Uwe Cantner
Elias Dinopoulos
Robert F. Lanzillotti
Editors

Entrepreneurship, the New Economy and Public Policy

Schumpeterian Perspectives

Springer

Entrepreneurship, the New Economy and Public Policy

Uwe Cantner · Elias Dinopoulos
Robert F. Lanzillotti
Editors

Entrepreneurship, the New Economy and Public Policy

Schumpeterian Perspectives

With 59 Figures
and 52 Tables

Springer

Professor Dr. Uwe Cantner

Friedrich-Schiller-University of Jena
Faculty of Economics and Business Administration
Chair for Microeconomics
Carl-Zeiss-Str. 3
07743 Jena, Germany
E-mail: uwe.cantner@wiwi.uni-jena.de

Professor Elias Dinopoulos

University of Florida
Department of Economics
Gainesville, FL 32611, USA
E-mail: elias.dinopoulos@cba.ufl.edu

Professor Robert F. Lanzillotti

University of Florida
Graduate School of Business
201 Bryan Hall
Gainesville, FL 32611-7150, USA
E-Mail: robert.lanzillotti@cba.ufl.edu

Some of the contributions habe been published in
"Journal of Evolutionary Economics", Vol. 13, No 5, 2003

Cataloging-in-Publication Data applied for

Library of Congress Control Number: 2004113933

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data available in the internet at *http://dnb.ddb.de*

ISBN 3-540-22613-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Cover design: Erich Kirchner, Heidelberg
Production: Helmut Petri
Printing: betz-druck

SPIN 11307396 Printed on acid-free paper – 42/3130 – 5 4 3 2 1 0

Table of Contents

Editorial

In 2002, the 9th International Joseph A. Schumpeter Society Congress was hosted in the University of Florida, Gainesville. The theme of the conference was *Entrepreneurship, the New Economy and Public Policy: Schumpeterian Perspectives* which served as an umbrella to more than 160 participants who delivered presentations on several current research topics in the fields of evolutionary economics, technological change and industry evolution, industrial organization and public policy, economic growth, and historical perspectives on Schumpeter. The papers in this volume are representative of the above mentioned topics.

The selection starts with a highlight of the conference's plenary sessions, Paul Samuelson's presentation *Reflections on the Schumpeter I knew well*. It contains a very personal reminiscence and specific view on the personality of Joseph Schumpeter including their last lengthy conversation in New York City, two weeks before Professor Schumpeter died in his sleep on January 7–8, 1950 at the age of 67.

The conference's presidential address by Robert Lanzillotti as well as the paper by Mariana Mazzucato are concerned with the *innovation driven* development of industries. Robert Lanzillotti, in *Schumpeter, product innovation and public policy: the case of cigarettes*, directly takes up the conference's theme by relating Schumpeterian innovation and competition to *antitrust policy* in the U.S. tobacco industry. Mariana Mazzucato's paper *Risk, variety and volatility: growth, innovation and stock prices in early industry evolution* applies the Schumpeterian framework of *creative destruction* to the co-evolution of the volatility of firms' growth rates and their stock prices.

Olav Sorensen as well as Junfu Zhang addressed the conditions under which *entrepreneurs* emerge and the resulting *geographical distribution* of entrepreneurship. Olav Sorensen's contribution *Social networks and industrial geography* puts forward the hypothesis that current geographic distributions of production are an important factor that place constraints on entrepreneurial activity. For the Schumpeterian mechanism of swarming – that is one or a few entrepreneurs facilitates the appearance of additional entrepreneurs – Junfu Zhang in *Growing Silicon Valley on a landscape: an agent-based approach to high-tech industrial clusters* conquers the traditional view that locally concentrated entrepreneurship is based on entrepreneurs' choices to move to a high-tech cluster. He also contrasts the traditional view to the impact that social contacts have on the formation of new firms.

The following four papers are concerned with the circumstances, conditions and organizational modes under which innovations can be successfully commercialized. In this respect, Gunnar and Åsa Eliasson in *The Theory of the firm and the markets for strategic acquisitions* address the determinants of innovative-firm formation, such as venture capital competencies, competencies for balancing innovative and productive efficiency, and the role of strategic acquisitions. Their analysis includes a comparison of the markets for innovation, entrepreneurship and venture capital in the US and Europe. In *The growth of commercialization-facilitating organizations and practices: a Schumpeterian perspective*, Sten Thore and Robert Ronstadt analyze the role of new financial vehicles, new organizational forms, and new practices promoting the transfer of new technology from the laboratory to the market as "commercialization facilitators". Their discussion is exemplified with a prominent "second-order" facilitator, the IC^2 Institute at the University of Texas. The problem of appropriating rents from intangible assets, which is a prominent feature of the New Economy, is addressed in Gunnar Eliasson and Clas Wihlborg's contribution entitled *On the macroeconomic effects of establishing tradability in weak property rights*. The authors provide valuable insights on the difficulty to define property-rights and all resulting incentive problems that might lead to under-investment when intangible assets become more prevalent. Innovative ways to price intangible assets are proposed and the so-called enabling laws are suggested. Jim Stewart's paper *Capital in the New Economy: a Schumpeterian perspective* is also concerned with the issue of investment, the uncertainty involved and the value of capital in a Schumpeterian world of creative destruction and uncertainty. Stewart argues that the entrepreneurial investment decision is difficult to reconcile with rational choice and the use of real options could be a solution to this dilemma.

A *Schumpeterian* view on productivity growth across different sectors is presented by Hans Lööf in his contribution *A comparative perspective on innovation and productivity in manufacturing and services* and by Michael Peneder in *Tracing empirical trails of Schumpeterian development*. Lööf investigates the R&D-innovation-productivity relationship in manufacturing and in services and discovers a surprisingly high similarity in this relationship between firms in each sector. Michael Peneder's contribution focuses on the interplay of the intersectoral qualitative changes and economic growth, which constitute two key concepts of *Schumpeterian* development. With a major focus on manufacturing versus services, he shows that variations in industrial structure are significant determinants of aggregate income levels and growth.

The effect of micro structures and their change on macro outcomes - as already included in Michael Peneder's study – is central to the papers by Giorgio Fagiolo, Giovanni Dosi and Roberto Gabriele and by Carolina Castaldi. In *Towards an evolutionary interpretation of aggregate labor market regularities*, Fagiolo et al. use an agent-based simulation methodology to examine how the Beveridge, Wage and Okun curves are related to micro-dynamics on labor and product markets when agents are non-optimizers. In *An evolutionary model of international competition and growth* Castaldi shows – also using an agent-based simulation analysis – how international growth differentials among countries can be traced back to differences in the innovative and imitative activities of firms and their respective effect on a countries international competitive position. It is shown that the international technology diffusion has a positive effect on average growth performance but contributes to growth differences among countries due to positive feedback mechanisms.

The papers by Hariolf Grupp, Icíar Dominguez Lacasa, Monika Friedrich-Nishio and Andre Jungmittag, by Masaaki Hirooka, and by Iordanis Petsas deal with the macroeconomic effects of innovative activity and technological progress. Grupp et al. present an empirical study on *Innovation and growth in Germany in the past 150 years* and investigate the sources and effects of technological activities in Germany. They show that the long period of 150 years can be divided into two sub-periods, 1850–1913 and 1951–1999, which differ considerably with respect to the public contribution to innovations. In the former period the number of patents granted is mainly influenced by demand, whereas in the latter period public and private R&D expenditures are of major importance. Schumpeter's theory of the *business cycle* is the template used by Masaaki Hirooka's paper entitled *Nonlinear dynamism of innovation and business cycles*. He investigates the relationship between innovations, new-market formation, the bubble economy and economic depression. In *The dynamic effects of general purpose technologies on Schumpeterian growth*, Iordanis Petsas analyzes the s-curve diffusion path of general purpose technologies and the resulting effects on long-run and transitional *Schumpeterian growth* rate.

Each paper by its own, but also the selected collection of papers, provide an excellent guide to the range of issues discussed at the conference and demonstrate the high relevance and still thought-provoking nature of Schumpeterian economics and its policy relevance.

Finally, we want to thank our referees for their assistance in improving the papers of this volume, Olga Gässner for her careful handling of the correspondence with the authors and editors, and Holger Graf for his thoughtful and diligent preparation of the final layout.

Gainesville and Jena,
June 2004

Robert F. Lanzillotti
Elias Dinopoulos
Uwe Cantner

Reflections on the Schumpeter I knew well

Paul A. Samuelson

Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive,
Cambridge, MA 02142, USA

> *"Ah, did you once see Shelley plain,*
> *And did he stop and speak to you,*
> *And did you speak to him again?*
> *How strange it seems, and new!"*
>
> *Robert Browning, Memorabilia, 1855.*

In December of 1934, when I was an acne aged senior at the University of Chicago cleaning off pictures of Adam Smith, Böhm-Bawerk, and Alfred Marshall for the Economics Department, George Stigler and Allen Wallis were gigantic graduate students exercising squatter sovereignty over a basement storage room in the new Social Science Research Building. They told me that the American Economic Association was holding its Christmas annual meeting downtown in the Palmer House and suggested that I might want to pay that zoo a visit.

That is how I first saw Shelley plain. In one statistics section a roly-poly Harold Hotelling introduced Bill Madow who put some of the biggest matrices known to man on the blackboard. (That left its mark on me.)

Then, down the hall, to an overflow audience a florid Arthur Marget sang the flowery praises of a Harvard speaker whose name I could not catch. From Harvard he may have come, but I could catch no meaning from his energetic gibberish about "kitchens" and "spaghettis". Only a year later did I come to realize his Germanic-English was preaching about short-term Crum-Kitchen business cycles, intermediate-length Juglar cycles and longest-run Kondratief waves. He spoke dynamically and dramatically, and since his own jokes made him laugh, I nervously joined in with the crowd's frequent applauses. It was not love at first sight but he did capture my interest.

Still, the following week when Walllis and Stigler asked what I had learned at the AEA, I replied "Harry Carver from Michigan math department suggested, 'to avoid the sample assumption of normality, permute the sample's measured properties with that universe's means properties'." Wallis then observed, "Samuelson, that's the silliest idea I've ever heard." This is really a story about Allen Wallis since the famous bootstrap technique did become important in statistics only some forty years later.

When I finally came to mention Schumpeter, George Stigler snorted, "Isn't he the nut who believes the interest rate to be zero in the stationary state?" I didn't then have the wit or the brashness to reply, "Yes, and Frank Knight is the nut who believes that the interest rate can *never* be zero in the stationary state." Knight was then our local Chicago Isaac Newton.

A year later I won a juicy SSRC fellowship that would pay all my graduate school expenses provided I went to a different university. So I was bribed to leave the midway Valhalla. In June 1935, without exception my Chicago teachers – Simons, Director, Knight, Douglas, Viner, Gideonse, Mints, Nef, and Yntema – recommended Columbia over Harvard. By lucky miscalculation I brashly ignored their wisdom. But it was not in order to sit at the Schumpeter knee. My hallucination was that the Harvard Yard would, like the Dartmouth lawns at Hanover, New Hampshire: a pretty white church on the hill, and much green ivy. Ed Chamberlain's 1933 *Theory of Monopolistic Competition*, which had never been assigned in any of my numerous Chicago courses, I found on a Reserve shelf and much enjoyed. But it took scarcely a month in the busy Harvard Yard to realise that Chamberlin was indeed a one-book-only man.

By sheer luck – my good luck – Harvard was about to come out of a lean period, led by an infusion of European talent: Schumpeter, Leontief and Haberler; buttressed too by the powerful soon-to-be arrival of Alvin Hansen, and my discovery of Edwin Bidwell Wilson, mathematician, physicist, statistician and only protégé of Yale's Willard Gibbs. Had I stayed at my beloved Chicago I would have missed completely, or had to fabricate alone, three great revolutions: the Keynesian revolution, the mathematical revolution, and the imperfect-competition revolution; besides, the Chicago ideology was an infantile eczema okay for one's teens but much in need of outside sunshine.

Today's lecture began with meeting Schumpeter for the first time. Lest the fast-moving clock choke off an account of how exactly fifteen years later I was the last professional economist to talk at length with Joseph Schumpeter, let me postpone for another occasion those intervening years. In 1949, this time in New York City, the AEA again met just after Christmas. Joseph and Paul had no inkling then that within a dozen days Professor Schumpeter would die in his sleep at the Taconic, Connecticut country estate of his American third wife, Elizabeth Boody Schumpeter. It can be said that Schumpeter's death was the gift of the gods – it came unannounced; with his boots on, he died going full-tilt. Born February 8, 1883 (four months before Keynes), Schumpeter died January 7-8, 1950 at but age 67.

(Keynes, perhaps from a heart damaged at Eton by rheumatic fever, was granted only 63 crowded years.) Best of all for a scholar like Schumpeter, his huge *History of Economic Analysis* was near enough to completion that his economic historian widow could supervise its final editing with the help of a few friends (Wassily Leontief, Richard M. Goodwin, Gottfried Haberler,...).

Pascal and Felix Mendelssohn were prodigiously precocious. But when each died before reaching age forty, each was physiologically an old man. Not so with Mozart – from him could have been extrapolated as much again in the future as had generously erupted in the past.

Schumpeter used to joke that in his seventies he would write his treatise on logic. In his eighties would come his sociological novel, and in preparation for it, he would for a second time ride the Boston subway. That was not to be.

When Schumpeter received an honorary Columbia degree – at age 30! – old Frank Fetter took him to a Princeton football game. Once was enough for a quick learner like Joseph Alois. Less promising was the attempt by Bob Bishop and me to initiate Schumpeter into the intricacies of poker following a semiriotous cocktail party. The gallant scholar could not seem to realise that if he financed everbody's losings, the game might lose some of its zest.

I recollect that Schumpeter was an active participant at several of the 1949 AEA sessions. His *Capitalism, Socialism, and Democracy,* which he professed to despise as an off-the-cuff pot-boiler, had been a great success in the 1940s. By contrast the two volumes of *Business cycles*, which drained much of his energies in the 1930s, made no considerable splash. Perhaps Keynes'1936 *General Theory* made it seem anti-climatic.

When we met in the hotel bar, Schumpeter told me of plans to prepare an important invited Chicago lecture during the coming January. Significant to me was a seeming recantation expressed by Schumpeter at an autumn 1949 NBER Business Cycles conference held just prior to the Christmas AEA meeting. Remember that Schumpeter, from the time of his first German book in his Edwardian-Age youth, had been the Viennese heretic who did not believe Walrasian general equilibrium; moreover, Schumpeter was the electic who shocked continental contemporaries by praising J.B. Clark and Knut Wicksell. As far as my own career was concerned, he egged me on to discover and utilize new mathematical tools in economics – even though, Moses-like, JAS himself was never to cross over into the Promised Land of Pareto, Hotelling, Tinbergen, and Frisch – to say nothing of Arrow, Debreu and Koopmans.

Imagine then the surprise that greeted his 1949 NBER statement, which I paraphrase as follows :

> Yes, econometric mathematics and statistics are fruitful tools for
> the future science of economics.
>
> But if the good fairies will allot you only one of *economic history* or
> *Mathematical econometrics,* then to become an outstanding economist,
> master the corpus of economic history.

Was this the ranting of a decaying arteriosclerotic mind, poised two months from extinction? My evidence is against that. The terminal Schumpeter was lucid and witty and often wise. Wicksell, late in life, lost some of his earlier unearned faith in Say's Law and Neutral Money. The realities of the 1930s Great Depression, which Irving Fisher personally suffered from especially, left Fisher before his death with a wiser and more qualified version of the *Quantity Theory of Money* than graced the pre-1927 macro writings of Fisher, Marshall, Pigou and even pre-1925 John Maynard Keynes.

Let me return to, so to speak, our last supper. One topic the two of us discussed was if I could help Dick Goodwin get a good job, one that he would think good enough for him to accept.

> JAS: I wish Harvard would give Goodwin a tenured chair. But that limited crew will veto any such nomination. As an alternative, I have offered to fight for a lifetime fellowship for this worthy scholar of simple tastes and a desire to be a modern painter part time. But Dick is a proud man and says, "If I cannot be a first-class member of the club, it is not the club for me."

Amherst, or Williams, or the University of Michigan Schumpeter mentioned as possible destinations but Goodwin (this Rhodes Scholar from a small Indiana high school that had never previously sent a graduate to Harvard) was just not interested.

> PAS: I am at a loss. Perhaps going abroad to Oxbridge will be the most hopeful exploration.

I digress to report that Goodwin was welcomed to Peterhouse, Cambridge University, where he taught and did research until his middle sixties, when he retired to a second career as an Italian professor in Sienna. This was an Indian summer for Richard Goodwin. Senator Joseph McCarthy's early 1950s research witch hunts played some role in the Goodwin saga, although in late 1949 I had no inkling of that. I ought to add that although Schumpeter made plain after 1932 that he himself had not been a refugee from Adolph Hitler's Bonn University, many refugees were helped by Schumpeter to settle in an American university.

At this point the two of us were joined at the hotel bar by Gottfried Haberler, Schumpeter‘s younger Vienna colleague at Harvard and by the colourful Imre de Vegh, who became Schumpeter‘s financial executor and who was a Hungarian aristocrat trained at Cambridge University. De Vegh was one of the first of the post-World War II "performers" as a Wall Street money-manager investor, first at Scudders, Stevens and Clark, and then later for his own firm. De Vegh had a great penchant for academies and was prone to press his hospitality on visitors from out of town.

To put it plainly, de Vegh then proceeded to kidnap us three academics, pouring us into a taxicab on the way to his penthouse apartment where his (surprised) wife was to cook us dinner. When we arrived, Mrs. de Vegh was in her kimono, curling the hair of her cute young daughter. She assured us that it was no imposition and that she did have a frozen Bird's-Eye chicken (then an innovation) to cook up for our dinner. In the meantime, our kidnapper host lay down for a noisy nap, leaving us three to amuse ourselves in the living room.

Crafty Gottfried Haberler soon escaped quietly. Schumpeter and I were again a twosome. We talked of many things. One, I remember, was Estelle Leontief‘s novel that John Day was contemplating publishing. Schumpeter judged it to be promising, but he didn't know whether she could meet the publisher's request to add more sex to it.

Finally, my wise master said: "What are we doing here? Gottfried is smarter than us." So we sneaked into the bedroom where our coats were hung and where our host dreamed on. Tiptoeing, we made our way to a taxi. That was the last I saw of my master. And a dozen days later, on snowy roads from Boston to Connecticut, Gottfried and wife Friedl in the front seat drove me surrounded by Goodwin on the right and by Schumpeter's then current assistant Alf Conrad on the left, to Schumpeter's funeral.

Of course we gossiped of good old days in the Harvard Yard and Littauer Hall. Did we find many good things to say about the late departed? Indeed we did. Was he completely without faults? Of course not, and long ago I've commented on some of them. Remember this, though: In chess you are only as good as your worst move. In creative science you are as good as your best moves.

Let me again solemnly affirm, at this place and before this company, that Joseph Schumpeter greatly enriched my life and enriched that of my late wife Marion Crawford who was perhaps his earliest research assistant. His support for my scholarly career was intense and without limits.

When Schumpeter died, he may have been the most frequently cited living economist.

Now, at the turn of the millennia, when total-factor-productivity has remarkably soared in America and abroad, both fools and sages sing Schumpeter's praise. That would have amused and pleased this worldly scholar who in some dark hours of the night used to despair in his German-shorthand diaries of justly deserved praises passing him by. So Keynes was wrong: in the long run not all of us are dead.

Schumpeter, product innovation and public policy: the case of cigarettes*

Robert F. Lanzillotti

Department of Economics, University of Florida, Gainesville, FL 32611-7140, USA
(e-mail: robert.lanzillotti@cba.ufl.edu)

Introduction

Perhaps the most fundamental proposition advanced by Joseph Schumpeter is that competition drives innovation and innovation drives progress. Strangely, although "competition" was a key ingredient to the processes of his capitalistic world, Schumpeter was not concerned about private monopolies and evidently had little stomach for the antitrust policy practiced over the past 50 years in the United States, or for the rather aggressive policy that is emerging currently in the European Common Market.[1] At the same time, a careful reading of his writings discloses that Schumpeter's benign view of monopolies and restraints of trade was limited to **product markets**; he expressed a definite concern about monopoly and restraints of trade in **R & D-innovation markets**. Schumpeter also displayed a very prophetic insight about R & D/innovation and public policy, observing: "... *Surely nothing can be more plain or even more trite common sense than the proposition that innovation ... is at the center of practically all the phenomena, difficulties, and problems of economic life in capitalist society*."[2]

As keen a vision as Schumpeter displayed about economic processes and public policy, he could not have imagined that his theories of innovation and "creative destruction" would become a centerpiece during the 1990s for the widespread antitrust prosecutions by several states against U.S. tobacco companies alleging a conspiracy in restraint of trade and other types of unlawful conduct, and for recovery

---

* The author gratefully acknowledges helpful comments and suggestions of Joel B. Dirlam, James T. McClave, Paul A. Samuelson, David Sappington, and Thomas R. Saving.

[1] "... There is no general case for indiscriminate 'trust busting' or for the prosecution of everything that qualifies as a restraint of trade... Pure cases of long-run monopoly must be of the rarest occurrence..." Joseph A. Schumpeter, Capitalism, Socialism, and Democracy, 3rd edn. (Harper & Bros, New York, 1950; orig. pub. 1942).

