since the collapse of the USSR and the challenges that remain.

Restructuring of the Russian Economy During Transition

The Russian economy has undergone deep restructuring since the collapse of the USSR. During the 1990s most prices were freed from state control, entrepreneurial activities were legalised, private ownership of capital assets became the norm, and services and household consumption as a share of GDP increased dramatically at the expense of heavy industry and investments. While more than 90% of fixed capital in Russia was stateowned in 1990, this share had declined to 25% by 2000 and to 20% in 2010 (Rosstat 2011; these figures do not include municipality-owned property, otherwise, the state ownership in 2000 would have been 43%: Goskomstat 2003). Between 1991 and 2010 the industry share in GDP declined from 47% to under 37%, while the share of services increased from 38% to almost 60% (WDI 2013; note that 'industry' includes both manufacturing and extractive industries). During the same period, gross investment declined from 42.9% of GDP to 22.3% (IMF 1993; IEP 2012), while household consumption increased from 37.5 to 50.3% of GDP (WDI 2013). The bulk of these structural changes occurred in the 1990s, and while the overall economic performance during that period was quite poor, these transformations laid the foundation for the rapid economic growth that followed the 1998 crisis (see Table 1).

Post-1998 Growth and the Current State of the Economy

The post-1998 growth was accompanied by prudent macroeconomic policies and significant institutional reforms, including reforms of taxation and of intergovernmental fiscal relations. Russia faced three main macroeconomic challenges prior to the 2009 crisis: (1) how to use the surge in state revenue, which came from high oil prices for accumulating reserves, to cushion the economy in case of the deterioration of the terms of trade and to address the impending pension crisis; (2) how to prevent a significant increase in the exchange rate that would have damaged Russian exporters; and (3) how to manage ruble inflation. All three tasks were accomplished successfully, although the annual inflation during 2000–2008 ranged from 9 to 20%. Most notably, as of 1 January 2009, the two government reserve funds (National Welfare Fund and Reserve Fund) accumulated a total of $225 billion (Minfin 2013).

Russia's outstanding economic performance in 2001–2008 was in part due to the radical tax reform of 2000–2001. Among other changes, this reform introduced the 'flat' personal income tax (PIT), and reduced corporate income tax (CIT) and turnover tax. The PIT reform eliminated the 30% and 20% tax brackets and set a single 13% tax rate for all income over the zero bracket. This reform did not seem to affect labor supply much, but it did result in significantly higher compliance rates (Gorodnichenko et al. 2009). The main component of CIT reform was the reduction of the standard top rate from 35 to 24% and unification of tax rates for different types of enterprises. In addition, turnover tax rate was reduced from 4 to 1%. Overall, Russia currently has a conventional tax system (at least from the statutory point of view), although tax administration is widely seen as problematic (see Martinez-Vazquez et al. 2008; Alexeev and Conrad 2013).

The reforms of the early 2000s not only changed the tax structure, but also altered the allocation of tax revenues between the federal government and the regions, reducing tax base sharing and stabilising the revenue sources of each level of government. Expenditure assignments were also changed with the adoption of the Budget Code in 2000 (amended in 2003, 2004 and 2007) that included provisions for budget formulation and execution, accounting and auditing, and other financial relations between the various levels of government. While the Code and other laws appear to have increased fiscal centralisation, the conventional quantitative measures present an ambiguous picture, showing that the regional share of revenues is increasing and the regional share of expenditures is declining since 2001. However, the extent of fiscal independence of the regions measured by the elasticity of transfers from the centre to regional revenues diminished during the 2000s, although it could have happened due to revenue equalisation efforts by the federal government (Alexeev and Weber 2013). In any case, the political power in Russia has clearly been centralised significantly relative to the 1990s.

The early 2000s also saw significant deregulation affecting mainly small businesses. The regulation of entry (registration and licensing) and of existing businesses (inspections) was streamlined and simplified. These reforms apparently facilitated the growth of small business even though in an uneven manner across regions (Yakovlev and Zhuravskaya 2013).

Despite reforms, some of the most pernicious legacies of the Soviet period remain in the form of state control of the key sectors of the economy, weak market institutions and a natural resource rents-based economic system. Although the share of the private sector in asset ownership and in output has increased dramatically since the early 1990s (see above), the 'commanding heights' of the economy largely remain under the government's control. The state has a majority stake in the three dominant Russian banks (Sberbank, VTB and Gazprombank). Russian railways and most of the rolling stock are owned by a state-controlled corporation, RZhD, that has so far defied most attempts at reform (Pittman 2013). Despite significant recent reforms, the state owns or controls about 60% of electricity generation assets and all major transmission facilities, and the trend appears to be towards an increased state ownership in the sector (Cooke et al. 2012). Moreover, even private electricity producers can be pressured by the state to increase capacity or cap prices in certain regions (Vasin 2013). The state owns a majority stake in Gazprom, which produces almost all natural gas and has a monopoly on natural gas exports, and in Rosneft, the largest company in the most important segment of the Russian economy – the oil sector. The second largest Russian oil company, LUKoil, is 'state-influenced'. Moreover, all trunk pipelines are 100% state-owned, giving the state control over access to oil export markets (over 95% of Russia's oil exports are via pipelines), that is used to favour 'state-influenced' companies (Berkowitz and Semikolenova 2007). Most of these assets have remained under federal government control throughout the transition, although state control of much of the oil production was consolidated after the YUKOS affair and the subsequent purchase of its main assets by the government controlled companies in 2004–2005.

State control of the key economic sectors is presumably motivated in part by their 'strategic nature'. In the case of oil and natural gas, however, a major consideration appears to be unfettered access to, and control of, natural resource rents. In fact, Alexeev and Conrad (2009) found that the Russian government collected more than 90% of the net present value of a typical oil deposit – more than any other country – provided that this oil is exported. The revenue from hydrocarbons (the tax on the extraction of oil and gas and the export taxes on oil, gas and oil products, but not including profits taxes of oil and gas firms) constitutes almost 50% of Russia's federal budget revenues, or 10.5% of GDP (Minfin 2011).

An excessive dependence on oil and gas rents may be problematic for an economy, resulting in the 'oil curse' (for a recent review of the topic, see Frankel 2012). Whether Russia suffers from the oil curse is debatable, but it certainly does not appear to use oil and gas rents efficiently. Gaddy and Ickes (2013a) argue that both the Soviet and now Russian economies rely heavily on hydrocarbon and mineral rents and point out that Russia's resource rents constitute a much larger share of the economy than in Soviet times. These rents include not only formal revenues derived by and from resource extracting industries, but also the subsidies provided by these industries to their customers and suppliers. For example, Russia's customers pay much less for natural gas than their western counterparts. Also, Russia's refineries receive crude oil at prices considerably below those on the world market, although Russian drivers pay about as much for gasoline as American ones. In addition, Russian equipment suppliers appear to have preferential access to contracts with the oil and gas industry. Gaddy and Ickes view the entire Russian economy as an inverted funnel that channels resources from resource extractors to the transportation, electricity and metals sectors, which in turn end up passing some of the subsidies to heavy machinebuilding and military enterprises. Naturally, both formal and informal rents support a large and inefficient bureaucracy. The number of employees at all levels of government, including municipalities, rose from about one million in 1994 to almost 1.65 million in

2010, even though during the same period the population declined by almost 4% (Goskomstat 2003; Rosstat 2011). Meanwhile, the perceptions of government effectiveness and regulatory quality in Russia between 1996 and 2010 have not improved and have remained well below the world median (WGI 2012).

The oil and gas sector itself does not appear to be particularly efficient either. Its physical infrastructure bears the heavy imprint of Soviet legacy, particularly of poor well conditions, layout of pipeline network and location of refineries. Despite the significant changes that took place in the sector in the 1990s, including the emergence of vertically integrated oil companies, many of the institutional arrangements and regulations within the industry can be traced to Soviet times. In addition, insufficient incentives for attracting outside investment, exploration expertise, and technology exacerbate the serious problems facing the industry, particularly, the declining availability of relatively cheap oil and gas (Gustafson 2012; Moe and Kryukov 2013a, b). Moreover, competition from shale gas deposits outside of Russia, which could be developed via hydraulic fracturing, poses an additional challenge.