[2] Joseph A. Schumpeter, Business Cycles: A theoretical, historical, and statistical analysis of the capitalist process (McGraw Hill, New York, 1939), p. 87.

of damages resulting from health-care costs incurred. In what ranks as the largest product liability case in U.S. history, in 1998, the tobacco companies entered into a multi-billion settlement with the states for Medicaid costs attributed to treatment of tobacco-related illnesses.[3]

The complaints contend that in 1953 the tobacco companies agreed not to compete with one another on R & D covering smoking and health effects, and not to develop "safer" cigarette products.[4] To support the constrained R & D contention, plaintiff economists testified that the conspiracy resulted in lower research and development expenditures by the tobacco companies; and that, absent the conspiracy, higher levels of R & D expenditures would have been made, resulting in a much "safer," or even a completely "safe" cigarette product.[5]

This paper addresses the major hypothesis underlying the antitrust complaint, namely that the new cigarette products introduced by the tobacco companies between 1955 and 1995 do not meet the Schumpeterian standard for innovation. First, we briefly review Schumpeter's concept of innovation and creative destruction. Second, we examine findings of earlier studies on non-price competition in the cigarette industry since the 1950s *vis a vis* the "constrained R & D" contention. Third, we present data on R & D activities of the tobacco companies, on cigarette products which lowered tar and nicotine levels, and on programs promoting the sale of new products. Fourth, we discuss the nature of R & D processes generally, and the tobacco companies' role in basic medical research on tobacco-related diseases. Fifth, we examine the plaintiff experts' delineation of the product and geographic markets relevant to the antitrust complaint.

1 The meaning of innovation and "creative destruction" in Schumpeter's world

In *Capitalism, Socialism and Democracy,* Schumpeter explained that the fuel which propels capitalism is the constant injection of new innovations in the form of new consumer goods, new production techniques, new modes of transportation, and

---

[3] An estimated $206 billion is projected to be distributed to 46 states by 2025, unless tobacco sales decline, in which case that figure will be reduced. The other four states – Florida, Minnesota, Mississippi and Texas – entered into separate settlements with the leading tobacco companies.

[4] Our analysis of the merits of the complaint is based upon an examination of the economic reports, depositions and trial testimony of the following experts: Jeffrey M. Harris and Keith Leffler on behalf of the State of Washington; Adam B. Jaffe on behalf of the State of Minnesota, and John L. Solow on behalf of the State of Iowa. Jeffrey E. Harris, one of several economists who testified on behalf of the plaintiff states in support of these allegations contended that there was "...a conspiracy to mutually avoid health claims about cigarettes, to avoid admissions about the risks of smoking, not to develop innovative products..."Harris deposition, p. 208, State of Washington v. American Tobacco Co., and "Health-Care Spending Attributable to Cigarette Smoking and To Cigarette Manufacturers' Anti-Competitive Conduct: State of Washington Medicaid Program, 1997–2001", Damage Expert's Disclosure in: State of Washington v. American Tobacco, Inc., et al., Jeffrey E. Harris, MD PhD, November 3, 1997. See State of Texas v. American Tobacco Co., et al.; State of Washington v. American Tobacco Co., et al., 96-2-15056; State of Connecticut v. Phillip Morris, Inc., et al., CV-96-0072414-S; and State of Minnesota and Blue Cross & Blue Shield of Minnesota v. Phillip Morris, Inc., et al., C1-94-8565; State SEA.

[5] See Harris, op. cit.

new forms of industrial organization. He argued further that the innovations which emerge from this process "revolutionize ... the economic structure from within, incessantly destroying the old one, incessantly creating a new one" – a process he described as *"creative destruction"* – creative in the sense that it creates new value [i.e.,what contemporary economics characterizes as increased consumer welfare] and destructive in the sense that the economic returns to capital/labor producing obsolete products are lowered or eliminated entirely.

Schumpeter evidently was not of a monolithic mind on what it takes to meet his innovation test. His theory of creative destruction treats innovation in both a "broad" and a "narrow" context. For example, the innovation discussion in his 1939 work on business cycles mentions "big" developments that trigger sweeping economic effects, to wit: *"... Individual innovations imply, by virtue of their nature, a'big' step and 'big' change. A railroad through a new country, i.e., a country not yet served by railroads, as soon as it gets into working order upsets all conditions of location, all cost calculations, all production functions within its radius of influence; and hardly any 'ways of doing things' which have been optimal before remain so afterward.*[6] Later [1950] Schumpeter narrowed his model, describing innovations as: "*... These revolutions periodically reshape the existing structure of industry by introducing new methods of production – the mechanized factory, the electrified factory, chemical synthesis and the like; new commodities, such as railroad service, motorcars, electrical appliances ...*"[7]

Accordingly, in order to determine whether the changes made in cigarette products introduced by the tobacco companies satisfy either of the Schumpeter innovation and creative destruction criteria, and to assess the validity of the conspiracy allegation, the paper first addresses these questions: (1) What constitutes "*innovation*" and "*creative destruction* " in the JAS sense? (2) Can the process be measured in some systematic manner? (3) What level of credibility should be accorded to expert testimony in the antitrust cases that none of numerous innovations relative to improvements in cigarette filtration designs, tobacco content, and papers qualify as Schumpeterian innovations? (4) In an economic sense, if sales of existing products are significantly affected [adversely] by the introduction and sale of a new cigarette product, is it fair to conclude that new value [i.e., higher consumer welfare] has been created, and Schumpeterian "innovation/creative destruction" has occurred? (5) Likewise, if the sales of new cigarette products increase significantly, displacing sales of existing products, does that constitute innovation/creative destruction in the Schumpeterian sense?

2 Non-price competition in the cigarette industry

Economic studies published in the 1950's and 1960's disclose that the nature of competition in the post-World War II cigarette industry was undergoing fundamental change, with product competition assuming increasingly greater importance in

---

[6] Business cycles, p. 101.

[7] Capitalism, socialism and democracy, p. 68.

the mix of competitive strategies pursued by the tobacco companies. This literature casts doubts on the validity of the basic antitrust complaint – that the tobacco companies were engaged in a non-compete agreement to stifle innovation and non-price (product) competition. More specifically, these studies make two important disclosures: (1) that up until the 1940's, the leading cigarette brands were so similar physically that blindfold tests revealed experienced smokers could not distinguish among them, and (2) following World War II, there was a dramatic change in tobacco companies' product policies: product differentiation (non-price competition) emerged as the major competitive weapon for maintaining or increasing company market share, and traditional industry leadership yielded to the new challenges.[8]

These studies also make clear that no single company was able to dominate the cigarette market across the board. In the non-filtered products category, American Tobacco Company maintained the leadership role, while Reynolds was the leader on filter tips. However, by the 1970s, after Marlboro became the world's best-selling cigarette, Philip Morris became the industry leader. The changing character of competition among the tobacco companies led to a proliferation of brands, featuring both filtered and non-filtered products, king-size, extra-long, mentholated, and low-tar. In consequence, following standard oligopoly theory, economists consistently cited the cigarette industry as a leading example of intensive product differentiation and strong advertising-promotions campaigns to increase or protect market share.[9]

Notwithstanding this historical record, various economists testified that product competition stopped, or was severely retarded after the Hill & Knowlton meeting in 1953.[10] At the same time, a fair assessment of the purpose of the meeting of tobacco company representatives with officials of Hill & Knowlton would recognize the increasing contemporaneous public concern about smoking and health that emerged in the 1950's, which posed a challenge of such fundamental importance to all the tobacco companies to justify both an *industry response* and development of a defensible *industry* position on smoking and health. Thus, the creation of Tobacco Institute Research Committee and its successor, the Council on Tobacco Research to deal on an *industry-wide* basis with the scientific issues involved in the emerging controversy over smoking and health was a legitimate activity, consistent with the *Noerr-Pennington* doctrine [that activities designed to influence legislative, judicial, or administrative decisions or actions are exempt from U.S. antitrust laws, even if the effect of the actions is to limit competition]. Finally, the tobacco companies

---

[8] See, for example, William Nicholls, Price policies in the cigarette industry (Nashville, 1951); R. B. Tennant, The cigarette industry, in: Walter Adams, The structure of American industry, 3rd edn. (New York, 1961); M. A. Alemson, Advertising and the nature of competition in oligopoly over time, Economic Journal, vol. 80 (June 1970); Lester G. Telser, Advertising and cigarettes, Journal of Political Economy, vol. 70 (October 1962); and James L. Hamilton, The demand for cigarettes, advertising, the health scare, and the cigarette advertising ban, Review of Economics and Statistics, vol. 54 (November 1972).

[9] Ibid.

[10] On cross-examination, however, these same witnesses acknowledged that the "gentlemen's agreement" was honored more in the breach, as the tobacco companies continued to work on the development of new products with improved filters and other designs to lower tar and nicotine levels. See the testimony of Professor Adam B. Jaffe in the Minnesota case, relative to Plaintiff Exhibit 18905 [Background material on the cigarette industry client, dated December 15, 1953], Trial transcript, pp. 8170–8179.

began to factor the health concerns into their overall market strategy, and at the same time undertake appropriate competitive actions necessary to maintain their commercial viability against rivals.[11]

3 Cigarette product innovations, 1955–1995

Court records and company documents disclose that the tobacco companies did not abandon competitive product strategies, including R & D and the development of "safer" products with lower tar and nicotine levels. Notable examples were the RJR "Premier"and the Phillip Morris "Saratoga," which represent perhaps the most dramatic change in cigarette design in the history of the industry.[12] These new products, and many others introduced over the 30-year period alleged in the complaint [e.g., Lark, Merit, Parliament, Next, Salem, and Viceroy] as well as special product development projects "Ariel", "Janus" and "Batflake," were characterized by plaintiffs' economic experts as "aberrational defections" from the alleged "gentlemen's agreement" to constrain R & D efforts and not legitimate product competition.

Once again, a fair assessment of the "constrained R & D" allegation would have to acknowledge that no tobacco company, no scientific laboratory, nor any other research organization, has been able to unlock the secrets for the design and development of a completely "safe" cigarette, mainly because the basic biomedical and technical knowledge needed to produce such a product did not exist in the decades covered by the complaints, nor does it exist today. At the same time, numerous products with lower and lower tar and nicotine levels were introduced by tobacco companies prior to and during the 1954-1995 time span covered by the state complaints. Companies continuously experimented with and introduced new products with improved filters, resulting in lower tar and nicotine levels.[13]

---

[11] In his supplemental report, Professor Harris acknowledged that "...As mounting scientific evidence led to increasing consumer demand for less harmful products, the explicit agreement not to perform independent biological research became increasingly difficult to enforce. From the standpoint of the economics of imperfect competition, such a development should come as no surprise... in a cartel to stifle innovation, at least one member firm may need to retain an inventory of research techniques and findings, as well as potential new products, as insurance against a deviant firm's introduction of a risk-reducing cigarette or tobacco substitute." Damage expert's supplementry disclosure, in: State of Washington v. American Tobacco, Inc., et al.," Jeffrey E. Harris, Md, Phd, January 5, 1998, pp. 17–18.

[12] Physiologically the "Saratoga" constituted an outstanding innovation in health properties as a cigarette, as did RJR's "Premier." However, test marketing disclosed that these "non-burning" products did not have good taste, were unacceptable to the public, and consequently were not commercially-viable products. See: Operations Department Presentation to the Phillip Morris Board of Directors, October 28, 1964: Research and Development, pp. 1–2.

[13] As of 1964, it was common knowledge in the industry that all tobacco companies were engaged in some forms of chemical research, with most directed to commercial and quality purposes. Also, the companies had allied themselves with biological research laboratories. Dr. Helmut Wakeham reported to the Phillip Morris board that the company's "...Research and Development Department is working to establish a strong technological base with both defensive and offensive capabilities in the smoking and health situation. Our philosophy is not to start a war, but if war comes, we aim to fight well and to win." Operations Department Presentation Phillip Morris Board of Directors, October 28, 1964: Research and Development, p. 6.

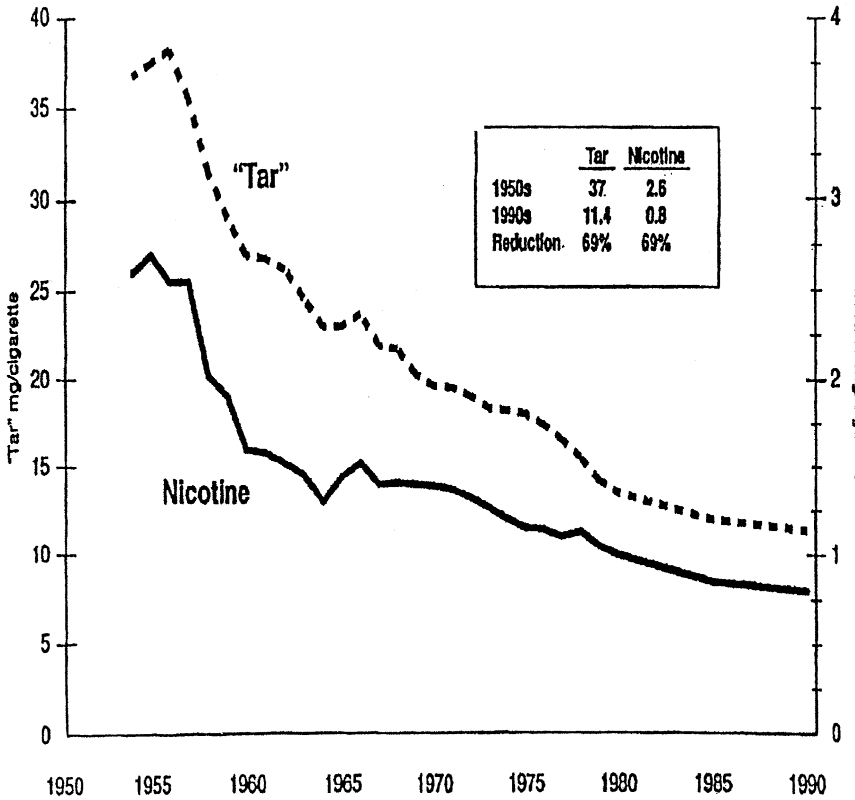


Fig. 1. "Tar" and nicotine yields of U.S. cigarettes sales weighted average basis, 1954–1990. Source: Same as Figure 2

A. Reduction of emissions: tar and nicotine levels

Tar and nicotine levels generally have been regarded by medical researchers and the U.S. Surgeon General as the prime indicators of the toxic content of cigarettes.[14] Figure 1 displays the reduction in tar and nicotine levels in cigarettes between 1954 and 1990, the time period covered by the antitrust actions. Tar content was reduced by 69% [37 mg per cigarette in 1954 to 11.4 mg in 1990]; and nicotine levels were reduced by 69% [2.6 mg per cigarette to 0.8 mg]. These reduction were made possible by new designs incorporated in cigarette products over the past 45 years, including filtration, reconstituted tobacco, paper porosity, reduced tobacco, expanded tobacco, and filter ventilation. Reductions in tar and nicotine levels were attained principally by new types of filters [e.g., (a) the 'micronite' filter in Lorillard's Kent, Newport Lights, Old Gold, and True; (b) the charcoal filter in Phillip Morris' Marlboro, Merit, Parliament, Next; (c) RJR's Winston, Salem, Premier, Eclipse and projects "Premier" and "Eclipse;" (d) Brown & Williamson's Viceroy and Kool; (e) projects "Ariel," "Janus," and "Batflake;" and L & M's Lark]

---

[14] See Surgeon General's Report of 1964.

plus research on filtration devices, nicotine analogs, and other products having "low" or no-ignition properties.[15]

A review of internal documents released by the tobacco companies, memorializing various company R & D activities over the 50-year period, discloses that while some of the new products represented largely "cosmetic" changes, many others involved significant "health-related" features, including (a) dozens of experiments with filtration materials, varieties of tobacco, and temperature reduction, with the target of reducing harmful emissions; (b) experiments with at least 10 different radical design concepts involving "non-burning" products; (c) just under 100 design and brand changes [including filter design, filtration materials, tobacco, paper type, and other features to reduce nicotine, tar and other emissions], and (d) at least 40 new products incorporating the new designs, improved filtration and lower toxic emissions (tar, nicotine and phenols). By any reasonable standard, this volume of product competition on new designs, improved filtration, and lowering of harmful emissions provide clear economic evidence of innovations containing health-related changes.

B. Public disclosure of product quality improvements regarding health effects: the "tar derby" and "Ad Wars" of the 1950's and 1960's

The antitrust complaints also allege that the conspiracy constrained advertising to smokers on health benefits associated with their respective improved products. However, Lorillard company documents disclose that within the five-year period coincident with the introduction of the new lower tar and nicotine products, more than 50 "health-related" messages were placed in advertisements announcing improved products with lower tar and nicotine content.[16] Similar ads were placed by Phillip Morris and other companies. Moreover, during the 1950's, tobacco companies engaged in an aggressive "tar derby" and intensive ad war, highlighting reductions in tar and nicotine levels. In 1960, the Federal Trade Commission halted competitive advertising based on tar and nicotine levels and health effects of smoking various brands. In 1967, the FTC began publishing the tar and nicotine levels of each brand of cigarettes, as determined by FTC labs, and all cigarette company ads were required to include the tar and nicotine content levels based on the FTC tests. Thus, the foregoing evidence undermines the contention that cigarette companies

---

[15] A chronology of significant tobacco industry product developments and related events for the 1953–1995 period is displayed in Figure 7.

[16] For example, Lorillard ads featuring "True," "Kent," and "Newport" contain the following kinds of health-related claims:

"Here's proof Kent gives greater filter protection than any other cigarette"

"True: America's no. One low tar and nicotine cigarette –

Fact is that True is lower in both tar and nicotine than 98% of all other cigarettes sold"

"Another development from Lorillard research: Kent's 'micronite' filter reduces phenol, as well as tars and nicotine in cigarette"

"The American medical association voluntarily conducted in their own laboratory a series of independent tests of filters and filter cigarettes, as reported in the journal of the American Medical Association. The tests proved that of all the filter cigarettes tested, one type was the most effective for removing tar and nicotine. This type filter is used by Kent ... and only Kent" had an agreement not to compete using informational ads about improved cigarette products, lower tar and nicotine levels, and less risky health effects.

Aside from those ads placed by the tobacco companies, additional information on lower tar and nicotine levels in cigarettes was published in the Surgeon General's Reports of 1964, 1972, and1979, the FTC annual reports on the ratings of cigarette brands, as well as articles in various medical journals, popular magazines, and newspapers.[17] In view of these different independent sources providing information to the public about cigarettes and health effects, it is difficult to accept at face value the contention that consumers were not receiving up-to-date information on improved filtered cigarettes. Moreover, given the high degree of cynicism expressed by plaintiffs about the credibility of tobacco company research and development activities, it stands to reason that reports on cigarette product innovations and health effects issued by organizations and agencies independent of the tobacco companies would be expected to have greater credibility with the public.

C. Demand side effects of new cigarette products

Figure 2 displays some selected product innovations that lowered tar and nicotine levels (shown in mg/cigarette): namely, reconstituted tobacco sheet, porous papers, expanded tobacco, and filter ventilation. Between the early 1950's and 1990, competition among the new brands containing improved filtration, lower tar and lower nicotine levels almost completely displaced older, standard cigarettes. In consequence, there was significant year-to-year variations in the market shares among the brands of major producers, notably RJR's "Winston" and "Salem" *vis a vis* Phillip Morris' "Marlboro" and "Merit", Brown & Williamson's "Kool", and Lorillard's "Newport," as shown in Figure 3.

Cigarette production and sales statistics disclose the rather dramatic response of consumers (smokers) to the introduction of new filtered cigarette products with lower health risks. Between 1954,when the first filtered cigarettes were marketed, and 1959, the market share of filter-tip products [Winston, Pall Mall, Viceroy, Kent, Marlboro and L & M] rose from a mere 3% to almost 50%, as shown in Figure 4.[18] Likewise, there was a dramatic shift of smokers to "low-tar" cigarettes (i.e., products with 15 mg tar or less, per cigarette, as measured by the FTC method) after the FTC began issuing public reports of its tar and nicotine ratings of all brands, beginning

---

[17] See, for example, Cigarette smoking and lung cancer, Consumer Reports, 1954 (February), pp. 54–92; C. W. Lieb, Can poisons in cigarettes be avoided? Readers Digest, 1953 (December), pp. 45–47; J. Monahan and L. M. Miller, The facts behind the cigarette controversy, Readers Digest, 1954 (July), pp. 1–6; E. L. Wynder, E. A. Graham and A. B. Croninger, Experimental production of carcinoma with cigarette tar, Cancer Research, 1953, pp. 855–864; R. Doll and A. B. Hill, A study of the aetiology of carcinoma of the lung, British Medical Journal, 1952, pp. 1271–1286; Smoking and health: a report of the Advisory Committee to the Surgeon General of the Public Health Service, U.S. Department of Health, Education and Welfare, Washington, D.C., 1964; and D. Hoffman and I. Hoffman, The changing cigarette, 1950–1995, Journal of Toxicology and Environmental Health, 1997, pp. 307–364.

[18] The data in Figure 4 reflect Department of Agriculture production statistics for filter-tip products, rather than sales data. However, Maxwell and Wooten sales reports confirm identical trends for production and sales. See: H.M. Wooten, Cigarette sales turn up again in '55 as filters boom, Printers Ink, December 30, 1955; and Cigarette output up 4.8% – filters up 59.8%, Printers Ink, December 28, 1956.