In addition to its vast energy reserves Russia has large deposits of valuable ores (e.g. iron, nickel and copper), but they generate much smaller rents than the hydrocarbons and are also becoming less plentiful and more difficult to mine (Fortescue 2013).

The challenges of the energy and minerals sector notwithstanding, Russia's manufacturing industry is arguably experiencing more serious problems. Although the share of manufacturing in GDP exceeds the share of extractive industries, manufacturing is often directly and indirectly subsidised by the oil and gas sector and some of its parts are protected by explicit and implicit tariffs. Without these subsidies and tariff protection, much of Russia's manufacturing would presumably be uncompetitive, at least in the short run. Another significant problem, also attributed to the Soviet legacy, is poor R&D performance, particularly at the product development stage. As part of its diversification drive away from oil and gas, the government has been trying to develop the 'innovation economy' mainly by encouraging investment in R&D activities and technology diffusion. However, tax incentives and direct investments through the creation of such vehicles as Rusnano Corporation and the Direct Investment Fund have not been successful so far. Gaddy and Ickes (2013b) argue that the wrong types and geographic allocation of physical and human capital inherited from Soviet times and the institutionalised 'rent addiction' of Russia's economy rule out successful diversification and modernisation. Under current circumstances, greater investments into 'modernisation' would only waste resources on reinforcing the existing inefficient capital and labour.

Agriculture

The agricultural sector, arguably the most problematic part of the Soviet economy, has experienced a major turnaround during the last 20 years. Through privatisation and the elimination of centralised planning, price and trade controls, agriculture has been mostly weaned off explicit farm and consumer subsidies and has undergone land reform and farm restructuring. Food shortages have been eliminated and Russia has transformed itself from the largest grain importer in the world to one of the principal exporters. Nonetheless, Russia's agricultural sector requires significant structural changes to reach world productivity levels. The woefully underdeveloped land markets have led to a highly unusual concentration of land in a small number of large corporate farms and a very large number of small farms occupying a small share of land. To a large extent, this sector's structure is a holdover from Soviet times that differs markedly from the pattern observed in market economies. Land productivity in the corporate farm sector is quite low, while it is remarkably high among the small farms. Despite using outdated technology, individual farms ranging in size from 0.5 to 10 ha produce 50% of gross agricultural output on about 20% of the land. At the same time, the 23 million small household farms are highly labour-intensive (Lerman and Sedik 2013). This situation is unlikely to change without a radical overhaul of the land market structure.

Informal Economy and Corruption

Pervasive state controls in most areas of the economy, large natural resource rents, and excessive and inefficient regulations combined with the old habits and mindset from the Soviet and even earlier times provide fertile ground for corruption and a large informal economy. The magnitudes of these phenomena are difficult to estimate with precision, but the available data show that Russia performs poorly in these respects relative to countries at the similar and even lower stage of development. According to WGI (2012), Russia ranks in the bottom 15% of the countries in terms of corruption control. Moreover, Russia's current scores are statistically significantly lower than they were in the early 2000s, returning to the level of the mid-1990s. The pervasive nature of corruption in Russia is confirmed by the survey conducted by Levin and Satarov (2013) who find that more than 50% of individuals and more than 80% of businesses have paid bribes. While these percentages have remained relatively stable over the last few years, the average size of the bribe (in US dollars) paid by businesses increased more than tenfold (!) between 2001 and 2005, although the size of bribes paid by individuals decreased in real terms. Of course, bribes do not necessarily impose net costs on businesses, and, in fact, bribes often lower the firm's costs by letting them obviate regulations or evade taxation. Also, firms may pay large kickbacks to obtain state procurement contracts that end up being highly profitable for the contractors. Nonetheless, corruption usually creates inefficiency in the economy relative to a well-functioning market system.

While the Russian government routinely announces various programmes to fight corruption, so far no effective approach has been developed. The difficulty of fighting corruption is exacerbated by the absence of genuine freedom of the press and other media, and the lack of an independent judiciary. According to the Press Freedom Index (http://en.rsf.org/pressfreedom-index-2013,1054.html) for 2013, published by the French-based Reporters without Borders, Russia ranked 148th out of 179 countries. Similarly, Russia was in the bottom quartile of countries for the 'voice and accountability' WGI (2012) indicator. With respect to *de facto* judicial independence Russia was ranked 60th out of 62 countries. Interestingly, Russia is ranked 7th in *de jure* judicial independence (Feld and Voigt 2003).