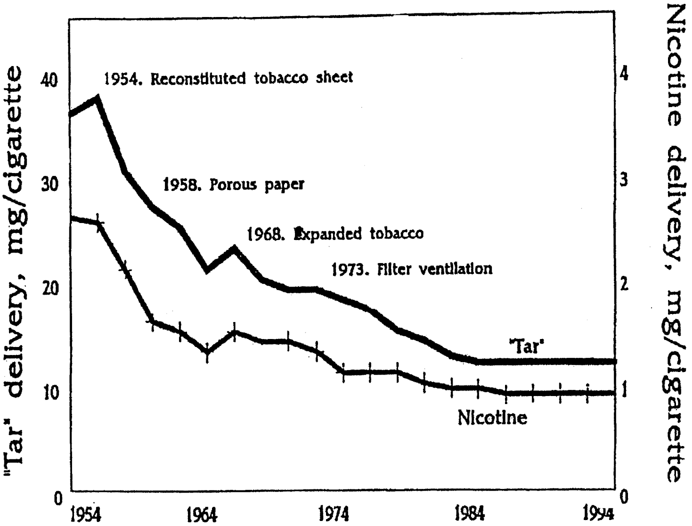


Fig. 2. 1954–1994 sales-weighted average “tar” and nicotine deliveries. Source: 1957–1987 reducing the health consequences of smoking, a report of the Surgeon General, 1989; prior to 1957 and subsequent to 1987, numbers calculated based on information similar to that used in the 1989 Surgeon General report

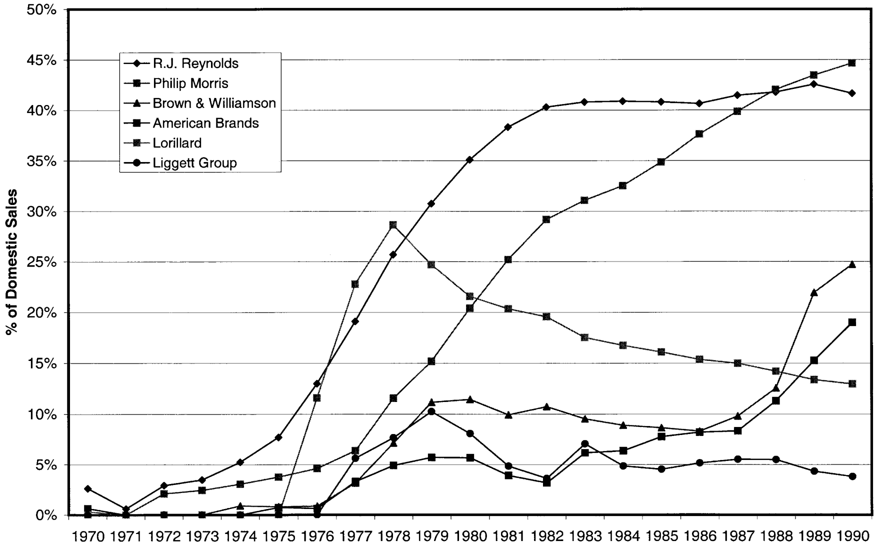


Fig. 3. Percentage of domestic sales from “light” (low tar) brands, by company 1970–1990. Source: Maxwell data produced in Minnesota case

in 1967.[19] Figure 5 shows the steady growth of low-tar cigarette products’ market share between 1967 and 1981, which amounts to a compound annual growth rate of approximately 25%. The FTC reported that by 1995, low-tar cigarettes accounted

---

[19] See: Federal Trade Commission Report to Congress for 1995 Pursuant to the Federal Cigarette Labeling and Advertising Act (Washington D.C.: Federal Trade Commission, 1997).

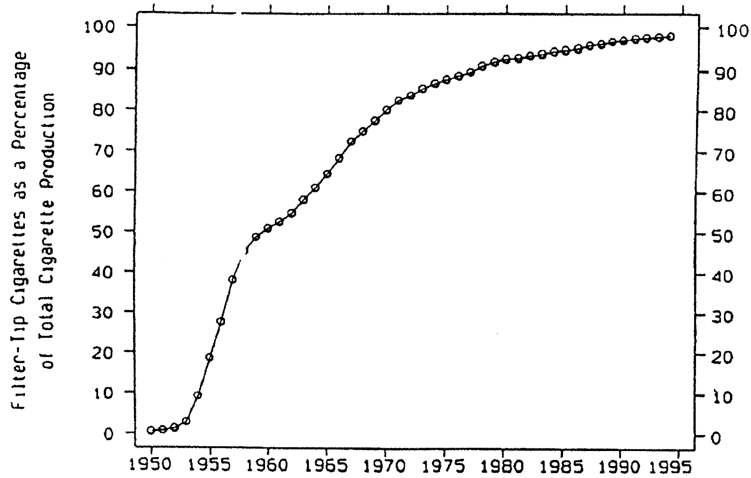


Fig. 4. Filter cigarette share of total cigarette production, 1950–1994. Source: U.S. Department of Agriculture, Economic Research Service. Archived at: http//mann77.mannlib.cornell.edu/datasets/specialty/94012/1/TAB003.WK1

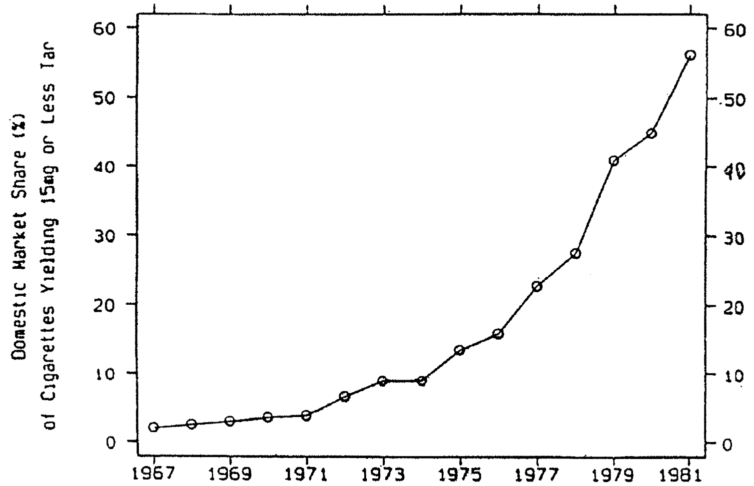


Fig. 5. U.S. market share of low-tar cigarettes, 1967–1981. Source: Federal Trade Commission Report to Congress for 1994 Pursuant to the Federal Cigarette Labeling and Advertising Act. Washington DC: Federal Trade Commission, 1996: Table 6

for approximately 73% of total cigarette sales.[20] In short, the demand side effects of new products vs. old products provide strong economic evidence of innovation.

The importance of this rather rich product innovation history was discounted in testimony by plaintiff economists, apparently because no company was able to design and produce the ultimate, ideal "safe" cigarette, which plaintiffs argue was the "public obligation" of the tobacco companies. In leveling this charge, plaintiff economists implicitly assumed (a) that the state of medical science had advanced

---

[20] Ibid.

beyond the possible adverse effects of "high" tar and nicotine levels on the lungs; (b) that scientists would have discovered and identified, with greater precision, at the time of the Surgeon General's Report of 1964 and thereafter, a deeper knowledge about the relationship between smoking and particular health effects; (c) the specific cigarette ingredients and cigarette emissions responsible for the suspected adverse health effects; and (d) the discovery and design of alternative products that minimized the risk of adverse health effects, while still providing the commercial requirements of an acceptable product from the standpoint of consumers.[21]

4 The nature of the R & D process

Economic reports and testimony presented by plaintiff economists, do not cite any references or literature dealing with innovation models, nor to a credible research and development model, which supports the assertion that tobacco companies *could have* developed "safer" cigarettes much sooner. Their analysis and conclusions are based on a theory of innovations and technological change driven principally by the level of research and development expenditures. By contrast, the literature on innovation models discloses that technological innovation is a far more complicated process than that described by plaintiff economists, and clearly not a process driven simply by the size of R & D expenditures.[22] Court records confirm that the tobacco companies did not abandon competitive product strategies, including the development of "safer" products with lower tar and nicotine levels, as previously noted.[23]

Nonetheless, plaintiff economists characterized these products and others introduced over the 30-year period alleged in the complaints [e.g., Lark, Merit, Parliament, Next, Salem, Viceroy and product development projects "Ariel", "Janus" and "Batflake"] as "aberrational defections" from the alleged "gentlemen's agreement" regarding R & D on cigarette products. Moreover, the analysis of cigarette product innovation proffered by plaintiff economic experts suffers from several other weakness:

---

[21] Of course, today's more advanced biomedical knowledge and technology (including the use of sophisticated "transgenic" biochemical manipulation and such animals as the famous "oncomouse") might permit improved testing and more precise identification of specific toxic substances in cigarette smoke, when identified as the proximate causes of lung, throat, and mouth cancers, might be eliminated in product composition and design. However promising these new avenues for cancer research, it is important to acknowledge that transgenics and the notion of transgenic pharmaceutical research were not available until the late 1980s. See Albert Rosenfeld, New breeds down on the pharm, Smithsonian, July, 1998, pp. 23–30.

[22] See, for example, the writings of Joseph Schumpeter, John Jewkes, David Sawyers, Richard Stillerman, Edwin Mansfield, Robert Solow, Edward Dennison, F. M. Scherer, D. Ross, and David Audretsch, among others, cited in footnote 24, below.

[23] Notable examples include the RJR "Premier"and the Phillip Morris "Saratoga," which represent perhaps the most dramatic change in product concept in the history of the industry. Physiologically the "Saratoga" was an outstanding cigarette, as was RJR's "Premier." However, as test marketed these products did not have good taste and consequently were unacceptable to the public. See: Operations department presentation to the Phillip Morris Board of Directors, October 28, 1964: Research and Development, pp. 1–2.

(1) The analysis glosses over the extensive time requirements involved in the complex multi-stage process that characterizes break-through bio-medical discoveries, beginning with the basic research stage to discover the keys for unlocking the doors to cellular biological mysteries, followed by the development of substitute ingredients, process, and finally the product development and the market test stages. To argue that higher levels of R & D expenditures alone would have resulted in "safer" products, sooner, without specifying with any precision at which stage(s) of the R & D process the shortfall existed, incorrectly implies that the research problem was largely a matter of product development.

(2) In contrast to the naive innovation model used by plaintiff economists, many scholars in this field note that there is a substantial random component in the discovery of an initial scientific break-through or innovation. Even though hundreds or thousands of researchers in private and government research laboratories may recognize the presence of an unsolved problem or unmet need [e.g., development of the components and design of a "safe" cigarette], only a fraction of those groups possess the technical skill and ability to devote serious effort to the basic research required and the ingenuity (and often, just sheer good luck) to develop the correct insight for solving the basic problem. The amount of resources devoted to R & D undoubtedly is an important variable in this equation, but the amount of private resources allocated will not necessarily insure success in solving difficult scientific problems, especially those dealing with human physiology, cellular biology, immunology and epidemiology.
Innovation theorists also explain that after the necessary conceptual advances have emerged and the scientific correctness of a concept [e.g., a "safe cigarette"] has been certified through test models or other demonstrations, the development process must next confront various uncertainties associated with perfecting the innovation and the development of prototypes, including:[24] (1) What is likely to be the composition, design and detailed configuration of the target product? (2) Will the prototype be commercially feasible [i.e., what are likely to be the development costs for perfecting the product, how long is the perfecting process likely to take]? (3) Will the product be acceptable to consumers in that form [e.g., the non-burning "Premier" and "Eclipse"]? (4) At what price can the new innovation be sold? (5) What is likely to be the market demand at that price?

(3) Plaintiff economists also testified that the tobacco companies should have spent much more, on the order of 10% of their sales volume, as is done in the ethical drug and pharmaceutical industry.[25] Close examination of data on industry

---

[24] See, for example, F. M. Scherer and David Ross, Industrial structure and economic performance, 3rd edn. (Houghton Mifflin, Boston, 1990), pp. 614–660; John Jewkes, David Sawers and Richard Stillerman, The sources of invention, 2nd edn. (Norton, New York, 1969; Abbott P. Usher, A history of mechanical inventions, rev. edn. (Harvard University Press, Cambridge, 1954; N. R. Hansen, Patterns of discovery (Cambridge University Press, Cambridge, 1958); Thomas S. Kuhn, The structure of scientific revolutions (University of Chicago Press, Chicago, 1962); Edwin Mansfield et al., Research and innovation in the modern corporation (Norton, New York, 1971; Edwin Mansfield, Industrial Research and Technological Innovation (Norton, New York, 1968; and Edwin Mansfield, Samuel Wagner, et al., The production and application of new industrial technology (Norton, New York, 1977).

[25] See especially the Harris, Jaffe and Leffler depositions and trial testimony in the cases cited above.

expenditures on R & D reveals that this is an arbitrary proposition. According to a study by the Federal Trade Commission, the median ratio of company-financed R & D-to-sales for all U.S. manufacturing industries was 1% or less.[26] Moreover, the central focus and business of the ethical drug and pharmaceutical companies traditionally has been grounded in both "basic" and "applied" research projects. By contrast, other industries, including those akin to the tobacco industry [e.g., the S.I.C. category, "*food and food products, beer, wine, liquor soft drinks and confections*"] are entirely different types of businesses, which historically have spent approximately 1% of sales revenues on R & D. Also, their research traditionally has been "applied" [product experimentation and development]; they typically have not engaged in bio-medical research, nor have they operated basic research laboratories such as those found in the ethical drugs and pharmaceutical industries.

(4) Plaintiff economists also contend that after the Surgeon General's Report of 1964, the tobacco companies should have re-invented themselves and directed a larger fraction of their sales revenues into basic bio-medical research. If this reasoning were valid, it would imply that beverage companies [beer, liquor and soft drinks], as well as candy, confections and snack food companies, should be allocating a substantial fraction of their sales dollars to basic medical research on alcoholism, obesity, and diabetes. Moreover, should they fail to do so would make them liable for adverse heath effect damages – a rather arbitrary, if not radical economic mandate for firms operating in a free enterprise economy. In short, economists testifying for plaintiffs contend that the tobacco companies should have reorganized themselves as bio-medical research companies and health advisory service businesses, promoting substitute cigarette products which focus groups found unacceptable. Thus, carried to its logical conclusion, this contention implies that any firm making legitimate products that meet consumer tastes and preferences has an obligation to issue health-advisory ads urging buyers to switch to other less-acceptable, or non-acceptable, foods, beverages or kindred products.

(5) Finally, one plaintiff economist apparently wants to have it both ways: He testified that the companies had a "gentleman's agreement" not to conduct in-house research for developing a safer product, and not to compete in advertising and promotion of safer products. At the same time, he contended that Phillip Morris' charcoal filtered "Saratoga" and "Next," BATco's prototype non-burning product "Ariel" [1964] and "Airbus" project [1980's], RJR's non-burning product "Premier," and its successor "Eclipse" [1987] – all developed from in-house research by Phillip Morris, RJR's biological research facilities in North Carolina, and various similar R & D projects – were isolated defections from the conspiracy, not serious R & D efforts to develop safer products. Thus, by asserting that (a) the research which developed new "safer" [lower tar and nicotine products] proves the companies had the ability to develop a safer cigarette, but (b) the products which evolved from that research really do not qualify as prod-

---

[26] See Federal Trade Commission, Statistical report: annual line of business report, 1977 (Washington, 1985) based on a study using "line of business data" for 238 manufacturing industries.

uct competition in the Schumpeterian sense, he turned his original argument on its head.[27]

5 Basic medical research and the tobacco companies

The contention that but for the alleged agreement to restrict R & D competition, the tobacco companies could have developed a safe cigarette product is based on several questionable assumptions: (1) that the companies' laboratories were sufficiently experienced in conducting bio-medical research; (2) that the cigarette companies were experienced in bio-medical research studies; (3) that they had established research laboratories dedicated to basic cellular biological research studies in-house, with an experienced staff of MDs and PhDs conducting basic biological studies; (5) that they had on-going contracts with private research laboratories covering basic research in cellular biology; or (6) they could have reorganized and re-directed their R & D into bio-medical research.

Prior to and after the 1964 Surgeon General's Report, the tobacco companies engaged in chemical research studies, principally for commercial and quality purposes,[28] focusing their R & D on the chemical composition of emissions, filtration, and product design. However, the companies did very little, if any laboratory biological research, e.g., RJR conducted some smoking exposure studies with rats, Liggett had a research contract with Arthur D. Little in the mid-1950's, and the industry's Council on Tobacco Research was familiar with Dr. Auerbach's research on "smoking dogs," sponsored by the National Cancer Institute. These in-house research efforts were more "applied" than basic, focusing on product development, e.g., re-designed filters to reduce nicotine and tar content, ventilating designs, treated and expanded tobacco, tobacco substitutes, and improving techniques for measuring nicotine and tar levels.[29]

It is speculative to argue that had the tobacco companies re-directed their research into basic bio-medical studies, their research scientists would have had any

---

[27] See testimony of Adam Jaffe in State of Minnesota, et al. v. Phillip Morris, et al. Trial transcript, pp. 8200–8226, and 8615–8695, especially pp. 8694–8695, which includes the following nonsensical assertion: "Well, I don't believe that's directly relevant, because they did research on filters, but they never made an attempt, other than the ones we have talked about, with specific products to figure out whether those products with improved filtration were in fact safer... So although that effort was, I believe, motivated by an attempt to respond to consumers' demand for safer products, I wouldn't characterize filter cigarettes as an attempt to develop a safer cigarette that really was the kind of thing that in terms of creative destruction would have been expected."

[28] The following was reported in 1964: All of the manufacturers are doing chemical research. Most of it is for commercial and quality purposes. Nevertheless, some of it is for smoking and health purposes – e.g., to enable them to alter quickly the constituents of the smoke if this should be required. Report on policy aspects of the smoking and health situation in USA (October, 1964), p. 15.

[29] The research limitation of the tobacco companies was confirmed in the letter dated December 15, 1968, from the head of the Operations Department of Phillip Morris, Dr. Helmut Wakeham, to President Goldsmith, in which he straightforwardly acknowledged that the tobacco companies did not have the expertise within their own research departments to carry out basic biological research studies on smoking and health, and urged that the research should be conducted by biological experts, not the tobacco companies. See Plaintiff Exhibit 10257 in the Minnesota case.

greater success in unlocking the unknowns than researchers and research laboratories with long-established reputations in this field. Moreover, given the highly-charged environment which emerged after the 1964 Surgeon General's report on smoking and health, it is highly doubtful that the tobacco companies could have escaped the charges leveled against them in the recent litigation [notwithstanding whatever biological research undertaken]. In short, the assertion that development of the prototype of a "safe" cigarette was largely a matter of expending more R & D dollars is simplistic and unrealistic from the standpoint of the scientific research, because it implicitly assumes that the basic epidemiological relationships regarding smoking and specific health effects had been established, and that all the tobacco companies had to do was to spend more money for product development.

On this score, in 1964, when the Surgeon General's Report was published, no scientific medical studies were extant identifying the direct connections regarding health effects and the specific ingredients that should or should not be used to produce a "safe" consumer-acceptable product [papers, filters, tobacco types, etc.] . Hence, if one seriously pursues the argument that all that was lacking to achieve the "safe product" goal were higher R & D budgets by the tobacco companies, one must identify, with some precision, the state of medical knowledge demonstrating scientifically-verified connections between cigarette products' ingredients and health effects, which might provide direction to product development efforts.[30]

6 Theoretical issues: defining the relevant product and geographic markets

The Sherman Act places the burden on plaintiffs to correctly identify the relevant product and geographic markets restrained by an alleged agreement. In addressing this issue, the economic reports and trial testimony in the cigarette litigation contain a common analytical error, namely that "cigarettes" constitute the relevant product market restrained by an alleged "gentlemen's agreement." One economist defined the relevant product market as the "United States cigarette market,"[31] and that the relevant geographic market was the United States [because of entry barriers, particularly the importance of brand names, product differentiation, and advertising expenditures]. Since the tobacco companies manufacture cigarettes, he assumed that "the cigarette market" is the product market relevant to the restraint alleged in the antitrust count. It is undisputed that the tobacco companies manufacture and distribute cigarette products. However, the Minnesota Complaint contends there was a restraint of trade not in cigarettes, but *"... in the R & D market for basic*

---

[30] The contention that production of a safe product was feasible, if only more R & D expenditures had been made by the tobacco companies, could be tested using a theory of innovation and data on the actual "success rates" of R & D expenditures in industries routinely involved in *applied research*. However, such a model implicitly assumes that *basic research* studies already had solved the puzzle regarding the particular cancer-causing agent(s) or the process triggered by specific constituents of smoke that are associated with the formation of carcinomas in human beings.

[31] State of Minnesota, et al. V. Phillip Morris, et al., Trial Volume Number 42, March 18, 1998, p. 8131.

research on smoking and health, and the discovery of the components and design of a "safer" cigarette product."[32]

The alleged conspiracy purportedly was designed to block research efforts to discover the ingredients and the development of an optimal cigarette product, risk-free of possible adverse health effects. No theoretical or empirical analysis and data were offered to support the allegation that the tobacco companies possessed and were capable of exercising *monopoly power* to restrain competition among entities engaged in biomedical and bio-technical research on smoking and health effects, and design elements for a "risk-free" product. Plaintiff economists argued that (a) because the concentration ratios for the *production of cigarettes* are high, a comparable level of concentration exists in the conduct of basic medical research and development related to smoking and health, and (b) the tobacco companies had the requisite *market or monopoly power* to restrain research activities in that market. This heroic leap from *cigarette production* to supply of *research services on smoking and health* constitutes a major flaw in plaintiffs economic analysis of the relevant product market.

A. The R & D market supplying discoveries on smoking and health effects

The economic market supplying research services on neoplasms of the respiratory system and the thoracic organs, discoveries on smoking and health effects, and other bio-tech research consists of a broad complex of institutions, laboratories and researchers, that is international in geographic scope. This market includes public and private scientists and laboratories conducting directed and non-directed research on the causes and treatments of all sorts of diseases. The laboratories are funded by public and private universities, government agencies, for-profit corporations, and non-profit entities, including the National Institutes of Health [National Cancer Institute], Sloan-Kettering, Battelle Institute, among others. Although they have engaged in applied research on chemical emissions of cigarettes, filtration, and product design, the tobacco companies are not competitors with entities conducting basic biomedical research. Indeed, no tobacco company, or any other company, laboratory or institute has a dominant share of the market supplying research studies on the causes, prevention and treatment of various forms of cancer, nor do tobacco companies individually, or as a group, have the "market power" to block or retard scientific research dealing with basic medical knowledge about smoking and health.