Many of the factors that determine the extent of corruption also affect the size of the unofficial economy, which facilitates and is facilitated by corruption. It is not surprising, therefore, that Russia's 'shadow economy' is estimated at almost 50% of GDP, placing it among the 25 largest of 162 countries (1999–2007 average; Schneider et al. 2010). Similarly to corruption, unofficial economic transactions may sometimes improve efficiency, given the existing excessive regulations. However, the shadow economy also generates significant welfare losses, which include depriving the state of resources for the production of public goods, inefficiently small scale of production, obstacles to the flow of information among economic agents, and distortions based on the differential ability to hide transactions from the state (Alexeev 2008; Kim 2013). On the other hand, the large unofficial sector implies that the official numbers may underestimate the size of the Russian economy. Although the Russian statistical agency Rosstat already incorporates estimates of the informal sector in its GDP data (Rosstat 2012), these estimates are less than half of those made by Schneider et al. (2010) and while the latter source shows a growing shadow economy throughout the 2000s, Rosstat estimates that the informal sector has been shrinking as a share of GDP.

Integration into the Global Economy

One of the most dramatic transformations of the last 20 years has been the elimination of the state monopoly on foreign trade early in the transition, leading to a radical increase in trade openness of the economy and the emergence of foreign direct investment (FDI) as a significant economic factor, although the stock of FDI remains relatively low (Tarr and Volchkova 2013). Russia's recent accession to the WTO will likely increase trade openness and FDI even further, and will restrain government attempts to engage in import substitution policies. Tarr and Volchkova (2013) estimate that in the medium term WTO membership will raise Russia's GDP by 3.3% of GDP per year, and in the long term the annual gain should reach 11% of GDP. These benefits will accrue across the board, but will be higher in the regions with more competitive business environments and for businesses that already operate in internationally competitive sectors, such as telecoms. The government is also trying to increase Russia's integration into the global financial system and develop local capital markets. As a part of this agenda, the government has adopted several financial liberalisation measures and aims to shift to a floating ruble by 2015 (EBRD 2012). There is a hope that increased trade and financial openness would help improve institutions in the country in the long term (Levchenko 2013). Another important trade-related development has been the creation of the Customs Union with Belarus and Kazakhstan in 2010, with the aim of establishing a common economic space that would harmonise macroeconomic and structural policies, and provide for the free movement of goods, capital and people.

Income Inequality

One positive aspect of the Soviet legacy – relatively egalitarian income distribution – has not survived market-oriented reforms. The pre-perestroika measured inequality in the USSR and Russia in particular was much lower than in the USA and only somewhat higher than in Sweden or Austria. However, unlike other East European economies that have managed to preserve their low pre-reform income inequality, in Russia the Gini coefficient for nominal per capita incomes increased from about 0.25 in 1991 to 0.42 in 2010 (Lokshin and Yemtsov 2013; CIA 2013) which is similar to the USA and considerably higher than in both East and West European countries such as Poland (0.34), Romania (0.30), Austria (0.26), Sweden (0.26) and OECD average (0.30) (WDI 2013; OECD database: http://stats.oecd.org/Index.aspx?QueryId=26067&Lang=en). The increase in income inequality since 1991 is due to the expanding gap between the income share of the top quintile and the rest of the population. The major jump in inequality occurred in the first 3 years of transition and is explained by the increasing dispersion of wages, the emergence of unemployment, a decline in labour force participation, and the low incomes of individuals engaged in subsistence agriculture and parts of the unofficial sector. Poverty rates also increased substantially at the beginning of the transition, although they declined considerably during the growth spurt of the 2000s. While government transfers, which have risen since 2007, contributed to the reduction of poverty and income inequality, further significant progress in improving income distribution can be achieved via better targeting of social assistance (Lokshin and Yemtsov 2013).

Some of the challenges of the Russian economy described above could in principle be solved if there is the will for genuine reform. Regulations could be scaled back, state-owned enterprises could be broken up and privatised, bureaucracy and the military could be reduced, further land reform could promote agricultural productivity, and better targeting of social programs could reduce income inequality and poverty. Reduced bureaucracy and regulations may alleviate corruption and improve the rule of law, although the latter would also require an overhaul of the judicial system and the police. There are, however, at least two other challenges, of demography and geography, that may be more difficult or at least considerably more expensive to solve.

The Demographic Challenge

The main demographic problem facing Russia is the rapid aging of the population and the steady contraction of working-age cohorts, mainly caused by a long-term reduction in birth rates and some recent improvements in mortality rates. The medium-term demographic situation is exacerbated by the extremely low birth rates in the 1990s and early 2000s, which, combined with particularly low life expectancy during the 1990s, led to the decline of Russia's population every year from 1993 through 2009 (Table 1). The recent net population increase, which is mainly due to a slight increase in birth rates and immigration, is almost certainly temporary, because the number of women born in the 1990s is quite small (Rosstat 2012). The government was initially slow to respond but has recently initiated generous baby bonuses (up to $13,000 for the second and each additional child) in an attempt to raise fertility. This programme appears to be aimed at the middle class, because rather than giving cash payments, the bonus can be applied only to home purchases, pension funds or child care (Denisova and Shapiro 2013). Although the programme initially resulted in significant increases in second and third births, these effects are likely to be short term, influencing only the timing of births rather than their numbers.

The immediate problem created by an aging population is the strain on the pension system, which is funded by the social payroll tax with the shortfall covered by general government revenue. The problem is exacerbated by the low retirement ages, which are set at 55 and 60 for women and men, respectively. As of 2010, the average pension amounted to almost 36% of the average wage, which is in line with international standards. The low retirement age and the aging population have resulted in a significant deficit of the pension fund, which was projected to reach 3% of GDP or 14% of the federal budget in 2012 (http://www.itar-tass.com/en/c39/370605.html). Moreover, under the current policy, the pension fund deficit will rise dramatically in the next few decades, mainly because of the expected doubling of the ratio of the non-working age population (those 65 and older) from 18% in 2010 to 36% in 2050 (Eich et al. 2012).

Population aging is in part a result of improvements in healthcare. After a significant deterioration in the 1990s, the healthcare system appears to be improving as a result of significantly increased government spending and organisational reforms. In fact, public spending on healthcare, which dipped substantially below the 1991 level in 1999–2000, then rose above this level by 2010. Nonetheless, at about 3.1% of GDP, public expenditure on healthcare in Russia lags well behind all of the OECD countries (with the exception of Mexico), and individuals' out-of-pocket expenses still constitute almost 40% of the total expenditure (Shishkin 2013). The main organisational reforms of healthcare since Soviet times have been the development of officially sanctioned private for-profit provision of health services and the introduction of compulsory health insurance funded mainly by employer contributions and by the regional and federal governments. However, despite reforms and a substantial increase in public funding, health outcomes lag those of other OECD countries (Shishkin 2013; Denisova and Shapiro 2013).

Geographic Challenges

One of the most important 'geographic' problems is the distorted distribution of the population bequeathed to Russia by Soviet policies. Most important, the share of the population living in colder areas is much higher than in other countries (Hill and Gaddy 2003). This is largely due to the Soviet policies of building labour camps in the Far North and shifting some strategic industries away from the European part of Russia, aided by the location of important natural resource deposits such as oil, natural gas and nickel. Despite the (modest) government subsidies, migration from the Far North during the transition has reached only about 8% of the region's population in the early 1990s, and even declined afterwards. Another geographic problem left over from the Soviet era is the distorted distribution of city sizes, which deviates significantly from those in most other large countries (World Bank 2005). It is expected that migration will eventually alleviate these distortions. However the migration may be disruptive and costly and, at this time, no significant trend for reversing the Soviet geographic distortions appears to be emerging (Markevich and Mikhailova 2013). Russia's geographic problems are exacerbated by its poor transportation infrastructure, especially the road network. Since 1990, the number of passenger cars has increased almost fourfold while the length of paved roads has risen only about 20% (Rosstat 2012). The situation is particularly dire in urban areas (World Bank 2012).