Therefore, the reports of plaintiff economic experts are flawed analytically because they gloss over fundamental economic differences between the cigarette product market and the market supplying biomedical and bio-technical research services on smoking and health effects. The tobacco companies have an obvious interest in research findings on diseases associated with cigarettes and health, which may provide new information that could be useful for cigarette product improvements. However, it is quite a stretch to argue that the cigarette companies as a group

---

[32] Ibid.

exercised *market or monopoly power* over the supply of biomedical and bio-tech research studies dealing with smoking and health effects. In short, (1) plaintiffs failed to correctly identify and define the relevant product and relevant geographic market which the various state complaints allege was restrained; (2) failed to present any economic analysis or data measuring the degree of market power the tobacco companies had in the correctly-defined market; and thus (3) failed to demonstrate how the tobacco companies could have restrained competition in the correctly-identified market.

The diverse market supplying basic and applied biomedical research in ethical drugs, pharmaceuticals, chemicals, new metals and alloys, plastics, and diverse consumer and producer goods: aircraft, autos, television, electronics, computers, satellites and communications does not have any artificial entry barriers, geographical or otherwise. The principal economic barrier essentially lies in scientists' ability to surmount existing frontiers of medical knowledge through basic research, which has very little to do with the tobacco companies. The tobacco companies are not players in this market, but rather are buyers, potential creators, and end-users of new materials, new processes and new components for cigarette manufacturing.

B. Why have scientific laboratories failed to develop a "safe" cigarette?

Research publications by various scientists working in the field of smoking and health effects disclose that medical technology, especially that concerned with cellular biology and the application of that knowledge to the development and production of a risk-free product was not fully known and not available to scientists inside and outside the tobacco industry during the alleged conspiracy period.[33] Moreover, it is questionable whether the tobacco companies could unilaterally retard the development of cigarette-making technology, because very limited, precise medical knowledge existed disclosing the connections between cigarette smoking and health effects during the alleged conspiracy period, and the design and specifications for manufacturing a risk-free product. Likewise, today there still does not exist any cigarette-making technology based on firm medical knowledge that can guarantee a risk-free product, beyond what already has been introduced via competitive innovations of cigarette companies, i.e., advanced designs utilizing improved filters, better papers, and other components.

1) R & D expenditures by the tobacco companies. Would a higher level of R & D expenditures by cigarette companies beginning in the 1960's and thereafter have produced a "safe" product? The answer to this question is not simply a matter of throwing large sums of money into tobacco companies' research activities *per se*. Rather, the answer is largely a function of the then current, and present state of bio-medical scientific knowledge, particularly in cellular biology, epidemiology, and pharmacology regarding the formation of carcinomas in human beings, and the attendant scientific knowledge about how to prevent those formations, including

---

[33] See citations listed in footnote 17, supra.

the composition and design of a product that provides the features smokers enjoy without risk of adverse health effects.[34]

The contention that the discovery and development of a "safe" cigarette [or, more precisely, the ingredients, components and manufacturing process for making a "safer" cigarette] was/is purely a function of higher and higher R & D expenditures thus is simplistic and questionable analytically. If such a straightforward, linear relationship existed for the development of innovations that depend upon solving biological and chemical unknowns that are associated with the formation of cancer, given the higher and higher levels of public and private R & D expenditures on various diseases over the past several decades, by now scientists also should have discovered the causes and prevention of many other serious diseases, such as AIDS, ALS (Amyotropic Lateral Sclerosis, popularly known as "Lou Gehrig disease"), DIABETES, ARTHRITIS, and ALZHEIMER's DISEASE, among others,

2) Why not a government "mandate" to develop a "safe" cigarette. For the sake of argument, assume an extreme regulatory scenario, namely that following the issuance of the 1964 Surgeon General's Report on Smoking the U.S. Congress passed a statute, effective in 1965, which *mandated* that cigarette companies produce "a safe cigarette" by the year 1970 – i.e., one that would not cause carcinomas in mice or men – which is essentially equivalent to the assumption underlying Professor Harris' analysis, which he states: "...The evidence reviewed in the previous section supports the conclusion that U.S. cigarette manufacturers had the technical capability to achieve present-day tar levels by the late 1960s..."[35]

Using regression analysis, Harris determined that the tobacco companies actually achieved an average rate of reduction in tar delivery per cigarette (as measured by the FTC method) between 1955 and 1995 of 3.4 mg per year, and then concluded that the average rate of reduction should have been 7.0 mg per year. Figure 6 displays the slope of the regression line Harris fitted to the *actual* tar levels achieved by manufacturers [-0.034]. Harris contends "... that innovation in the tobacco industry was no more than one-half as rapid as it could have been ..." and the tobacco companies should have achieved lower tar levels twice as fast, but for the alleged agreement to restrain innovation [namely, a slope $= -0.07$, as displayed in Fig. 6].[36]

The Harris slope represents one plausible alternative innovation scenario, given his assumption that the companies could have reduced tar levels at twice the actual rate "... had a competitive market prevailed."[37] The product development history reviewed earlier indicated a fairly brisk competition among tobacco companies to

---

[34] In short, the research process that will lead to the design and production of a "safe" cigarette depends crucially upon (a) the state of scientific knowledge, (b) the technical feasibility of developing the desired materials, design, and manufacturing process for the target product, and (c) the availability of supporting innovations of greater or lesser magnitude to make the basic innovation commercially viable.

[35] See Jaffe Supplementary disclosure statement (January 5, 1998), p. 36.

[36] In this connection, Harris assumes a linear regression correctly fits the data displayed in Figure 6. However, a "goodness-of-fit" test indicates that there is good reason to believe the correct fit is non-linear (if all the data are used, including figures for 1985–1995), in which case the slope is not nearly as steep as Harris reports.

[37] Ibid.

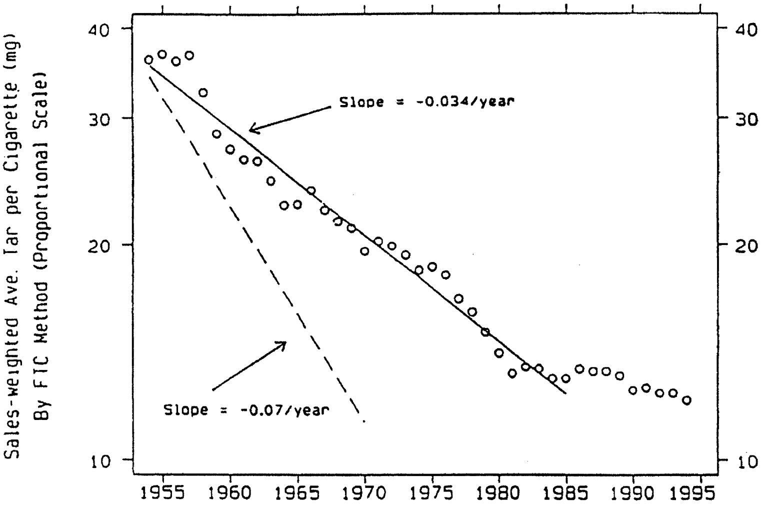


Fig. 6. Average sales-weighted tar delivery per cigarette by the FTC method, United States, 1954–1994

reduce tar levels, which raises serious doubts about the validity of Harris' contention that innovation should have been "at least twice as rapid."[38] Such a rate might have been attained under a government-approved, cooperative industry program, but it is speculative to argue that mandated legislation could have guaranteed either a more rapid rate of reduction in tar levels or the development of a "safe" cigarette product by Harris' arbitrary date of 1970.

Thus, the argument – that had the cigarette companies demonstrated the "will," or had they been "mandated" by federal statute – a "safe" cigarette would be available today, makes sense only if the necessary scientific research had been completed, as was true in the case of various other U.S. "mandates" – e.g., on the auto companies [to develop "safer" bumpers, passenger restraints, and cleaner emissions], on utilities [to add scrubbers to stacks for cleaner emissions from coal-burning plants], and on sewage treatment facilities [to develop systems for cleaner effluents].[39] In the case of autos, utilities, and sewage treatment plants, however, the necessary scientific knowledge for developing those target products was extant, i.e., the mandates were feasible because requisite underlying electrical and mechanical engineering know-how existed, was available, or clearly developable at a predictable cost.. In fact, it was cost, not availability of technical knowledge, that was responsible for the delay in developing these innovations. Companies simply did not have any incentive to internalize these economic externalities until the government mandates

---

[38] Although concern about health effects of smoking is not limited to the U.S., it is instructive to note that the improvements in cigarette quality has largely been made by U.S. tobacco companies.

[39] It should be noted that if, under a mandated scenario, individual tobacco firms believed that any health-related discovery would unavoidably be made available to competitors, thereby rendering their own R & D unprofitable, they understandably would have refrained from reaching "ideal" levels of R & D, even without a conspiracy. Although this outcome would not be ideal from a social perspective, such behavior is not uncommon in industry, certainly not illegal under the antitrust laws.

were promulgated, after which some of the development and production costs were shifted to consumers.

By contrast, developing a “safe” cigarette represents a world of difference from the above examples, principally because the scientific knowledge and technology regarding how various cigarette ingredients and components affect the development of carcinomas in rats and human beings was not and still is not known. Ever since the publication of the Surgeon General’s Report of 1964 and thereafter, we know that there is a higher incidence of carcinomas in persons who smoke versus non-smokers, but no one to date has been able to identify *how* they develop, nor the particular ingredient(s) responsible, which is the basic scientific knowledge required to produce a completely “safe” product.

Even if one accepts the argument that the tobacco companies did not have an interest in discovering how cigarette smoking and cancer are related, the companies surely would have commercial interest in a “safe” product innovation that would make existing products obsolete, whether developed by current or potential new competitors. Economic analysis and Schumpeterian logic instructs us that if the scientific knowledge were extant and there for the R & D expenditure, the innovations would have emerged in the U.S. or elsewhere, and exported to the U.S.. The economic returns to such an innovation would be huge, because the latent demand is large and patently inelastic. Moreover, given the imperfect nature of the U.S. patent system, once the new design, ingredients, components, and manufacturing process were patented, there would be “leakage” of information, and others would use the patent disclosure to immediately undertake further research to develop products providing similar or even improved results.

One corollary to this proposition is that tobacco companies would be eager to compete for head-start advantages, through outright purchase of patent rights, acquisition of licenses to produce the ultimate completely “safe” cigarette that would make most existing cigarette products obsolete. However, no evidence was presented by during the trial demonstrating that after the release of the 1964 Surgeon General’s Report the tobacco companies had the market power to prevent the development or introduction of such an innovation. This follows from the fact that U.S. tobacco companies do not have a monopoly on the supply of scientists doing research in this field, nor of basic research on health effects of smoking. Numerous independent researchers in private and public laboratories around the world are engaged continually in research on various diseases, including cancer. It is silly to argue that the tobacco companies individually, or as a group, are able to stifle this activity, even if they wished to engage in some sort of “gentlemen’s agreement” not to compete on product development. The design of a “safe” cigarette will be discovered and developed when the basic research issues have been solved, an innovation process over which the tobacco companies have very little, if any, control.

Conclusion

The drive by tobacco companies to develop and market new products would appear to be incontrovertible: “new” [i.e., post-1953] products virtually completely displaced “old” [pre-1953] products. The data displayed in Figure 3 (sales growth of new brands), in Figure 4 (the dramatic shift of smokers to filtered products), and the spectacular growth in the sales of low-tar products (Fig. 5) on their face provide a strong market test refutation of the contention in the *Minnesota* case that the new products marketed by the tobacco companies during the 1954–1995 period really did not meet the "innovation/creative destruction" standard promulgated by Schumpeter.[40] On their face, therefore, these data on shifts in cigarette product sales would appear to provide ample evidence to meet the Schumpeterian test for innovation and creative destruction, namely the continual development and introduction of improved cigarette product design and the inter-firm non-price (quality) competition which ensued over several decades.

Appendix A

Chronology of significant tobacco industry product developments, 1953–1995

Date	product development, innovation or related event
1953	Sloan-Kettering report on carcinogenicity of cigarette tars. American Cancer Society report issued on dangers of smoking. Reader's Digest publishes article on dangers of smoking.
1954	Tobacco Industry Research Committee [”TIRC”] is formed. 1st successful filtered cigarette [RJR's ”Winston”] is marketed. Competitors response: American's "Pall Mall," B & W's "Viceroy," Lorillard's "Kent," Phillip Morris' "Marlboro," and L & M's "L&M."
1957–1960	Cigarette product competition [”tar derby" and advertising "war"] leads to one-third reduction in average tar levels of cigarettes.
1958	The Tobacco Institute formed by cigarette manufacturers.
1960	FTC tar/nicotine regulations promulgated; tobacco companies cease advertisements containing tar & nicotine levels.
1960	Batco launches project to develop a "smokeless" cigarette [project "Ariel"].
1964	1st Report of U.S. Surgeon General states smoking is "habituation," not "addiction," and smoking is causally-related to lung cancer.
1966	Surgeon General reports that low tar and low nicotine cigarettes provide benefits to smokers by reducing the probability of disease.

---

[40] See this colloquy: "Q. ...The defendants did research on the development of filter cigarettes, right? A. Well, I don't believe that's directly relevant, because they did research on filters, but they never made an attempt other than the ones we talked about with specific products to figure out whether those products with improved filtration were in fact safer products... I wouldn't characterize filter cigarettes as an attempt to develop a safer cigarette that really was the kind of thing that in terms of creative destruction would have been expected." Trial Transcript, State of Minnesota, et al. V. Phillip Morris, et al., pp. 8694–8695.

1967	FTC measures & publishes tar and nicotine levels for all cigarette brands.
1968–1979	L&M develops a smokeless cigarette from "Project XA"/decides marketing product is not commercially feasible.
1970	Tobacco companies voluntarily agree to include tar/nicotine levels in ads.
	RJR terminates "Mouse House" research.
1970–1979	B&W launches project "Janus" to isolate & remove harmful substances in tobacco.
1972	Surgeon General Report states smoking is "associated" with a list of diseases.
1979	Surgeon General Report reports: nicotine is addictive; smoking is associated with substance-abuse dependency; and smoking reduces life expectancy.
1982	Merrill Dow introduces nicotine gum.
1980–1981	RJR experiments with "smokeless" cigarette "Premier."
1986	Average tar & nicotine levels reduced by 69% from 1954 levels.
1987–1988	RJR test markets "Premier"/FTC regulations prohibit advertising as "safer" product.
1991	Phillip Morris introduces "nicotine-free products [< 1mg]: "Next," "Merit Free," & Benson & Hedges "De Nic"/ nicotine "patch" marketed.
1992	Phillip Morris develops safer cigarette ["Table"].
1994	Brown & Williamson internal documents leaked to public.

Risk, variety and volatility: growth, innovation and stock prices in early industry evolution[⋆]

Mariana Mazzucato

Economics Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
(e-mail: m.mazzucato@open.ac.uk)

Abstract. The paper studies the patterns of volatility in firm growth rates and stock prices during the early phase of the life-cycle of an old economy industry, the US automobile industry from 1900–1930, and a new economy industry, the US PC industry from 1974–2000. In both industries, firm growth rates are more volatile in the period in which innovation is the most "radical". This is also the period in which stock prices are more volatile. The comparison sheds light on the co-evolution of industrial and financial volatility and the relationship between this co-evolution and mechanisms of Schumpetarian creative destruction. Results provide insight into the debate on whether the statistical behavior of firm growth rates is well represented by Gibrat's Law.

Keywords: Growth rates – Stock prices – Innovation – Volatility

JEL Classification: L11, 030, G12

1 Introduction

The paper studies patterns of volatility in firm-specific growth rates and stock prices during the early phase of the life-cycle of an old economy industry, the US automobile industry from 1900–1930, and a new economy industry, the US PC industry from 1974–2000. The firm level analysis builds on and supports the industry level analysis found in Mazzucato (2002). Results shed light on the co-evolution of growth rate volatility and stock price volatility and the relationship between this co-evolution and mechanisms of Schumpetarian creative destruction.

[⋆] I thank Massimiliano Tancioni for his excellent research assistance. Support from the following grants is much appreciated: European Commission Key Action "Improving the socio-economic knowledge base"contract HPSE-CT-2002-00146, and the Open University RDF Grant contract no. 793.

In so doing, it adds new insights to the literature on Gibrat's law and firm growth (i.e. whether firm growth rates follow a random walk). The question explored is not whether Gibrat's law holds or not but under which conditions it is most likely to hold.

After reviewing data sources in Section 2, Section 3 studies industrial volatility in both industries by focusing on the statistical properties of firm-level growth rates. Both absolute and *relative growth* rates are explored, with the latter being more central to an evolutionary analysis of change (Dosi and Nelson, 1994). The null hypothesis on the unit root tests is that growth rates follow a random walk. In Section 4, statistical tests are used to study the volatility of firm-level stock prices and dividends. Results indicate that, for most firms, stock prices are most volatile during the same decades in which relative growth rates (e.g. market shares) are most volatile. In Section 5, innovation dynamics in the two industries are used to interpret these patterns of volatility: in both industries the decades in which relative growth rates and stock prices were the most volatile were the same decades in which innovation was the most radical and competence-destroying (Tushman and Anderson, 1986). Building on this result, Section 6 uses panel data analysis to test whether in the early phase of the industry life-cycle, changes in firm stock prices are related to variables describing industrial instability more so than in the mature phase. The firm-level results confirm the industry level results in Mazzucato (2002) where changes in industry structure (e.g. number of firms, entry/exit rates, concentration, etc.) and innovation are related to changes in stock price volatility.

2 Data

The study focuses on the US market for automobiles and personal computers (including both domestic and foreign producers). The firm-level and industry-level data is annual. Sales are measured in terms of annual units of automobiles (cars and trucks) and personal computers (all microcomputers:desktops and notebooks) produced. In both industries, units produced follow the same general qualitative dynamic as that of net sales in dollars but is preferred due to its greater precision (sales figures are affected by idiosyncratic accounting items).

Automobiles. Individual firm units and total industry units from 1904–1999 were collected from annual editions of *Wards Automotive Yearbooks* (first editions with data starting in 1904 were published in 1924). Although firm-level units were collected for only 8 domestic firms and 5 foreign firms (the first foreign firms entered in 1965), the *total* industry sales include the units shipped by all existing firms (e.g. in 1909 that includes the output of 271 firms). Firm stock prices and dividends figures were collected from annual editions of *Moody's Industrial Manual*. Industry-specific per share data was collected from the *Standard and Poor's Analyst*

Handbook.[1] However, since all financial data, except stock prices, only goes back to 1946 for the automobile industry, the data for the pre-war period was aggregated from the firm-specific data gathered from Moody's.[2] The first year the auto firms went public was 1918, with Ford following only in 1956 (see endnotes for exact dates for each firm).

PCs. Annual firm-level data on the total number of personal computers produced from 1974–2000 was obtained from the International Data Corporation (IDC), a market research firm in Framingham, Massachusetts. Although this database is very detailed (including brand, form factor, processor speed, region and customer segment), for the purpose of this study firm-level units were aggregated across different models and brands produced by a given firm. Firm stock prices and dividends were obtained from *Compustat.* Industry-level financial variables were obtained, as for the post-war auto industry, from the Standard and Poor's *Analyst's Handbook* (2000). The firms which define this index are all included in the firm-level analysis, except for Silicon Graphics and Sun Microsystems (the only two firms in the S&P index for the computer industry which don't produce personal computers).[3] Hedonic prices were obtained from the Bureau of Economic Analysis (BEA). An index measuring quality improvements, used here as a proxy for innovative activity, was obtained from Filson (2001).

The data is analyzed during the first 30 years of each industry's history. This represents the "early" phase in the industry life-cycle, i.e. the phase that encompasses the introductory phase when the product first emerges and the initial growth phase (Gort and Klepper, 1982). The data for the "mature" phase of the automobile industry's life-cycle is also analyzed to gather some insight into what might lie ahead for the PC industry. Although the depression years, 1929–1933, are omitted from the sample, results for relative growth rates are not altered when these years are included. To control for movements in the general market, analysis is also done on the units data divided by GDP and on the financial data divided by the S&P500 market aggregate (e.g. GM stock price divided by the S&P500 stock price).

---

[1] The firms used to create the S&P index for automobiles are (dates in parentheses are the beginning and end dates): Chrysler (12-18-25), Ford Motor (8-29-56), General Motors (1-2-18), American Motors (5-5-54 to 8-5-87), Auburn Automobile (12-31-25 to 5-4-38), Chandler-Cleveland (1-2-18 to 12-28-25), Hudson Motor Car (12-31-25 to 4-28-54), Hupp Motor Car (1-2-18 to 1-17-40), Nash-Kelvinator Corp (12-31-25 to 4-28-54), Packard Motor Car (1-7-20 to 9-29-54), Pierce-Arrow (1-2-18 to 12-28-25), Reo Motor Car (12-31-25 to 1-17-40), Studebaker Corp. (10-6-54 to 4-22-64), White Motor (1-2-18 to 11-2-32), and Willy's Overland (1-2-18 to 3-29-33).

[2] Since in the post-war period, the results were not sensitive to whether we used the aggregate industry data (provided by S&P) or the average of the firm-specific one, this suggests that the pre-war data is robust.