Some of the important strengths and weaknesses of Russia's economy were revealed by the 2008–2009 global economic crisis. Largely sound economic policies, combined with substantial resources, allowed the government to handle the initial stage of the crisis well. The government was able to prevent financial and banking collapse, and, moreover, to maintain various social programmes, increase pensions and subsidise large enterprises to limit unemployment while refraining from the nationalisation of struggling firms. The crisis highlighted the heavy dependence of the economy on oil prices, whose dramatic drop in the second half of 2008 was presumably the main reason for Russia having the largest decline in GDP among G-20 countries in 2009 (Guriev and Tsyvinski 2010). Moreover, the reliance on oil prices seems to be increasing. Whereas in 2007 the $69 per barrel price of Russian oil resulted in a budget surplus of 5.5% of GDP, the oil price required to balance the budget in 2011 was $108, and the Russian Ministry of Finance estimates that a $117 per barrel price would be needed in 2013 (Minfin 2011). These developments are due mainly to the significantly increased expenditure commitments of the federal government, including a substantially higher military budget and expanded social programs. On the other hand, government reserve funds have been growing again, reaching almost $151 billion at the beginning of 2013 (Minfin 2013) and the postcrisis inflation fell to 6.1% in 2011 and 6.6% in 2012 (CBR 2013).

Conclusion

We conclude that despite Soviet legacies, Russia has shown strong economic performance since 1998. If oil prices increase modestly from their current levels, the economy should be able to maintain respectable growth rates in the near future. However, Russia continues to face social and economic challenges that cannot be resolved without significant institutional reforms. The conventional view is that dramatic reforms are unlikely as long as oil prices remain relatively high, which could create the possibility of exacerbating economic and social difficulties in the more distant future. If the Russian economy does not reform and is hit by another economic crisis, it may not have the same luxury of large financial reserves to meet the challenges, thus, raising the danger of the repeated historical pattern of growth spurts followed by stagnation that have characterised Russia over the last 300 years.

See Also

▶ Command Economy
▶ Second Economy (Unofficial Economy)
▶ Soviet Economic Reform
▶ Soviet Growth Record
▶ Transition and Institutions

Bibliography

Alexeev, M. 2008. *Second economy. New Palgrave dictionary of economics*. New York: Palgrave Macmillan.

Alexeev, M., and R. Conrad. 2009. The Russian oil tax regime: A comparative perspective. *Eurasian Geography and Economics* 50(1): 93–114.

Alexeev, M., and R. Conrad. 2013. The Russian tax system. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Alexeev, M., and S. Weber. 2013. Russian fiscal federalism: Impact of political and fiscal (de)centralization. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Berkowitz, D. and Y. Semikolenova. 2007. *Is Russia's growth a 'flash in the pan'?*. Working paper 318. Department of Economics, University of Pittsburgh.

BP. 2012. BP *Statistical review of world energy*. Available at: http://www.bp.com/sectiongenericarticle800.do?categoryId=9037130&contentId=7068669.

CBR. 2013. *Central Bank of Russia* website. http://www.cbr.ru/statistics/.

CIA. 2013. *The World Fact Book*. https://www.cia.gov/library/publications/theworld-factbook/.

Cooke, D., A. Antonyuk and I. Murray. 2012. *Toward a more efficient and innovative electricity sector in Russia*. Consultation Paper, International Energy Agency, Insights Series.

Denisova, I., and J. Shapiro. 2013. Recent demographic developments in the Russian Federation. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

EBRD. 2012. *Transition report* 2012: *Integration across borders, European bank for reconstruction and development*. Available at: http://www.ebrd.com/transitionreport.

Eich, F., C. Gust, and M. Soto. 2012. *Reforming the public pension system in the Russian Federation*. IMF working paper, WP 12/201.

Ericson, R. 2013. Command economy and its legacy. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Feld, L., and S. Voigt. 2003. Economic growth and judicial independence: Cross-country evidence using a new set of indicators. *European Journal of Political Economy* 19(3): 497–527.

Fortescue, S. 2013. The economics of mineral resources. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Frankel, J. 2012. The natural resource curse: A survey of diagnoses and some prescriptions. In *Commodity price volatility and inclusive growth in Low-income countries*, ed. R. Arezki, C.A. Pattillo, M. Quintyn, and M. Zhu. Washington, DC: International Monetary Fund.

Gaddy, C., and B. Ickes. 2013a. Russia's dependence on resources. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Gaddy, C., and B. Ickes. 2013b. *Bear traps on Russia's path to modernization*. London: Routledge.

Gorodnichenko, Y., J. Martinez-Vazquez, and K.S. Peter. 2009. Myth and reality of the flat tax reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy* 117(3): 504–554.

Goskomstat. 2003. *Russian statistical handbook* 2003 (in Russian). Moscow: Goskomstat.

Guriev, S., and A. Tsyvinski. 2010. Chapter 1. Challenges facing the Russian economy after the crisis. In *Russia after the global economic crisis*, ed. A. Aslund, S. Guriev, and A. Kuchins. Washington, DC: Peterson Institute for International Economics.

Gustafson, T. 2012. *Wheel of fortune: The battle for oil and power in Russia*. Cambridge, MA: Harvard University Press.

Hill, F., and C. Gaddy. 2003. *The Siberian curse: How communist planners left Russia out in the cold*. Washington, DC: Brookings Institution Press.

IEP. 2012. *Russian economy in 2011: Trends and outlooks*. Moscow: Gaidar Institute for Economic Policy.

IET. 2006. *Russian economy in 2005: Trends and outlooks*. Moscow: Institute for the Economy in Transition.

IMF. 1993. *Russian federation*, IMF economic reviews, No. 8. Washington, DC: International Monetary Fund.

Kim, B.-K. 2013. The unofficial economy in Russia. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Lerman, Z., and D. Sedik. 2013. Russian agriculture and transition. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Levchenko, A. 2013. International trade and institutional change. *Journal of Law, Economics, and Organization*.

Levin, M., and G. Satarov. 2013. Corruption. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Lokshin, M., and R. Yemtsov. 2013. Poverty and inequality in Russia. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Markevich, A., and T. Mikhailova. 2013. Economic geography of Russia. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Martinez-Vazquez, J., M. Rider, and S. Wallace. 2008. *Tax reform in Russia*. Northampton: Edward Elgar.

Mau, V., and T. Drobyshevskaya. 2013. Modernization and the Russian economy: Three hundred years of catching up. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Minfin. 2011. *Main results and directions of budget policy for 2012 and through 2014* (in Russian). Moscow: Ministry of Finance of the Russian Federation.

Minfin. 2013. *Russian Ministry of Finance websites*: http://www1.minfin.ru/en/reservefund/statistics/amount/ and http://www1.minfin.ru/en/nationalwealthfund/statistics/amount/.

Moe, A., and V. Kryukov. 2013a. The Russian oil sector. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Moe, A., and V. Kryukov. 2013b. The natural gas sector. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Pittman, R. 2013. Blame the switchman? Russian railways restructuring after ten years. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Rosstat. 2011. *Russian statistical handbook 2011*. Moscow: Rosstat.

Rosstat. 2012. *Russian Statistical Agency Website* (data placed into the website in 2012): http://www.gks.ru/wps/wcm/connect/rosstat/rosstatsite/main/account/#.

Schneider, F., A. Buehn and C. E. Montenegro. 2010. *Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007*. Policy research working paper series 5356. The World Bank.

Shishkin, S. 2013. Russia's healthcare system: Difficult path of reform. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Tarr, D., and N. Volchkova. 2013. Russian trade and foreign direct investment policy at the crossroads. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

Vasin, A. 2013. Russian electricity market: Development scenarios. In *Oxford handbook of Russian economy*, ed. M. Alexeev and S. Weber. New York: Oxford University Press.

WDI. 2013. *World Development Indicators database*: http://data.worldbank.org/data-catalog/world-development-indicators.

WGI. 2012. *World Governance Indicators database*: http://info.worldbank.org/governance/wgi/index.asp.

World Bank. 2005. *Russian Federation: From transition to development, a country economic memorandum for the Russian Federation*. Washington, DC: World Bank.

World Bank. 2012. *Reinvigorating the economy*, Russian economic report, No. 28. Washington, DC: World Bank.