[3] The computer industry was first labelled by S&P as *Computer Systems* and then in 1996 changed to *Computer Hardware*. Firms included in this index are:
Apple Computer (4-11-84), COMPAQ Computer (2-4-88), Dell Computer (9-5-96), Gateway, Inc. (4-24-98), Hewlett-Packard (6-4-95), IBM (1-12-19), Silicon Graphics (1-17-95), and Sun Microsystems (8-19-92).

3 Firm growth rates

The literature on firm growth contains abundant empirical evidence, across different industries and countries, that – at least for the short term and for the case of new firms – the growth rate of firms is principally affected by stochastic shocks (Gibrat, 1931; Ijiri and Simon, 1977 and for an excellent review of the literature see Marsili, 2002). For this reason, *Gibrat's Law of Proportionate Growth,* which states that firm growth rates are i.i.d. random variables independent of firm size (constant returns to scale) is often used as a benchmark model to analyze firm growth patterns. According to Gibrat's Law, the size of a firm at time $t + 1$ is taken to be a function of its size at time t subject to random variation. Taking xi to denote firm size, the size of firm i is governed by the following equation:

$$x_i(t) = \alpha + \beta_i x_i(t-1) + \varepsilon_i(t) \tag{1}$$

where $x_i(t)$ is the log size of firm i at time t, and α is a growth component common to all firms. Gibrat's law assumes that ε is an i.i.d random variable and for all i, and $\beta_i = 1$, that is, the expected rate of growth is independent of the present size. The principal result in such models is that although firms might begin ex-ante with equal growth prospects, differences in initial conditions and the presence of random events cause firms to quickly diverge in size and market shares, causing a skewed size-distribution (log-normal) to emerge. The empirical evidence on Gibrat's Law is mixed, with some studies showing that growth rates and their variance tend to fall with size and age (Hall, 1987; Evans, 1987) and others which find evidence for the law by focusing on the large percentage of exits. Geroski and Machin (1993) claim that Gibrat's Law is better suited to describe the growth process of relatively large firms.[4]

Gibrat's law is difficult to reconcile with those studies which find that there are *persistent* and *cumulative* patterns in firm profits and innovation (Bottazzi et al., 2002; Mueller, 1990). While the random story suggests that there is some kind of reversion to the mean, the persistence story suggests that there are strong positive feedback mechanisms at work. Positive feedback may arise from learning by doing in which firm growth depends on cumulative output, or from more complex reasons related to the dynamics of absorptive capacity (Cohen and Levinthal, 1990) and the way that firm-specific capabilities develop (Teece, Pisano and Shuen, 1990). Geroski and Mazzucato (2002) find that Gibrat's law is better suited to describe firm growth during the early phase of the industry life-cycle when technological opportunity is greater and concentration is lower. This last finding is returned to in the discussion of the results below.

Gibrat's law can be tested for in various ways. As is well known, testing for the "beta" coefficient in Equation (1) contains a bias towards accepting the random walk hypothesis (Geroski and Mazzucato, 2002). Testing for a unit root (for each firm),

---

[4] Recent empirical studies have suggested that the best simple generalization is that, on average, smaller firms have a lower probability of survival, but those that survive grow proportionately faster than large firms. The real problem lies not in characterizing what happens on average but the fact that a wide range of different patterns occurs across different markets, so that it is difficult to generalize as to the normal size-growth relationship or the 'typical' shape of the distribution.

using different variations of the Dickey-Fuller method can provide an alternative way to test for the random walk hypothesis (Dickey and Fuller, 1979). This is the method chosen here. If a time-series has a unit root then it shows a *systematic* pattern in its movement (a stochastic trend), but such movement isn't predictable because it is the effect of random shocks on a long-memory process. Dickey-Fuller testing strategies are generally addressed for discriminating between the trend-stationary processes and the difference-stationary processes. They can be modeled in many ways, depending on the consideration of drifts, trends, and the number of augments of the lagged dependent variable (ADF: Augmented D-F tests). It has been observed that the size and power properties of the ADF test are sensitive to the number of lags (Agiakoglou and Newbold, 1992). There are two main approaches: (1) the general-to-specific technique (Hall, 1994) which starts with a large number of lagged terms that are iteratively reduced until a significant statistic is encountered, and (2) the model selection information criteria (Akaike Information Criterion – AIC and the Schwarz Bayesian Information Criterion – SBC). Information Criterion methods based on small values can, in the presence of MA errors, result in size distortions (Maddala, 1998). General to specific approaches tend to define higher dimensions, but this can produce important losses of power, especially in small sample sizes. The large (and increasing) number of unit root tests is a consequence of the fact that there is no uniformly powerful test of the unit root hypothesis. Given the moderate sample dimensions in the data used here, the Schwarz Bayesian Information Criterion (SBC) is used to handle the tradeoff between fit and parsimony. In the results that follow, both absolute and relative growth rates are studied. The latter are more relevant for an evolutionary understanding of inter-firm competition.

Results. Tables 1–4 contain the standard deviations, means and unit root tests of firm-level and industry-level units and market shares, where the latter are taken as proxies of *relative* growth rates. Results for firm level growth rates divided by industry growth rates were qualitatively the same as those for market shares. To maximize the degrees of freedom, the unit root tests are done on the entire 30 years of the "early" phase of industry history (or the maximum number of years for which the firm existed).The descriptive statistics are done on the three individual decades as well as the entire pre-war and post-war periods. When unit root tests identified a trend, the descriptive statistics are performed on the detrended data. But since no qualitative difference was found in the different periods between the detrended and the non-detrended series (after the logs and differences were taken), statistics only for the non-detrended data are reported. Data is presented for the top 8 firms in the automobile industry: GM, Ford, Chrysler, American Motors, Studebaker, Packard, Hudson, and Nash (with unit root tests also presented for the foreign firms as a group), and the top 10 firms in the PC industry: Apple, Compaq, Dell, Everex, Gateway, Hewlett-Packard, IBM, NEC, Toshiba, and Unisys.

Automobiles. Table 1 indicates that units contain a unit root only in the pre-war period, i.e. they follow an I(1) process in the pre-war and an I(0) process in the post-war period. This is true for all firms except for Ford which does not contain a unit root in either period, most likely due to its stable dominance of the industry in the pre-war period (and the general stability of the industry in the post-war period). The

Table 1. Analysis of the statistical properties of the series (processes) of units produced by US and foreign firms: DF and ADF tests

Firm	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Chrysler	1925–1941	SBC	ADF(1)+drift+t	−2.6184	−3.7612	I(1)
	1948–1998	SBC	DF+drift	−3.1572	−2.919	I(0)
Ford	1904–1941	SBC	DF+drift	−4.0172	−2.9446	I(0)
	1925–1941	SBC	ADF(1)+drift	−3.3364	−3.0522	I(0)
	1948–1998	SBC	DF+drift	−3.3243	−2.919	I(0)
GM	1909–1941	SBC	DF+drift	−1.6327	−2.9591	I(1)
	1925–1941	SBC	DF+drift	−2.0393	−3.0522	I(1)
	1948–1998	SBC	DF+drift	−3.1688	−2.919	I(0)
Hudson	1910–1926	SBC	ADF(1)+drift	−3.6729	−3.7612	I(1)
	–	–	–	–	–	
Nash	1917–1926	SBC	DF+drift+t	−2.1548	−4.1961	I(1)
	–	–	–	–	–	
Packard	1904–1926	SBC	DF+drift+t	−2.4551	−3.6454	I(1)
	–	–	–	–	–	
Reo	1905–1928	SBC	DF+drift+t	−2.3108	−3.6331	I(1)
	–	–	–	–	–	
Studebacker	1911–1941	SBC	DF+drift	−1.8746	−2.9665	I(1)
	1925–1941	SBC	DF+drift	−1.5779	−3.0522	I(1)
	–	–	–	–	–	
American	–	–	–	–	–	–
	–	–	–	–	–	
	1948–1985	SBC	ADF(1)+drift	−2.5329	−2.94	I(1)
Total units (US)	1899–1941	SBC	ADF(1)+drift	−2.3311	−2.9339	I(1)
	1925–1941	SBC	DF+drift	−1.9753	−3.0522	I(1)
	1948–1998	SBC	DF+drift	−4.2927	−2.919	I(0)
Honda	1971–1998	SBC	DF+drift	−6.7257	−2.9798	I(0)
Mazda	1985–1998	SBC	DF+drift	−1.2576	−3.1485	I(1)
Mitsubishi	1985–1998	SBC	DF+drift	−2.6298	−3.1485	I(1)
Nissan	1965–1998	SBC	DF+drift	−4.9077	−2.9558	I(0)
Toyota	1966–1998	SBC	DF+drift+t	−6.174	−2.9591	I(0)
Volkswagen	1965–1998	SBC	DF+drift	−1.4288	−2.9558	I(0)

DF = Dickey-Fuller Test, drift stands for intercept, t for trend
ADF = Augmented Dickey-Fuller Test, drift stands for intercept, t for trend
SBC = Schwarz Bayesian Criterion (based on ML)
I(1) = Integrated process of order 1, or Difference Stationary (DS)
I(0) = Stationary process, or Trend Stationary (TS)

same result holds for relative growth rates (market shares) and whether or not auto units were divided by GDP. As regards the different decades, Table 2 indicates that firm-level and industry-level units were most volatile in the period 1918–1941, with most volatility occurring between 1918–1928. Studebaker was the only exception, with its most volatile period after World War 2.

Personal Computers. Table 3 indicates that the growth rates of all firms, except Apple and Dell, contain a unit root. The descriptive statistics in Table 4 indicate that all firms experienced higher mean growth in the most recent decade (1990–2000) but more volatile growth (standard deviation) in the first decade (1970–1980) or the second decade (1980–1990). Relative growth, unlike absolute growth, was the most

Table 2. Descriptive statistics of units (logs of differences) in the US auto industry

		1904–18	1918–28	1918–41	1948–00	1948–70	1970–00	1970–80	1980–90	1990–00
gm	st. dev	0.1650	0.1745	0.1548	0.0798	0.0933	0.0732	0.1114	0.0502	0.0229
	mean	0.1001	0.0597	0.0579	0.0052	0.0153	−0.0088	−0.0029	−0.0158	−0.0162
ford	st. dev	0.2382	0.2734	0.2307	0.0837	0.1045	0.0572	0.0735	0.0683	0.0359
	mean	0.1722	0.0011	0.0173	0.0072	0.0283	−0.0103	−0.0277	−0.0054	−0.0161
chrysler	st. dev		0.1035	0.1415	0.0897	0.1149	0.0615	0.0717	0.0708	0.0536
	mean		0.1019	0.0654	0.0006	0.0130	−0.0089	−0.0210	−0.0090	−0.0155
amc	st. dev			0.2039	0.1067	0.1136	0.0913	0.0673	0.1288	
	mean			0.0971	0.0021	0.0182	−0.0192	−0.0135	−0.0197	
studeb	st. dev		0.1565	0.1639	0.3023	0.3023				
	mean		0.0409	0.0470	−0.1038	−0.1038				
packard	st. dev	0.2096	0.1936							
	mean	0.1460	0.0262							
hudson	st. dev		0.2549							
	mean		0.1178							
nash	st. dev		0.1784							
	mean		0.1169							
industry	st. dev		0.1569	0.1500	0.0638	0.0759	0.0523	0.0768	0.0428	0.0231
	mean		0.0305	0.0378	0.0070	0.0172	−0.0034	−0.0050	−0.0054	−0.0088

Table 3. DF-ADF tests for logs of units produced in the US PC industry

Firm	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Apple	1979–2000	SBC	DF+drift	−7.127	−3.004	I(0)
Hewlett-Pack	1979–2000	SBC	ADF(1)+drift+t	−2.725	−3.633	I(1)
IBM	1979–2000	SBC	ADF(1)+drift	−2.101	−3.004	I(1)
NCR	1979–2000	SBC	DF+drift+t	−0.16	−3.659	I(1)
Unisys	1979–2000	SBC	ADF(1)+drift	−2.628	−3.012	I(1)
Commodore	1979–1994	SBC	DF+drift+t	0.169	−3.735	I(1)
Compaq	1985–2000	SBC	DF+drift+t	−2.045	−3.735	I(1)
Dell	1987–2000	SBC	ADF(1)+drift+t	−6.435	−3.792	I(0)
Gateway	1986–2000	SBC	ADF(1)+drift	−1.586	−3.082	I(1)
Toshiba	1983–2000	SBC	DF+drift	−1.747	−3.04	I(1)
Wang	1979–1993	SBC	DF+drift	−1.228	−3.082	I(1)
Wyse	1986–1994	SBC	DF+drift+t	−1.686	−4.081	I(1)
All firms	1979–2000	SBC	DF+drift	−3.358	−3.004	I(0)

DF = Dickey-Fuller Test, drift stands for intercept, t for trend
ADF = Augmented Dickey-Fuller Test, drift stands for intercept, t for trend
SBC = Schwarz Bayesian Criterion (based on ML)
I(1) = Integrated process of order 1, or Difference Stationary (DS)
I(0) = Stationary process, or Trend Stationary (TS)

volatile in the last decade. The industry results diverge from the firm-level results: total units experienced both higher average growth and more volatile growth in the first decade 1970–1980. This is most likely because in some years the aggregation may dampen inter-firm heterogeneity and volatility while in other years it may enhance it, depending on how the different series interact. Nevertheless, it appears clear in both the firm-level and industry-level data that the last decade was the least volatile in terms of absolute growth but the most volatile in terms of relative growth and market shares. Below we will see that this is because although entry rates were

Table 4. Descriptive statistics of units (logs of differences) in the US PC industry

		1970–00	1970–80	1980–90	1990–00
apple	st. dev	0.7958	0.6847	0.4118	0.1132
	mean	5.7186	4.1964	5.7176	6.2251
compaq	st. dev	0.6717		0.3192	0.4843
	mean	5.9325		5.3333	6.3406
dell	st. dev	0.8999		0.4933	0.6501
	mean	5.6426		4.7283	6.0815
everex	st. dev	0.2038		0.2768	0.1932
	mean	4.7409		4.6979	4.7762
gateway	st. dev	1.5397		0.9631	0.5243
	mean	4.9947		3.3857	6.0170
hpackard	st. dev	1.1976	0.4852	0.3910	0.7145
	mean	4.7415	3.2614	4.8468	5.8018
ibm	st. dev	1.3390	0.4552	0.8164	0.1464
	mean	5.3104	3.3484	5.6838	6.3037
nec	st. dev	0.3437			0.3437
	mean	6.1557			6.1557
toshiba	st. dev	1.1964		1.0799	0.3587
	mean	4.9549		4.0701	5.7926
unisys	st. dev	0.6569	0.3242	0.4711	0.6040
	mean	4.1503	3.3528	4.3014	4.4822
industry	st. dev	0.1758	0.2062	0.1884	0.0357
	mean	0.1504	0.2432	0.1450	0.0646

high in the first decade (leading to high absolute growth), the new entrants were only able to seriously compete with IBM in the third decade (leading to a shake-up of market shares).

Hence, Tables 1–4 indicate that the first 30 years in both the auto and PC industry were characterized by volatile growth and that Gibrat's Law is not a bad description of growth in the *early* years of industry evolution. The auto industry experienced the most volatile growth, in both absolute and relative terms, in the first decade of its existence, while the PC industry experienced the most volatile absolute growth in the first decade (when entry rates were highest) and the most volatile relative growth in the third decade. After looking at the statistical properties of stock prices, Section 3 will interpret these results in terms of innovation dynamics in both industries.

4 Stock price volatility

This section studies the co-evolution of the volatility of growth rates and the volatility of stock prices. Research into this question can benefit by linking two literatures that do not often talk to each other: the industry dynamics literature that looks at factors that determine industrial *instability* (Hymer and Pashigian, 1962; Gort and Klepper, 1982), and the finance literature that looks at factors that determine stock price *volatility* (Shiller, 1989; Braun et al., 1995, Campbell et al., 2000). The connection between the volatility of growth rates and stock prices lies in how "risk" and uncertainty evolve over the industry life-cycle – i.e. the dynamics of a time-varying (industry) risk premium – and how this risk is both a cause and an effect of the mechanisms that create differences and inequality between firms. It is this non-linearity that led the pioneer of the economics of risk to state: "*Without uncertainty it is doubtful whether intelligence itself would exist.*" (Knight, 1921, p. 268).

The volatility of economic variables, like firm growth rates and stock prices, is often attributed to "random" behavior. This was already seen above for the case of growth rates through the discussion of Gibrat's Law. In the case of stock prices, their volatility has been attributed by some authors, such as Shiller (1989), to the existence of animal spirits, herd effects, and bandwagon effects. By connecting both types of volatility to innovation dynamics (Sect. 5 below), the "real" production related dimension of the volatility is highlighted.

Tables 5–12 contain the standard deviations, means and unit root tests of firm-level and industry-level stock prices and dividends. Results are also included for the aggregate industry data, i.e. the average industry stock price and dividend per share computed by the S&P Analyst Handbook. To control for movements in the general market, analysis was also done on the firm-level data divided by the S&P500 equivalent. The results for these deflated (by the S&P500) values are found in italics. However, in both industries no qualitative differences were found between units that were not deflated and those that were.

Automobiles. The results in Tables 6 and 10 indicate that firm-level and industry-level stock prices and dividends were most volatile in the same period that units and market shares were the most volatile: the period 1918–1941, with most volatility of stock prices occurring between 1918–1928 and the most volatility of dividends in the period 1933–1941. Units were even more volatile in the period preceding 1918 but firms were not quoted on the stock market yet. This holds for all the firms except for Studebaker which instead experienced more volatility of both units and stock prices in the post-war period (1948–1970) but more volatility of dividends in the pre-war period. Division by the S&P 500 index does not alter any of the qualitative results between the two periods (i.e. the earlier period is still much more volatile), except again in the case of Studebaker whose dividends were more volatile in the post-war period when divided by the S&P500 and vice versa when not divided.

Whereas Table 2 indicates that firm-level units and market shares follow an I(1) process in the pre-war and an I(0) process in the post-war period, Tables 6 and 10 indicate that most of the stock prices and dividends follow an I(1) process in both periods (as does the S&P500 stock price index). This suggests the possible presence of "excess volatility" in both periods. Using the stock prices derived from the Efficient Market Model as a benchmark against which to compare the volatility of actual stock prices, Mazzucato and Semmler (1999) and Mazzucato (2002) find that the degree of excess volatility was higher during the early stage of the auto industry.

Table 6 indicates that the average relative automobile stock (i.e. the average auto industry stock price divided by the S&P500 stock price) grew much less than the economy average in the post-war period. This is also the period when average industry sales growth began to fall (the mean growth rate both at the firm and industry level is negative after 1970). As regards the last three decades, stock prices

Table 5. Analysis of the statistical properties of the series (processes) of the real stock prices (logs) by US firms: DF and ADF tests

Firm	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Chrysler/	1925–1941	SBC	DF+drift	−1.6451	−3.0819	I(1)
	1948–1997	SBC	DF+drift	−1.2896	−2.9202	I(1)
Ford	–	–	–	–	–	–
	–	–	–	–	–	–
	1956–1998	SBC	DF+drift	−1.0184	−2.9339	I(1)
GM	1918–1941	SBC	DF+drift	−1.85	−3.0039	I(1)
	1925–1941	SBC	DF+drift	−1.4806	−3.0522	I(1)
	1948–1998	SBC	DF+drift	−1.5644	−2.919	I(1)
Hudson	1922–1941	SBC	DF+drift+t	−2.4929	−3.6921	I(1)
	1925–1941	SBC	ADF(1)+drift+t	−3.747	−3.7119	I(0)
Nash	1922–1941	SBC	DF+drift+t	−4.2651	−3.6921	I(0)
	1925–1941	SBC	DF+drift+t	−3.9201	−3.7119	I(0)
Packard	1917–1941	SBC	ADF(1)+drift+t	−5.1512	−3.6592	I(0)
	1925–1941	SBC	ADF(1)+drift+t	−4.7759	−3.7119	I(0)
Studeback	1920–1941	SBC	ADF(1)+drift	−3.516	−3.0199	I(0)
	1925–1941	SBC	ADF(1)+drift	−3.2718	−3.0522	I(0)
	1948–1966	SBC	DF+drift	−2.2482	−3.0294	I(1)
American	–	–	–	–	–	–
	–	–	–	–	–	–
	1954–1986	SBC	ADF(1)+drift+t	−2.9128	−3.5615	I(1)
Industry	1918–1941	SBC	DF+drift	−1.4876	−3.0039	I(1)
	1925–1941	SBC	DF+drift	−2.8846	−3.0522	I(1)
	1948–1998	SBC	ADF(1)+drift	−2.181	−2.919	I(1)
S&P500	1913–1941	SBC	DF+drift	−2.3482	−2.975	I(1)
	1925–1941	SBC	ADF(1)+drift	−3.0079	−3.0522	I(1)
	1948–1998	SBC	DF+drift	−0.083576	−2.919	I(1)

DF = Dickey-Fuller Test, drift stands for intercept, t for trend
ADF = Augmented Dickey-Fuller Test, drift stands for intercept, t for trend
SBC = Schwarz Bayesian Criterion (based on ML)
I(1) = Integrated process of order 1, or Difference Stationary (DS)
I(0) = Stationary process, or Trend Stationary (TS)

– like units and market shares – were more volatile in the decade 1970–1980 than in the following two decades.

Personal Computers. Table 7 indicates that all firm stock prices, except those of Unisys and Gateway, contained a unit root. This is consistent with Shiller's finding that stock prices tend to move like a random walk (Shiller, 1989). Table 8 illustrates that in the PC industry, firm and industry stock prices were most volatile in the *last* decade (especially in the early 1990's). Thus, stock prices were the most volatile in the same decade that relative growth rates were the most volatile. Units, measured in terms of relative growth, were instead the most volatile in the first decade when entry rates were highest (Mazzucato, 2002).