Yakovlev, E. and E. Zhuravskaya. 2013. The unequal enforcement of liberalization: Evidence from Russia's reform of business regulation. *Journal of European Economic Association*, forthcoming.

Ryazanov, David (1870–1938)

D. J. Struik

Ryazanov was born David Borisovich Goldendach on 10 March 1870, in Odessa. Because of his connections, first with the Narodniks, then with the budding social democracy, he spent several years in prison. In 1898, he joined the new Russian Social Democratic Party, belonging after 1903 to the Menshevik wing. Between 1900 and 1905, he did research abroad on the labour movement and contributed to Kautsky's *Neue Zeit.* He participated in the Revolution of 1905, and by 1907 was again in Germany, doing that research on Marx and Engels on which his fame is mainly based.

By this time Franz Mehring and others had started the publication of works of Marx and Engels hidden in archives, private collections and often obscure periodicals. Ryazanov contributed two volumes, *Gesammelte Schriften von K. Marx und F. Engels, 1852 bis 1862* (1920, published in Stuttgart and translated into German by Luise Kautsky), which contain among others the writings on the Crimean War and on Palmerston. The war intervened with a study on the First International, which was only published in 1926 as *Die Entstehung der Internationalen Arbeiter Assoziation* (Marx–Engels Archiv I, Frankfurt am Main).

The revolution of 1917 brought Ryazanov back to Russia, where he joined the Bolsheviks, who formed the Communist Party in 1918. He placed all his knowledge at the service of the Soviet State, and in 1920 became director of the new Marx–Engels Institute. His main purpose was the preparation of the collected works of Marx and Engels. To this end, Ryazanov went travelling abroad, collecting, copying, buying whatever he could find, including material from the rich archives of the German Social Democratic Party. The Institute bought up whole libraries on economic and labour conditions in various countries. Starting from scratch, the Institute in 1930 possessed 55,000 pages of photostats, 32,000 pamphlets, 450,000 books and periodicals, and was growing.

The Russian edition of the works of Marx and Engels came out between 1931 and 1951 in 28 volumes. The edition in the original languages included only seven volumes, containing works up to 1848. It is known as the MEGA, short for *Marx–Engels Gesamtausgabe* (published in Berlin, Moscow and Leningrad, 1927–35). It made available the *Deutsche Ideologie* and the *Economic–Philosophic Manuscripts of 1844.* The *Dialectics of Nature* came out in *Marx–Engels Archiv* II (1927, 117–395). The Institute also published many other works of marxist authors, such as Plekanov and Liebknecht.

Ryazanov's lectures on Marx and Engels, published in 1923 and 1928 in Russian, were published in English as *Karl Marx and Friedrich Engels* (1927) and republished with a new preface in 1973. His remarkable edition of *The Communist Manifesto* appeared in English in 1930.

Because of his involvement in Menshevik activity, Ryazanov lost his position in 1931, and was succeeded by V.V. Adoratskilz (1878–1945). He spent some time in Saratov and Leningrad, doing research. He died in Saratov in 1938.

Selected Works

1917. (ed.) *Gesammelte Schriften von Karl Marx und Friedrich Engels 1852 bis 1862*. 2 vols. Trans. L. Kautsky. Stuttgart: J.H.W. Dietz.

1926. (ed.) *Die Entstehung der Internationalen Arbeiter Assoziation. Marx–Engels Archiv I*. Frankfurt am Main.

1927a. (ed.) *Dialectics of Nature. Marx–Engels Archiv II*, 117–395.

1927b. *Karl Marx and Friedrich Engels*. New York: International Publishers. Republished with new preface by D.J. Struik, New York and London: Monthly Review Press, 1973.

1927–35. (ed.) *Marx–Engels Gesamtausgabe (full title: Karl Marx, Friedrich Engels, Historisch-Kritische Gesamtausgabe, Werke, Schrifte, Briefe)*. 7 vols. Berlin/Moscow-Leningrad: Marx–Engels Institute.

1930. *The Communist Manifesto of Karl Marx and Friedrich Engels. Introduction and explanatory note by D. Ryazanov*. Trans. from the Russian ed of 1922. London: Martin Lawrence.