Table 6. Descriptive statistics of stock prices (logs of differences) in the US auto industry

		1918–28	1918–41	1948–00	1948–70	1970–2000	1970–1980	1980–1990	1990–2000
gm	st. dev	0.170964	0.147147	0.095173	0.080395	0.107536	0.096764	0.147021	0.038757
	mean	0.055331	0.043339	0.006101	−0.000395	0.00716	−0.017501	0.016151	0.01551
ford	st. dev			0.107571	0.08662	0.116013	0.125156	0.096478	0.128755
	mean			0.006197	0.008201	0.004956	−0.040533	−0.000331	0.036639
chrysler	st. dev	0.469634	0.248602	0.14847	0.08969	0.187895	0.154917	0.215675	0.185247
	mean	−0.080296	−0.000182	−0.001796	−0.016302	0.001725	−0.084603	0.03748	0.018132
amc	st. dev			0.170042	0.19818	0.13688	0.092815	0.188094	
	mean			−0.001156	0.020099	−0.027492	−0.020019	−0.038167	
studeb	st. dev	0.206597	0.195977	0.280321	0.280321				
	mean	−0.006913	−0.007087	0.011368	0.011368				
packard	st. dev	0.223114	0.187234	0.107029	0.107029				
	mean	0.057746	0.011162	−0.041811	−0.041811				
hudson	st. dev	0.200407	0.186195	0.098529	0.098529				
	mean	0.095105	0.010527	−0.01907	−0.01907				
nash	st. dev	0.390304	0.316962						
	mean	0.028112	0.017133						
industry	st. dev	0.145861	0.139372	0.079111	0.067126	0.088107	0.08955	0.088704	0.081875
	mean	0.093952	0.062004	0.029882	0.033595	0.024391	−0.019957	0.030865	0.045813

Auto stock prices divided by the S&P500 stock price

		1918–28	1918–41	1948–00	1948–70	1970–00	1970–80	1980–90	1990– 00
gm	st. dev	0.213194	0.172266	0.05879	0.066352	0.051239	0.044026	0.072389	0.02192
	mean	0.003571	0.002638	−0.01278	−0.02173	−0.00648	−0.00982	−0.00759	−0.00803
ford	st. dev			0.050352	0.046057	0.052478	0.063024	0.052042	0.045389
	mean			−0.0074	−0.00587	−0.00717	−0.02108	−0.0123	0.001149
chrysler	st. dev	0.444806	0.243829	0.065299	0.040836	0.079745	0.064677	0.097673	0.073517
	mean	−0.18241	−0.05268	−0.01673	−0.0318	−0.00736	−0.04255	0.007565	−0.00414
amc	st. dev			0.095735	0.120111	0.064283	0.042164	0.088971	
	mean			−0.00584	0.00264	−0.01593	−0.01062	−0.02475	
studeb	st. dev	0.207561	0.167586	0.18901	0.18901				
	mean	−0.04401	−0.03154	−0.01761	−0.01761				
packard	st. dev	0.187661	0.145828	0.07082	0.07082				
	mean	0.020267	−0.01436	−0.04794	−0.04794				
hudson	st. dev	0.188132	0.130295	0.079557	0.079557				
	mean	−0.0045	−0.0456	−0.0423	−0.0423				
nash	st. dev	0.392557	0.312088						
	mean	−0.08453	−0.10006						
industry	st. dev	0.125784	0.108988	0.035248	0.037218	0.033578	0.040372	0.038819	0.027427
	mean	0.061789	0.035215	−0.00204	0.000273	−0.00369	−0.01134	−0.00555	−0.00424

Table 13 summarizes the above results using aggregate volatility figures: market share instability (as defined in Hymer and Pashigian, 1962),[5] the standard deviation of the growth of total units, the standard deviation of the growth of stock prices and dividends, and the latter two divided by the S&P 500 equivalents. It is clear that in both industries stock prices were the most volatile in the period when market share instability was the highest. The next section documents that in both industries this was also the period in which innovation was the most radical, i.e. had greatest impact on the production process.

---

[5] Hymer and Pashigian (1962) define market share instability, I, as: $I = \sum_{i=1}^{n} [|s_{it} - s_{i,t-1}|]$, where s_{it} = the market share of firm i at time t.

Table 7. DF-ADF tests for the logs of real stock prices in the US PC industry

Firm	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Apple	1983–2000	SBC	DF+drift	−2.333	−3.04	I(1)
Hewlett-Pack	1979-2000	–	–	–	–	–
IBM	1979–2000	SBC	ADF(1)+drift	−1.822	−3.003	I(1)
NCR	1979–2000	–	–	–	–	–
Unisys	1979–2000	SBC	ADF(1)+drift+1	−3.855	−3.633	I(0)
Commodore	1979–1993	SBC	DF+drift	−2.071	−3.081	I(1)
Compaq	1985–2000	SBC	DF+drift	−2.117	−3.066	I(1)
Dell	1990–2000	SBC	DF+drift+t	−1.746	−3.927	I(1)
Gateway	1995–2000	SBC	DF+drift+t	−4.683	−4.581	I(0)
Toshiba	1992–1997	SBC	ADF(1)+drift	−1.498	−3.551	I(1)
Wang	1979–1998	SBC	DF+drift	−2.613	−3.019	I(1)
Wyse	1986–1988	SBC	DF+drift	−0.393	−4.706	I(1)

DF = Dickey-Fuller Test, drift stands for intercept, *t* for trend
ADF = Augmented Dickey-Fuller Test, drift stands for intercept, *t* for trend
SBC = Schwarz Bayesian Criterion (based on ML)
I(1) = Integrated process of order 1, or Difference Stationary (DS)
I(0) = Stationary process, or Trend Stationary (TS)

Table 8. Descriptive statistics of stock prices (logs of differences) in the US PC Industry (stock prices divided by the S&P 500)

		1970–00	1970–80	1980–90	1990–00	1970–00	1970–80	1980–90	1990– 00
apple	st. dev	0.209303		0.238666	0.19128	0.083065		0.09724	0.073131
	mean	0.043676		0.064494	0.005768	0.009145		0.020582	−0.00922
compaq	st. dev	0.225305		0.230429	0.218546	0.089324		0.099345	0.078982
	mean	0.091429		0.135991	0.068572	0.032226		0.057387	0.018271
dell	st. dev	0.370061		0.694571	0.350854	0.138582		0.281417	0.131854
	mean	0.201428		0.190106	0.247106	0.072183		0.086022	0.089015
everex	st. dev	0.239683		0.296755	0.104951	0.09334		0.117415	0.034097
	mean	−0.129761		−0.052906	−0.199099	−0.05629		−0.02735	−0.08481
gateway	st. dev	0.337883			0.337883	0.107928			0.107928
	mean	0.080564			0.080564	0.019431			0.019431
hpackard	st. dev	0.132265	0.134861	0.135744	0.150659	0.059641	0.070197	0.061558	0.063048
	mean	0.043979	0.025972	−0.000603	0.081044	0.011006	0.013164	−0.0052	0.019333
ibm	st. dev	0.122494	0.098844	0.103189	0.147733	0.054581	0.053856	0.056612	0.052748
	mean	0.027457	0.005942	0.022224	0.050709	0.001435	0.004051	−0.00277	0.004831
nec	st. dev	0.265431	0.307358	0.295656	0.171236	0.123413	0.152363	0.140936	0.056444
	mean	0.050035	0.275437	0.019097	0.007448	0.015864	0.1347	−0.00042	− 0.00681
toshiba	st. dev	0.298498			0.298498	0.119026			0.119026
	mean	0.077906			0.077906	0.02701			0.02701
unisys	st. dev	0.237456	0.090132	0.262856	0.353007	0.093811	0.054849	0.108797	0.133124
	mean	−0.013906	−0.040914	−0.092654	−0.000323	−0.01384	−0.01852	−0.05017	−0.00969
industry	st. dev	0.090528	0.070891	0.066299	0.119646	0.034966	0.029413	0.032475	0.044552
	mean	0.02582	−0.004754	0.015488	0.058539	−0.00383	−0.00399	−0.01369	−0.0003

5 Innovation

So far, it has been shown that (a) in both the auto and PC industry, relative growth rates were the most volatile in the same period that stock prices were the most volatile, and (b) in the auto industry this period was also the one in which entry rates were highest, while in the PC industry this period occurred a decade after the industry “shakeout”. Innovation dynamics in the two industries help to interpret these two results.

Table 9. Analysis of the statistical properties of the series (processes) of the real dividends (logs) by US firms: DF and ADF tests

Firm	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Chrysler	1925–1941	SBC	DF+drift	−1.5616	−3.0819	I(1)
	1948–1997	SBC	DF+drift+t	−3.0655	−3.5005	I(1)
Ford	–	–	–	–	–	–
	–	–	–	–	–	–
	1956–1998	SBC	DF+drift+t	−3.1455	−3.5217	I(1)
GM	1918–1941	SBC	DF+drift	−4.368	−3.0039	I(0)
	1925–1941	SBC	DF+drift	−3.5828	−3.0522	I(1)
	1948–1998	SBC	DF+drift+t	−4.1592	−3.4987	I(0)
Hudson	1922–1941	SBC	ADF(1)+drift+1	−2.0859	−3.6921	I(1)
	1925–1941	SBC	ADF(1)+drift+1	−2.9512	−3.7119	I(1)
Nash	1918–1941	SBC	DF+drift+t	−2.1926	−3.6921	I(1)
	1925–1941	SBC	DF+drift+t	−1.8594	−3.8731	I(1)
Packard	1917–1941	SBC	ADF(1)+drift	−2.6589	−2.997	I(1)
	1925–1941	SBC	DF+drift	−2.0046	−3.0522	I(1)
Studeback	1920–1941	SBC	DF+drift+t	−2.6264	−3.6746	I(1)
	1925—-1941	SBC	DF+drift+t	−2.5698	−3.7119	I(1)
	1948–1966	SBC				

DF = Dickey-Fuller Test, drift stands for intercept, t for trend
ADF = Augmented Dickey-Fuller Test, drift stands for intercept, t for trend
SBC = Schwarz Bayesian Criterion (based on ML)
I(1) = Integrated process of order 1, or Difference Stationary (DS)
I(0) = Stationary process, or Trend Stationary (TS)

In the auto industry, the period in which relative and absolute growth rates and stock prices were the most volatile was also the period in which entry rates were the highest, and technological change was the most radical, i.e. the first three decades of industry evolution. In an in depth survey on innovation in the auto industry, Abernathy et al. (1983) find that the innovations that impacted the production process of automobiles the most occurred between 1900 and 1935. Similarly, using the hedonic price index that they created, Raff and Trajtenberg (1997) find that most of the real change in automobile prices between 1906–1982 occurred between 1906–1940, and within that period most of the change occurred between 1906–1918. Since falling prices were mainly caused by technological change (as well as by the diffusion of mass production and the general expansion of the market), this means that most technological change occurred before 1940. This is confirmed by Filson (2001) using a "quality change" index, derived by dividing the BEA actual price index by the hedonic price index (Filson, 2001).

In the PC industry, the period in which growth rates and stock prices were the most volatile was also the period in which technological change was the most radical. This result only holds for *relative* growth rates since the period of radical change (1995–2000) occurred well *after* the surge in entry (1974–1984). As pointed out in Mazzucato (2002), the reason behind this difference is that unlike the auto industry, the PC industry emerged from a pre-existing industry (the mainframe industry) in which the industry leader (IBM) was able to control PC innovation and market shares for almost two decades. IBM's monopoly withered away due to

Table 10. Descriptive statistics of dividends (logs of differences) in the US auto industry

	gm	ford	chrysler	studeb	packard	hudson	nash	industry
1918–1928								
mean	0.014397		0.110924	−0.023247	0.066026		0.013197	0.028383
sd	0.482548		0.019819	0.536553	0.467959		0.248424	0.323424
1918–1941								
mean	0.033447		0.076592	−0.013561	0.042445	0.037086	0.013197	0.033497
sd	0.390041		0.293965	0.454954	0.421882	0.605477	0.248424	0.260793
1948–2000								
mean	−0.009355	−1.35E18	−0.007187	0.117701	0.174293	0.0546 (	−0.008183	
sd	0.236352	0.183059	0.283631	0.41256	0.501824	0.316417		0.166783
1948–1970								
mean	0.002363	−0.001266	−0.024958	0.117701	0.174293	0.0546		0.000777
sd	0.217482	0.115859	0.302836	0.41256	0.501824	0.316417		0.166347
1970–2000								
mean	−0.018327	0.000633	−0.011149					−0.021017
sd	0.249796	0.2078	0.284339					0.169351
1970–1980								
mean	−0.013752	0.027978	−0.027366					−0.00438
sd	0.17152	0.121243	0.415832					0.135948
1980–1990								
mean	0.02803	0.02128	0.070741					0.040017
sd	0.300927	0.276309	0.253263					0.141594
1990–2000								
mean	−0.039131	−0.045216	−0.008671					−0.066232
sd	0.300438	0.185551	0.163264					0.238726

Dividends divided by S&P500 dividends

	gm	ford	chrysler	studeb	packard	hudson	nash	industry
1918–1928								
mean	−0.20397		−2.97376	0.002687	−0.38466	−1.85289	−1.43912	−0.53587
sd	1.970619		1.351719	2.158434	1.807524	1.927482	2.179365	1.351978
1918–1941								
mean	−0.24482		−1.03621	1.1693	0.554859	−1.85289	−1.12314	−0.14841
sd	1.783283		3.996167	3.064914	2.807317	1.927482	1.935122	2.216052
1948–2000								
mean	0.038245	0.021935	0.153166	−1.55441	−2.30179	−0.75077		−0.08632
sd	3.860647	0.260755	2.010349	8.667584	31.47605	2.343213		1.120368
1948–1970								
mean	0.137328	0.064747	0.280318	−1.55441	−2.30179	−0.75077		−0.1759
sd	6.268709	0.309609	2.877157	8.667584	31.47605	2.343213		1.672397
1970–2000								
mean	−0.02449	0.001267	−0.01519					−0.02433
sd	0.291131	0.233028	0.459053					0.189174
1970–1980								
mean	−0.03281	0.036619	−0.05477					−0.01699
sd	0.305296	0.171906	0.654712					0.218123
1980–1990								
mean	0.019654	0.010679	0.179367					0.04059
sd	0.324001	0.311346	0.371452					0.165079
1990–2000								
mean	−0.0331	−0.04334	−0.00831					−0.05725
sd	0.262582	0.160829	0.148537					0.20252

three main events: (1) Intel's introduction of the 32-bit 386 chip in 1985, (2) the introduction of Windows in 1990, and (3) the commercial rise of the World Wide Web in 1995. Whereas the innovations introduced in the 1970's and 1980's were controlled by IBM (since everything had to be IBM compatible), the above quality changes in the 1990's shifted power from IBM to Microsoft and Intel (Bresnahan

Table 11. Descriptive statistics for logs of real dividends in the US PC industry

Firm	Market	Sample	Selected by	Reference test	Test stat.	95% crit. value	Process
Apple	US	1989–1996	SBC	DF+drift+t	0.856	−4.196	I(1)
Hewlett-Pack	us	1979–2000	SBC	DF+drift+t	−2.016	−3.633	I(1)
IBM	US	1979–2000	SBC	ADF1+drift+t	−2.54	−3.633	I(1)
NCR	US	1979–1991	SBC	DF+drift+t	−4.571	−3.828	I(0)
Unisys	US	1979–1990	SBC	DF+drift+t	−1.304	−3.873	I(1)
Wang	US	1979–1989	SBC	DF+drift	−2.876	−3.18	I(1)
Wyse	US	1979–2000	–	–	–	–	–
DF-ADF tests for E/S by firm							
Apple	US	1980–2000	SBC	ADF(1)+drift	−3.498	−3.011	I(0)
Hewlett-Pack	US	1979–2000	SBC	DF+drift	1.905	−3.004	I(1)
IBM	US	1979–2000	SBC	ADF(1)+drift	−2.195	−3.004	I(1)
NCR	US	1979–2000	SBC	ADF(1)+drift+1	−2.505	−3.633	I(1)
Unisys	US	1979–2000	SBC	DF+drift	−3.618	−3.004	I(0)
Commodore	US	1979–1993	SBC	ADF(1)+drift	−2.46	−3.082	I(1)
Compaq	US	1983–2000	SBC	DF+drift	−4.611	−3.04	I(0)
Dell	US	1987–2000	SBC	DF+drift+t	0.815	−3.792	I(1)
Gateway	US	1992–2000	SBC	ADF(1)+drift+1	−4.993	−4.081	I(0)
Toshiba	US	1979–2000	SBC	DF+drift	−3.354	−3.004	I(0)
Wang	US	1979–1998	SBC	DF+drift+t	−2.613	−3.659	I(1)
Wyse	US	1983–1988	SBC	DF+drift	−1.696	−3.551	I(1)

Table 12. Descriptive statistics of dividends (logs of differences) in the US PC Industry (dividends divided by the S&P 500 dividend)

		1970–00	1970–80	1980–90	1990–00	1970–00	1970–80	1980–90	1990–00
apple	st. dev	0.263737		0.20787	0.233295	0.2505		0.2272	0.1972
	mean	0		0.18809	−0.074697	0.0279		0.2286	−0.0531
compaq	st. dev	0.367919			0.367919	0.3147			0.3147
	mean	0.376664			0.376664	0.3283			0.3283
hpackard	st. dev	0.078744	0.05656	0.101334	0.040454	0.1401	0.1227	0.1166	0.0387
	mean	0.066423	0.08854	0.033962	0.085194	0.1460	0.3103	0.0926	0.0886
ibm	st. dev	0.116946	0.03359	0.017298	0.177071	0.1255	0.0739	0.0193	0.1613
	mean	0.008652	0.06962	0.013481	−0.040089	0.0299	0.1375	0.0149	− 0.0338
nec	st. dev	0.159784	0.03196	0.232574	0.056952	0.2020	0.0883	0.2796	0.0488
	mean	−0.003193	0.05077	−0.017321	−0.00702	0.0436	0.1812	0.0198	0.0107
toshiba	st. dev	0.080092	0.08009			0.1226	0.1226		
	mean	0.047025	0.04702			0.0756	0.0756		
unisys	st. dev	0.125461	0.11623	0.09945		0.1868	0.1933	0.0975	
	mean	0.03501	0.11917	−0.013284		0.0833	0.2308	−0.0065	
industry	st. dev	0.1107	0.0663	0.0364	0.1689	0.1161	0.1024	0.0425	0.1524
	mean	0.0091	0.0567	0.0138	−0.0410	0.0042	0.0663	−0.0096	−0.0443

Table 13. Standard deviation and mean of market shares, units, stock prices and dividends

	MS Inst.	Units	Stock	Dividend	Stck/SP500	Div/SP500
AUTO						
1908–1918	**25.2**	**0.1620**				
		0.0401				
1918–1928	**22.6**	**0.1569**	**0.1458**	**0.3234**	**0.1257**	1.3520
		0.0304	*0.0939*	*0.0283*	*0.0617*	*−0.5536*
1918–1941	17.9	0.1500	0.1393	0.2608	0.1089	**2.2161**
		0.0378	*0.0620*	*0.0334*	*0.0352*	*−0.1484*
1948–2000	7.6	0.0638	0.0791	0.1668	0.0352	0.9974
		0.0070	*0.0298*	*−0.0881*	*−0.0020*	*−0.0863*
1948–1970	10.3	0.0759	0.0671	0.1663	0.0372	1.5054
		0.0171	*0.0335*	*0.0007*	*0.0002*	*−0.1759*
1970–2000	5.6	0.0523	0.0881	0.1694	0.0335	0.1772
		−0.0030	*0.0243*	*−0.0210*	*−0.0036*	*−0.0243*
PC						
1970–1980	1.4	**0.2062**	0.0708	0.0663	0.0294	0.1024
		0.2431	*−0.0047*	*0.0567*	*−0.0039*	*−0.0663*
1980–1990	11.5	0.1884	0.0662	0.0364	0.0324	0.0425
		0.1450	*0.0154*	*0.0137*	*−0.0136*	*−0.0096*
1990–2000	**17.9**	0.0357	**0.1196**	**0.1689**	**0.0445**	**0.1524**
		0.0646	*0.0585*	*−0.0410*	*−0.0003*	*−0.0043*
1970–2000	28.9	0.1758	0.0905	0.1107	0.0349	0.1166
		0.1504	*0.0258*	*0.0091*	*−0.0038*	*0.0041*

italics=mean value
bold number=decade with highest value
MS lnst.= instability index from Hymer and Pashigian (1962)
Units=standard deviation and mean of units produced
Stock=standard deviation and mean of industry-level stock price
Dividend=standard deviation and mean of industry-level dividend
Stck/SP500= industry-level stock price divided by S&P500 stock price
Div/SP500=industry-level dividend divided by S&P500 dividend

and Greenstein, 1997) and caused a great upheaval in relative growth rates.[6] The "quality change" index indicates that most of the percentage change in the quality of personal computers occurred in the first and third decades: 34% between 1975–1986, and 38% in the period 1993–1999 (soon after Windows 3.0 was introduced),

---

[6] In the 1980's IBM focussed on incremental technical change with backward compatibility: all other firms' hardware and software products had to work with IBM equipment. IBM's power first came under threat when the Intel 80386 chip was used by Compaq's new computer, so that the computer was marketed for the quality of the chip not the IBM compatibility. Once the "industry standard" label became more important than the "IBM compatible" label, IBM became much weaker. The next shake-up to the power structure came when IBM split with Microsoft over operating systems in 1990. The industry standard now changed to the "Wintel" standard, finishing off what remained of IBM's special status. Another reason why the 1990's presented such disruptive change was due to the development of the new "client/server" platform (networked platform with highly intelligent terminals). This new platform was based on a vertically disintegrated structure which devalued traditional management causing the strengths of the incumbents (mainly IBM and DEC) to become obsolete (Bresnahan and Greenstein, 1997). The tradition of backward compatibility made the incumbent platforms particularly hard to change in reaction to the users' new needs.

with only 17% in the middle stage (Filson, 2001). Hence the last decade in the PC industry witnessed the most quality change and also the most volatility in relative growth and stock prices.

To summarize, in both industries, innovation was the most radical and "competence destroying" in the period when growth rates and stock prices were the most volatile. Whereas in the auto industry this applies to both relative and absolute growth rates, in the PC industry this applies only to relative growth rates. This is because the dynamic of absolute and relative growth only coincides if the period of market growth (when new firms enter) coincides with the period of market instability (when the market shares of those firms undergo turbulence). This holds for industries which follow a classical life-cycle pattern (Klepper, 1996), but not for those industries which, for example, emerge from a pre-existing industry and hence contain some of the characteristics of the mature phase in the very early years (i.e. the existence of an incumbent). The fact that it is volatility in relative growth rates that coincides with radical innovation, supports the emphasis in evolutionary economics that relative not absolute growth captures the real nature of competition (Dosi and Nelson, 1994).

6 Panel data analysis: inter-firm heterogeneity

The efficient market model (EMM) states that a firm's stock price is equal to the expected value of discounted future dividends:

$$v_t = E_t v_t^* \qquad \text{where} \tag{2}$$

$$v_t^* = \sum_{k=0}^{\infty} D_{t+k} \prod_{j=0}^{k} \gamma_{t+j} \tag{3}$$

where v_t^* is the ex-post rational or perfect-foresight price, D_{t+k} is the dividend stream, γ_{t+j} is a real discount factor equal to $1/(1 + r_{t+j})$, and r_{t+j} is the short (one-period) rate of discount at time $t+j$. The argument in Section 5 above amounts to saying the discount factor used in γ_{t+j} should, if markets take into account all information, embody an understanding of the risk and uncertainty that arises from radical innovation. Using this argument, one would expect the predictive power of dividends (and other fundamentals) to be greater in the period in which innovation is low. The fact that there are not many examples of finance models which study risk as an outcome of firm and industry specific innovation, suggests that there is much to be learned from the dialog between the Schumpetarian literature on innovation and the finance literature on risk and market valuation.

To better understand the degree to which stock price volatility and growth rate volatility *co-evolve*, panel data analysis is used to regress the rate of change of firm stock prices on life-cycle variables: firm numbers, market share instability, market concentration, and also on more traditional fundamentals at both the firm, industry and economy wide level (e.g. firm dividends, industry dividends, S&P 500 stock price, etc.). Given the results already obtained, the goal is to see whether in the early phase of industry evolution (i.e. the first three decades of industry evolution) stock prices are more related to variables defining industrial instability, and whether in the mature phase they react more to changes in fundamentals. We also explore the role of heterogeneity in the different periods of industry evolution, i.e. the degree to which firms differ from each other in how they are affected by the different variables.

Due to the low number of firms and the long time period, Seemingly Unrelated Regression estimations (SURE) are used (Zellner, 1962; Smith, 2000). Wald tests are used to test for inter-firm heterogeneity, both in terms of the differences between firms with respect to a single regressor (Wald test type 1) and the differences between firm-specific coefficients for all the regressors (Wald test type 2). That is, it tests for joint restrictions for homogeneity in the slope coefficients (i.e. the Fixed Effect hypothesis).

Firm-level stock prices were regressed on firm-level dividends (Div), firm market share (MktSh), the S&P500 index stock price (StPrSP500), the S&P500 index dividends (DivSP500), the number of firms in the industry (Nfirms), and the level of concentration (Herf). Other variables were also included but since no convergence occurred of the Maximum Likelihood (ML) algorithm, they had to be omitted. For reasons of convergence, multicollinearity and parsimony, different specifications were tried for the PC industry, where each specification includes a different sub-set of firms. The t-statistics in the SURE analysis are often very large because of the convergence procedure. Due to space limitation, only the tables for the restricted case are illustrated, the unrestricted case which requires also comparison with single equation OLS estimates are treated verbally.

Automobiles (1918–1941, 1948–2000). In Tables 14–15 Wald tests indicate that in the pre-war period the joint restrictions for homogeneity (the Fixed Effect hypothesis) and the restrictions of the firm-specific coefficients for the single regressors can both be rejected. In the post-war period we *cannot* reject this restriction on the whole set of parameters and we can also not reject the restrictions for homogeneity of the firm-specific coefficients for all the single regressors. This means that in the post-war period there is more homogeneity between firms in how stock prices are affected by the different variables. In the pre-war period, the rate of change of firm stock prices are significantly affected by changes in market shares, the number of firms and the herfindahl index. Neither the firm level, industry level nor market level fundamentals seem to be significant in this early period. In the post-war period there is increased significance of the fundamentals (both at the firm level and at the general market level) and no significance of the industrial dynamics variables (market shares, number of firms and herfindahl index).

Personal Computers (1975–1999). In each of the different specifications, the results were similar to those that emerged in the *pre-war* period for automobiles: rejection of the joint restrictions for homogeneity of the whole set of parameters (Fixed Effect panel hypothesis) and non-rejection of the restriction for homogeneity of the firm-specific coefficients for most of the single regressors. As in the pre-war auto industry, the most significant variables in this early stage of evolution are changes in market shares, the number of firms and the herfindhal index. The financial fundamentals both at the level of the firm and at the level of the general market were less significant.

Table 14. Auto industry 1927–1951 – Restricted SURE estimates (ML estimates for the Fixed Effect Panel), homogeneity restrictions for all the coefficients (variables in logs of first differences)

	int	StPrSP500	Div	Div SP500	MktSh	Nfirms	Herf
Chrysler	0.097211	–	–	–	–	–	–
t-value	1.802	–	–	–	–	–	–
Gen. Motors	0.029421	–	–	–	–	–	–
t-value	0.36932	–	–	–	–	–	–
Studebaker	0.092913	–	–	–	–	–	–
t-value	0.34865	–	–	–	–	–	–
Hudson	−0.012132	–	–	–	–	–	–
t-value	−0.13587	–	–	–	–	–	–
Packard	−0.048739	–	–	–	–	–	–
t-value	−0.11418	–	–	–	–	–	–
FE Panel	0.0317348	1.1901	0.22154	−0.75814	1.6945	0.10523	−2.5164
t-value	0.453984	6.7143	7.8966	−3.5449	4.3548	0.29019	−8.4993
Wald-test for Homog restr (1)		1.28E+11					
Wald-test for Homog restr (2)		31.7782	8.01 E+10	105.6889	2.28E+09	165.1449	3.53E+01
ML	−23.3301						
SBC	−38.2243						

Wald Homogeneity restriction test (1): Homogeneity restrictions (equality) for thefirm-level coefficients of all the regressors.
Wald Homogeneity restriction test (2): Homogeneity restrictions (equality) for the firm-level coefficients of the single variables.

These results indicate that in both industries stock price dynamics in the early phase of the industry life-cycle were affected significantly by the turbulence in market structure: changing number of firms, rising concentration and market share dynamics. On the other hand, firm level and market level fundamentals (dividends, earnings per share) had a greater effect on stock price dynamics in the mature phase than in the early phase. This can of course only be tested for in the auto industry which has already reached its mature phase. Furthermore, in the early phase of both industries it is easier to reject the joint restrictions for homogeneity of the whole set of parameters, indicating that in this phase, unlike in the mature phase (for automobiles at least), there is more heterogeneity between firms. The fact that there is more heterogeneity between firms in the early period and the fact that firm level stock prices react to changes in industrial turbulence in this period supports the cointegration results in Mazzucato (2002) where it is found that stock price behaviour is more idiosyncratic in the early life-cycle phase (i.e. more firm-specific risk, as defined by Campbell et al., 2000).[7]

---

[7] The capital asset pricing model (CAPM) measures *idiosyncratic risk* through the covariance between the firm-level (or industry-level) stock return and the market-level stock return: the lower is this covariance the higher is the unsystematic or idiosyncratic level of risk. Through cointegration analysis, Mazzucato (2002) finds that this idiosyncratic risk is higher in the early phase of the industry life-cycle.

Table 15. Auto industry 1957–1997 – restricted SURE estimates (ML estimate for the Fixed Effect Panel) – homogeneity restrictions for all the coefficients (variables in logs of first differences)

	int	StPrSP500	Div	Div SP500	MktSh	Nfirms	Herf
Chrysler	−0.071372	–	–	–	–	–	–
t-value	−1.406	–	–	–	–	–	–
Gen.Motors	−0.042365	–	–	–	–	–	–
t-value	−1.316	–	–	–	–	–	–
Ford	−0.063612	–	–	–	–	–	–
t-value	−1.8497	–	–	–	–	–	–
FE Panel	−0.05911633	0.94722	0.061502	0.0043001	0.50341	0.11348	0.26794
t-value	−1.5239	5.0049	2.2092	0.18261	0.70544	0.2642	0.75981
Wald-test for Homog restr (1)		11,5370**					
Wald-test for Homog restr (2)		0,70006**		2,6299**	1,4789**	0,21925**	2,3759**
ML	7.7857						
SBC	−8.9254						

Wald Homogeneity restriction test (1): Homogeneity restrictions (equality) for the firm-level coefficients of all the regressors.
Wald Homogeneity restriction test (2): Homogeneity restrictions (equality) for the firm-level coefficients of the single variables.

7 Conclusion

The paper began by considering the literature on firm growth and the "random walk" hypothesis which is embodied in Gibrat's Law of Proportionate Growth. Unit root analysis and descriptive statistics on firm-specific growth rates (both absolute and relative) were used to test for this hypothesis. It was found that in the early auto and PC industry the Gibrat hypothesis describes the statistical process of firm growth better in the *early* phase of industry evolution. This is most likely because this early phase is characterized by rapidly evolving technology and a general expansion of the market (creating opportunities for some and disadvantages for others). Once changes in technology and demand settle down and managers become more concerned with process innovation and achieving economies of scale, firm growth rates tend to be more stable and structured. While one can clearly see the different patterns in the early versus mature phase of the auto industry (e.g. the unit root hypothesis cannot be rejected for most firms in the early phase while it can be rejected in the mature phase), the stable pattern has yet to show up in the mature phase of the PC industry – perhaps not too long from now given that the industry has been experiencing slower growth and much less product innovation (see "Personal computer shipments suffer first fall in 15 years", *The Financial Times*, July 21/22, 2001).

Stock prices were found to be most volatile in the decades that relative growth rates were most volatile. While in the auto industry this coincides with the period in which absolute growth rates were the most volatile, in the PC industry absolute growth rates were most volatile in the first decade (1974–1984) while relative growth rates were most volatile in the third decade of industry evolution (the 1990's). This is because in the auto industry the phase of initial market expansion with new firm entry and exit (causing volatility in absolute growth rates) was also characterized by radical innovation, causing a shake up in market shares during the very early years (Epstein, 1928). Instead, in the PC industry, the initial phase of expansion and entry (leading to high absolute growth) was not characterized by radical technological change since IBM dominated the growth of the industry and the innovation process. Only in the 1990's – the decade of the "new economy" – did innovation become free from IBM's dominance allowing the firms that had already entered years before to finally compete for industry market share (leading to volatility in relative growth).

A look at the dynamics of innovation in the two industries (through qualitative case studies and quantitative "quality change" indices) confirmed that the periods in which relative growth rates and stock prices were the most volatile in both industries were also the periods in which innovation impacted the production process the most. This suggests that both types of volatility are related to mechanisms of Schumpetarian creative destruction, or in the words of modern strategy theory, to the mechanisms of "competence-destroying" innovations – that destroy the monopoly of incumbents (Tushman and Anderson, 1986).

The analysis has put forward the proposition that the volatility of growth rates (in units) and stock prices is linked to the dynamics of innovation during early industry evolution. This view of volatility as capturing a "real" production related dimension of economic behavior lies in contrast with the view that sees volatility as resulting from random behavior (e.g. shocks, animal spirits).

References

Abernathy WJ, Clark K, Kantrow A (1983) Industrial renaissance: producing a competitive future for america. Basic Books, New York

Agiakoglou C, Newbold P (1992) Empirical evidence on Dickey-Fuller type tests. Journal of Time Series Analysis 13: 471–483

Berndt ER, Rappaport N (2000) Price and quality of desktop and mobile personal computers: a quarter century of history. Paper presented at the National Bureau of aeconomic Research Summer Institute 2000 session on Price, Output, and Productivity Measurement, Cambridge, MA, July 31, 2000

Bottazzi G, Cefis E, Dosi G (2002) Corporate growth and industrial structures: some evidence from the Italian manufacturing industry, Industrial and Corporate Change 11: 705–723

Braun PA, Nelson DB, Sunier AM (1995) Good news, bad news, volatility and beta. Journal of Finance 50: 1575–1603

Bureau of Economic Research Summer Institute 2000 session on Price, Output, and Productivity Measurement. Cambridge, MA, July 31, 2000

Bresnahan TF, Greenstein S (1997) Technological competition and the structure of the computer industry. Working paper, Stanford University

Campbell JY, Lettau M, Malkiel BG, Yexiao X (2000) Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56: 1–43

Cohen W, Levinthal D (1989) Innovation and learning: the two faces of R&D, implications for the analysis of R&D investment. Economic Journal 99: 569–596

Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74: 427–431

Dosi G, Nelson R (1994) An introduction to evolutionary theories in economics. Journal of Evolutionary Economics 4: 153–172

Epstein R (1928) The automobile industry; its economic and commercial development. Arno Press, New York

Evans D (1987) The relationship between firm growth, size and age: estimates for 100 manufacturing industries. Journal of Industrial Economics 35: 567–581

Filson D (2001) The nature and effects of technological change over the industry life cycle. Review of Economic Dynamics 4(2): 460–494

Geroski P, Machin S (1993) Innovation, profitability and growth over the business cycle. Empirica 20: 35–50

Geroski P, Mazzucato M (2002) Learning and the sources of corporate growth. Industrial and Corporate Change 11(4): 623–644

Gibrat R (1931) Les inegalites economiques. Requeil Sirey, Paris

Gort M, Klepper S (1982) Time paths in the diffusion of product innovations. Economic Journal 92: 630–653

Hall A (1994) Testing for a unit root in time series with pretest data based model selection. Journal of Business and Economic Statistics 12: 461–470

Hymer S, Pashigian P (1962) Turnover of firms as a measure of market behavior. Review of Economics and Statistics 44: 82–87

Ijiri Y, Simon H (1977) Skew distributions and sizes of business firms. North Holland, Amsterdam

Klepper S (1996) Exit, entry, growth, and innovation over the product life-cycle. American Economic Review 86(3): 562–583

Knight FH (1921) Risk, uncertainty and profit. Houghton Mifflin, Boston

Maddala GS, In-Moo K (1998) Unit roots, cointegration and structural change. CUP, Cambridge

Marsili O (2001) The anatomy and evolution of industries. Edward Elgar Press, Northampton, MA

Mazzucato M, Semmler W (1999) Stock market volatility and market share instability during the US automobile industry life-cycle. Journal of Evolutionary Economics 9(1): 67–96

Mazzucato M (2002) The PC industry: new economy or early life-cycle? Review of Economic Dynamics 5: 318–345

Raff DMG, Trajtenberg M (1997) Quality-adjusted prices for the American automobile industry: 1906–1940. In: Bresnahan TF, Gordon RJ (ed) The economics of new goods. NBER Studies in Income and Wealth 58, pp 71–107. University of Chicago Press, Chicago

Shiller RJ (1989) Market volatility. MIT Press, Cambridge

Simon HA, Bonini CP (1958) The size distribution of business firms. American Economic Review 48(4): 607–617

Standard and Poor's Industry Surveys (2000) Computers Hardware, Automobiles

Teece D, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. Strategic Management Journal 18(7): 509–533

Tushman M, Anderson P (1986) Technological discontinuities and organizational environments. Administrative Science Quarterly 31: 439–465

Wards Automotive Yearbook (1936–1995) Ward's Communications, Detroit

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and test for aggregation bias. Journal of the American Statistical Association 57: 348–368

Social networks and industrial geography[⋆]

Olav Sorenson

Anderson Graduate School of Management, University of California, Los Angeles, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, USA
(e-mail: osorenso@anderson.ucla.edu)

Abstract. In many industries, production resides in a small number of highly concentrated regions; for example, several high tech industries cluster in Silicon Valley. Explanations for this phenomenon have focused on how the co-location of firms in an industry might increase the efficiency of production. In contrast, this article argues that industries cluster because entrepreneurs find it difficult to access the information and resources they require when they reside far from the sources of these valuable inputs. Since existing firms often represent the largest pools of these important factors, the current geographic distribution of production places important constraints on entrepreneurial activity. As a result, new foundings tend to arise in the same areas as existing ones, and hence reproduce the industrial geography. In support of this thesis, the article reviews empirical evidence from the shoe manufacturing and biotechnology industries.

Keywords: Social networks – Entrepreneurship – Agglomeration – Spin-offs

JEL Classification: L11, M13, R30

1 Introduction

Production in a wide range of industries resides in a limited number of highly clustered geographic locations often referred to as industrial districts. In California, for example, Silicon Valley has become famous for its dense concentration of high technology companies while Los Angeles provides a home to a large share of the entertainment industry. The same pattern appears in other countries as well. Tijuana, a Mexican city near San Diego, hosts a large contingent of electronics manufacturing

---

[⋆] Adapted from a plenary talk delivered at the 8^{th} annual meetings of the International Schumpeter Society in Gainesville, FL. Constança Esteves and Lee Fleming provided comments useful to developing this written version.

firms and the region around Geneva in Switzerland dominates the high end of the watch industry. Simply detailing the long list of documented examples would require more pages than this paper permits.[1]

Explanations for this spatial distribution of industrial production have focused on arguing that these dense concentrations might increase the efficiency of the production and distribution of goods. The earliest research by German scholars maintained that the minimization of transportation costs could explain the concentration of heavy manufacturing in Bavaria because locating there allowed these firms to benefit from their close proximity to coal and iron ore (von Thünen, 1826; Weber, 1909). These transportation cost-based arguments, however, fail to account for the concentration of a variety of light manufacturing and service industries, such as high technology or entertainment, where these costs make up a negligible fraction of the value of the good. Attempts to explain the clustering of these industries has lead researchers to revisit the agglomeration economies proposed by Alfred Marshall (1920). Thus, recent work has elucidated the potential benefits of an extended division of labor (Romer, 1990), labor pooling (Diamond and Simon, 1990; Rotemberg and Saloner, 1990) and informational spillovers (Arrow, 1962). The logic of this literature implies that managers accurately recognize the benefits of certain locations and therefore locate there (or that strong selection pressures produce an equilibrium geographic dispersion from random entry).

Although these factors undoubtedly play a role in the maintenance of many industrial districts, geographic concentration can persist even when economic efficiency (at least in production) does not support it. The explanation for this phenomenon comes from a more nuanced consideration of the process of entrepreneurship – specifically, the importance of social networks to it. Two factors must converge for a nascent entrepreneur to found a new firm. First, the potential entrepreneur must perceive an opportunity for profit in a particular segment, or market niche, of the economy. Since much of the relevant information only exists privately, awareness of potentially profitable opportunities requires connections to those with the pertinent knowledge, typically those currently engaged in business in a particular industry. Second, the individual that perceives an opportunity must build a firm – assemble the necessary capital, skilled labor and knowledge – to exploit it. Again, social relationships play a crucial role in acquiring tacit information and in convincing resource holders to join the fledgling venture, whether as employees or investors. Because the social ties that facilitate both of these antecedents rarely extend beyond the regions in which these relevant resources and knowledge reside, entrepreneurs within a given industry most frequently arise in close proximity to industry incumbents. This regularity implies that industries can remain geographically concentrated even when co-location disadvantages firms.

The remainder of this piece delineates the reasoning behind this idea. Section 2 reviews the relationship between social networks and geography. Section 3 details the various mechanisms through which social networks enable entrepreneurship. The subsequent sections present some examples of how these processes affect the dynamics of two quite different industries: footwear production and biotechnology.

---

[1] For those interested in additional cases, Porter (1990) provides numerous examples.

Section 6 summarizes the material and its implications for theory and regional policy.

2 The spatial distribution of social networks

Social networks influence the geographic distribution of industries because networks do not randomly link individuals. Rather, people interact most frequently with those who live in close geographic proximity and with whom they share backgrounds, interests and affiliations (often referred to as social proximity). The parallel patterns – relating social and physical distance to the likelihood of a relationship – reflect the fact that both arise from influencing the probability of random interaction. To form a relation, two individuals typically must meet in space and time. Because both physical and social locations strongly influence people's activities, proximity on these dimensions increases the likelihood of a chance encounter (Blau, 1977). Moreover, even after an initial contact, each of these factors importantly influences the likelihood of forming and maintaining a relationship. Confirming the adage that "birds of a feather, flock together," people appear to prefer to develop and maintain social ties with those of similar backgrounds and with related interests (Lazarsfeld and Merton, 1954). And geographic proximity also strongly influences the durability of relationships by reducing the costs of maintaining a relation. Distance increases the direct expenses associated with engaging in the frequent and extended interaction necessary for the maintenance of social relationships (Zipf, 1949), particularly close personal ties. It also escalates the opportunity costs of interaction, in the sense that the number of equally preferred but more proximate individuals increases with distance (Stouffer, 1940).[2] Hence, individuals' social networks primarily contain ties to others like themselves that live nearby.

Roughly 80 years of empirical research support this principle. Beginning with Bossard (1932), a number of researchers have investigated the role of propinquity in marriage and friendship, consistently finding that the likelihood of a relationship declines precipitously as the physical distance between the two parties increases (Festinger, Schacter and Back, 1950). Studies of interaction patterns within organizations similarly observe that employees communicate more frequently with co-workers in nearby offices (Allen, 1977). More recently, researchers have begun to show that business relationships also follow this pattern. For example, corporate board interlocks more frequently occur among firms with geographically proximate headquarters (Kono, Palmer, Friedland, and Zafonte, 1998). Floor traders on an options exchange prefer to deal with those situated nearby on the floor (Baker, 1984). And venture capital firms rarely invest far from their offices (Sorenson and Stuart, 2001). A parallel line of research demonstrates that the probability of forming a relationship rises when two individuals reside closer to each other in 'social space' (i.e. have similar backgrounds or share demographic characteristics). For

---

[2] Stouffer (1940) and Zipf (1949) derive formal models predicting the expected functional form relating distance to the likelihood of interaction – the former based on the opportunity costs involved and the latter focusing more on the direct costs. Both models predict that the likelihood of a relationship should change roughly proportional to the inverse distance between two individuals.

example, social similarity so strongly structures interaction that it explains most of the variation in social networks in the 1985 General Social Survey (Marsden, 1988). Numerous other studies also find proximity in social space a salient factor in explaining who interacts with whom (for a review, see McPherson, Smith-Lovin and Cook, 2001).

3 Social networks and entrepreneurship

To understand the importance of social networks to entrepreneurship, let us think about the process of starting a firm from the point of view of a nascent entrepreneur. In considering this issue, we can usefully divide the process of entrepreneurship into two stages: (1) opportunity identification, and (2) organization building. In other words, our potential entrepreneur must first come up with a concept for a new business, and then she needs to assemble a company that will allow her to pursue her idea. Social networks play an important role in each of these stages.

Identifying opportunities

Our hopeful entrepreneur must first access social networks to identify opportunities in the economy. In emerging industries, knowledge of the existence of the enterprise itself may elude attention in the media. Our potential entrepreneur, hence, will only consider these ventures if she knows someone in the business, either directly or through an acquaintance. Even in more mature lines of business, much of the information necessary to assess the potential value of an opportunity only exists privately. Incumbent firms notably have incentives to conceal, for example, the profitability of their operations lest it encourage others to enter their attractive market segment. Private companies can simply avoid making public such sensitive data, while public firms – particularly those operating in multiple lines of business – can obscure their accounting statements to minimize the amount of information they reveal. Regardless, companies likely have difficulty completely stemming the flow of such knowledge. Many employees, chiefly in the managerial ranks of the firm, necessarily have access to this valuable information, potentially allowing them to pass it on to others. As a result, private information regarding entrepreneurial opportunities flows through social networks.

Multiple lines of empirical research confirm the importance of social networks as conduits for communication flows. For example, a substantial body of research investigating the diffusion of innovations has demonstrated that these ideas progress along the transmission lines made available by existing social ties (Coleman, Katz, and Menzel, 1966; see Rogers, 1995, for a review). In the study of labor markets, Granovetter (1973) and others have shown that much of the information on the availability of jobs passes through informal social ties. Research outside of sociology also increasingly pays attention to the importance of these knowledge flows. For example, recent studies in finance find that investors who invest locally earn positive abnormal returns (Coval and Moskowitz, 2001; Garmaise and Moskowitz, 2003), presumably because social networks bring them preferential access to data regarding the attractiveness of local opportunities.

The localized structure of social networks thus implies that our potential entrepreneur will be most aware of opportunities in the industry in which she works, particularly if she works for a firm situated in a regional concentration of like firms. Many of our nascent entrepreneur's relationships will tie her to those with whom she works, making her most aware of opportunities in her industry of employment. If a number of rival firms in that industry also operate in the local region, the likelihood that her non-work contacts also connect her to information flows within the industry increases. However, not just the number, but also the pattern of relationships often matters in the identification of valuable opportunities. The expansiveness of a person's social network, in particular, determines the breadth of information available to them. Sitting at the nexus of diverse knowledge flows offers opportunities to bring these ideas together into novel combinations (Schumpeter 1942). Therefore, entrepreneurs in regions with a large population of firms in an industry tend to occupy positions in communication networks that lend themselves to identifying promising opportunities and assessing market conditions in the industry.

Building new organizations

After identifying a potentially profitable opportunity, our entrepreneur requires access to a variety of resources to begin operations. Even in an emerging industry, she will need capital and labor. As industries mature, the efficiency of production increases due to a combination of capital investments, improvements in human capital specific to the business and the accumulation of tacit knowledge through learning-by-doing. For new entrants to compete effectively with incumbents, they require access to each of these elements – a process referred to as resource mobilization in the sociology literature. Social relations facilitate the acquisition of each of these three elements: (1) tacit knowledge, (2) financial capital, and (3) human capital.

Social networks may operate most strongly in structuring the flow of tacit information. By its very definition tacit knowledge defies codification. Regardless, tacit information frequently underlies the profitability of enterprises in many industries; precisely because of the difficulty of replicating such knowledge, it can provide valuable rents to those that hold it. New firms that can access the existing knowledge in the industry enjoy a large advantage (Klepper and Sleeper, 2000; Klepper, 2001). Despite its value, tacit information does not lend itself to market-based exchange. Potential buyers may question the value of the knowledge, and sellers cannot easily assuage their concern without revealing their valuable information. Dense social networks often prove useful in such circumstances because they facilitate the trust necessary for exchange to occur (Coleman, 1990). Moreover, accessing this information typically requires strong social ties. Complex, tacit knowledge eludes transfer in the absence of the high bandwidth that face-to-face contact makes possible (Nelson, 1959). For example, ethnographic accounts of science and industrial R&D commonly note that individuals acquire research capabilities through hands-on experience and apprenticeships with skilled researchers

(Latour, 1989). Together these factors suggest that our potential entrepreneur will need strong ties to individuals within the industry to access the valuable tacit knowledge she needs.

In fledgling industries and in concentrated, mature ones, the limited geographic availability of this important resource will influence the ease across regions of beginning a new venture within an industry. Industry-specific tacit knowledge resides primarily in the incumbent firms within an industry; thus, potential entrepreneurs need strong social connections to employees currently working in the industry. This condition virtually requires that new entrepreneurs in an industry hail from the ranks of current employees (Sorenson and Audia, 2000). Although the likelihood of any tie declines with physical and social distance, the odds of having the strong ties needed to acquire this tacit knowledge likely decline particularly quickly. Developing and maintaining these strong ties requires frequent and intensive interaction, a situation unlikely to occur except among co-workers and close, personal friends.

Though more efficient markets exist for the distribution of capital and labor, two factors impede access to these resources in the absence of social networks. First, all new ventures represent fundamentally uncertain propositions – not only is the undertaking risky, but even the level of risk is unknown. Notably, experimental studies have demonstrated that even risk-neutral investors exhibit ambiguity aversion when faced with an inability to quantify the risks (Fox and Tversky, 1995). Hence, both potential investors and employees likely view opportunities to join fledgling ventures with considerable suspicion. Compounding these issues, resource holders also face an information asymmetry problem when assessing a potential entrepreneur: The hopeful founder likely knows more about the quality of her idea than the investors and potential employees that she approaches. Hence, they cannot simply rely on her judgment because she has clear incentives to overstate the attractiveness of her proposition (Akerlof, 1970). These factors create a friction in the flow of financial and human resources that social networks can help lubricate.

By mitigating the perceived risk associated with investing, social ties to capital holders elevate the likelihood of an entrepreneur obtaining financial backing. At least two dimensions of an entrepreneur's contact network facilitate this process. First, individuals place greater confidence in information collected from trusted parties, making investors more likely to rely on information garnered from entrepreneurs if they share a strong social bond. For example, studies of the venture capital industry find that VCs prefer to finance investments that come to them through referrals from close contacts (Fried and Hisrich, 1994). Second, lacking a strong tie, consistent information across multiple independent sources might offer the investor some assurance regarding the reliability of their information regarding the potential investment (Sorensen and Stuart, 2001).

The importance of social networks in access to capital further limits the choice of locations for new firm builders. Mutual social ties with potential investors improve our entrepreneur's odds of obtaining capital. Although the availability of financial capital in general may vary little from region to region, the usefulness of these close social relations in obtaining financing tends to bind entrepreneurs to the regions in which they have contacts, even if other locations might seem more attractive. Moreover, the venture capital critical to the financing of many high tech industries does not exist everywhere; hence, its availability may constrain the spatial distribution of industries that rely on it for financing (Sorenson and Stuart, 2001).

In addition to financial capital, our entrepreneur must also recruit human capital to join her venture. Established firms often provide the largest pool of labor to new ventures of like kind (Sorenson and Audia, 2000). Given the uncertainties faced by fledgling companies, it may take considerable persuasion on the part of entrepreneurs to attract highly skilled workers away from secure positions at established organizations. Unlike investors, employees cannot easily diversify their risk making them even more sensitive to these problems. To acquire staff for her new company, our entrepreneur must leverage her network of industry contacts to persuade employees to leave their current employers to join the new firm. Without trust in the founder of the company and confidence in her ability and judgment, managerial and technical workers will not likely leave their secure positions to join a new startup. Again, strong social relationships facilitate the trust necessary to obtain these scarce human resources, particularly in the face of competition in the labor market (Løvås and Sorenson, 2003). Assuming that these relationships concentrate spatially, then proximity to existing firms should greatly expedite the recruitment of a workforce.

To summarize, nascent entrepreneurs have the best information on opportunities and the best ability to mobilize the resources necessary to build a firm (1) in the industries in which they have experience, and (2) in the regions in which they live. Together, these propositions imply that the geographic distribution of production tends to reproduce itself even when this spatial pattern offers no particular advantage to the firms involved. The next two sections detail how these forces play out in two industries: footwear manufacturing and biotechnology.

4 Example 1: Footwear manufacturing[3]

Footwear manufacturing provides a particularly appropriate industry for examining these processes because small firms continue to account for a large portion of manufacturing, making the geography of entrepreneurship relevant to the distribution of the industry as a whole. Non-rubber shoe manufacturers turn leather into footwear through the application of more than 200 mechanized processes and a great deal of labor. Small firms dominate the industry; as late as 1991, roughly half of all firms still employed fewer than fifty workers (Raehse and Sharkley 1991). Two factors likely contribute to the continuing pervasiveness of small manufacturers. One, economic studies have repeatedly attributed rather insignificant cost savings to increasing scale (e.g., Szenberg, Lombardi and Lee 1977); particularly in the medium- and high-quality segments of the market, plants can operate efficiently with a small number of employees. Two, potential entrepreneurs face relatively low financial barriers to entry. With a modest deposit, anyone can lease equipment from the United Shoe Machinery Corporation for a small royalty on each pair of shoes

---

[3] Sorenson and Audia (2000) provide a far more detailed exposition on the forces driving geographic concentration in the U.S. footwear industry from 1940 to 1989. The empirical evidence reported in this section largely summarizes their findings.

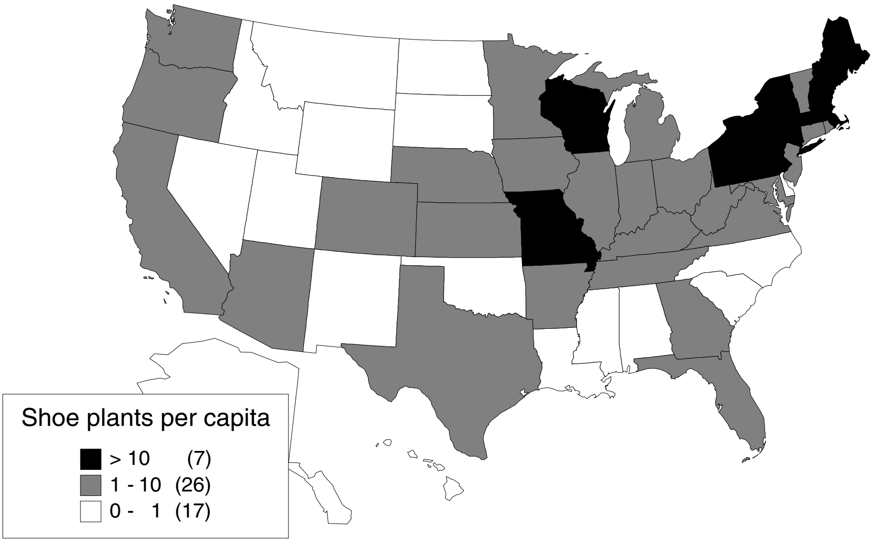


Fig. 1. Distribution of shoe plants in 1940

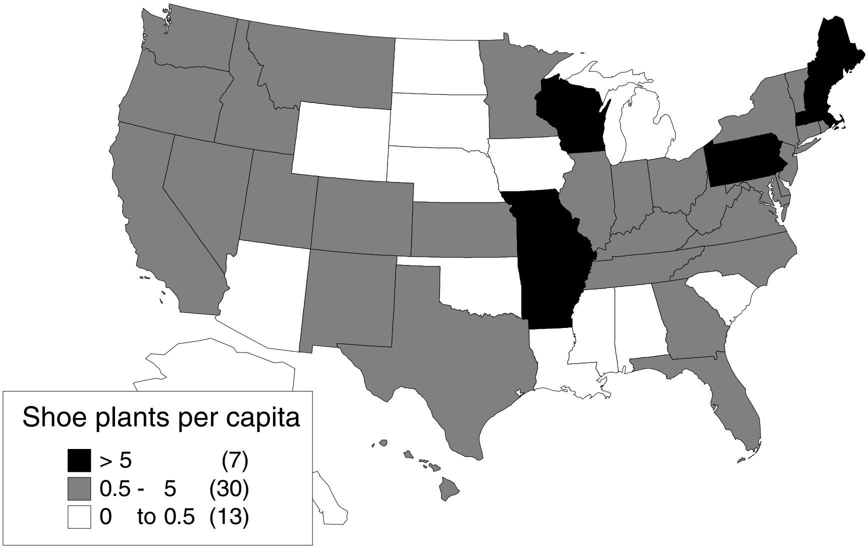


Fig. 2. Distribution of shoe plants in 1989

(just under 2% of total costs). The absence of strong scale economies and barriers to entry allows small, independent plants to continue playing an important role in this industry.

The shoe industry displays a high degree of spatial concentration. Figure 1 illustrates the geography of shoe production in 1940. Darker shadings denote states

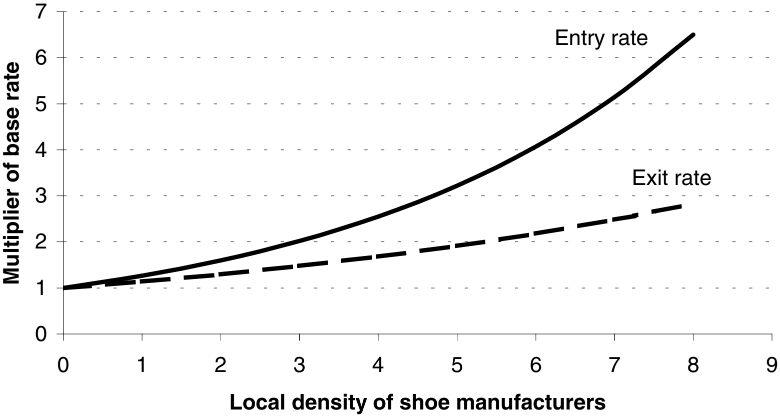


Fig. 3. Entry and exit rates of footwear manufacturers as a function of local density

with larger numbers of shoe plants per capita. Several states have no production, while the most crowded states, Massachusetts and New York, have 281 and 264 plants respectively in 1940. Figure 2 depicts the distribution of shoe plants per capita in 1989. Although the total number of plants declines due to an influx of imports, the states with heavy concentrations of plants in 1940 generally continue to have the heaviest concentrations in 1989.

Though the raw per capita counts of plants reveal strong regional heterogeneity in density, these graphs actually understate the concentration of production activity. First, these maps mask the concentration of production within states; even within states, plants typically cluster in small regions: around Boston in Massachusetts, near St. Louis in Missouri, and close to Milwaukee in Wisconsin. Moreover, individual towns even specialize in the types of shoes they produce. In Massachusetts, for example, Haverhill and Lynn produce women's shoes, while the South Shore makes men's shoes.

At least two processes could account for this geographic concentration. On the one hand, co-location may offer benefits to the firms that reside in these industrial districts, presumably externalities resulting from economies of agglomeration. Alternatively, social networks may constrain the geography of entrepreneurial opportunities such that new firms rarely arise in regions far from incumbents. These two possibilities suggest a critical test: If firms benefit from co-location, then the local concentration of plants in the industry should correlate positively with firm performance. Conversely, if companies do not benefit from locating near to one another, constraints on entrepreneurship likely drive the geographic distribution of industry.

To determine which of these two processes explained the stability of geographic concentration among U.S. footwear manufacturers, Sorenson and Audia (2000) estimated the relationship between the density of existing firms in a region and the rates of both firm failure and new firm founding. Figure 3 displays the relationships between average state-level failure and founding rates predicted by their results.[4]

---

[4] Density in this graph corresponds to the mean local density in the state, where local density refers to the plant-level sum of inverse distances to all other plants. The failure rate models estimated the

Failure rates clearly rise with the local density of firms – shoe manufacturers located in the densest states fail at nearly three times the rate of those in the most isolated ones – suggesting that economies of agglomeration cannot account for the geographic distribution of footwear production. Furthermore, additional analyses revealed that no identifiable subset of firms (e.g., larger firms, older firms, multi-unit organizations) benefited from locating near to other firms. These differential failure rates would lead to a rapid dispersion of production if new firms appeared at random on the map; however, concentration stimulates founding rates even more strongly – the densest states experience founding rates nearly seven times higher than the most remote ones. Since locating among dense concentrations of rivals offers no clear benefits, one might question why entrepreneurs (if unconstrained in their geographic choice) would locate there.[5] Social networks, however, can explain this phenomenon: Most entrepreneurs arise from the ranks of employees currently working in the industry; when they found new firms, they tend to do so in close proximity to where they live, thereby reifying the distribution of production.

Qualitative accounts of entrepreneurship in the shoe industry tell a consistent story. The biographies of the founders of various shoe companies all report that the individual had worked in the industry, usually as a plant manager, prior to starting their own firm. Moreover, when starting their new firms, these entrepreneurs usually located them in the same town as their previous employer, or in a neighboring one. Thus, a spin-off process with offspring landing close to their parents appears to account for the geographic concentration of non-rubber footwear production in the United States.

5 Example 2: Biotechnology[6]

The second fitting example comes from an examination of the dynamics of the biotechnology industry. On the one hand, biotechnology may seem different in virtually every respect from non-rubber footwear manufacturing; its highly skilled labor force primarily engages in innovation rather than in the production of a consumer product that has changed little over the last two centuries. On the other hand, the two industries have very similar structures when we look at the size distribution of firms. Small, young firms account for a large proportion of each population; hence, entrepreneurship plays an important role in the geography of both industries.

---

continuous hazard of firm closure and included controls for firm age, plant size, imports, exports, total domestic production, and the number of shoe firms at a national level. To compare these firm-level rates to the state-level founding models, I averaged the predicted failure rates across all firms in a state; this averaging actually somewhat truncates the relationship between the local concentration of production and failure rates. Using negative binomial regression, the founding rate models estimated the number of new firms founded in a state, explicitly accounting for imports, exports, total domestic production, state population, wages, the availability of leather suppliers, national density, lagged failures and time-invariant state effects (through the use of state-level fixed effects).

[5] Other possibilities clearly exist: a risk-return tradeoff, local institutional factors, etc. Sorenson and Audia (2000) consider each of these in detail and provide evidence that suggests that these factors cannot account for the dynamics of the shoe industry.

[6] The empirical evidence reported in this section largely reviews the more detailed findings summarized in Stuart and Sorenson (2003).

Like the shoe industry, biotechnology in the United States clusters in a relatively small number of regions. Biotechnology firms first appeared in force in San Diego, the San Francisco Bay Area and around Boston. Today, these regions remain home to the largest concentrations of biotech firms, though important clusters have also emerged near to Oakland, Seattle, and Washington, DC. The dispersion beyond its initial origins has been slow. Historical accounts have attributed this evolution to a spin-off process whereby early incumbent firms, such as Hybritech in San Diego, act as incubators for generations of future founders who locate their firms nearby their parent.

The biotechnology study sought to address the same issue as the shoe industry analysis: namely, do social networks, or externalities in production, account for the concentration of the industry. Once again, the researchers estimated the effects of the local density of firms in the industry on performance and entrepreneurship rates, though in this case time-to-IPO rather than firm failure provided the performance metric.[7] Stuart and Sorenson (2003) also extend the shoe industry analysis by looking not just at the proximity of firms in the biotech industry, but also at the local presence of skilled labor (biotech patent holders), venture capital firms and universities with leading biological sciences departments.

The results established three findings.[8] First, confirming the importance of access to vital resources, proximity to three of these four types of resources – incumbent biotech firms, venture capitalists and universities – stimulates the founding of new ventures. Second, the importance of proximity to these resources declines as the industry matures – possibly due to the emergence of industry-specific institutions (e.g., conferences, professional societies and industry associations) that extend the geographic reach of individuals' networks. Third, parallel to the shoe industry findings, locating near rivals hurts firm performance (i.e. increases the time-to-IPO).

Figure 4 helps us understand the implications of these results. The dark bars illustrate the rate at which new firms appear in each zip code, while the light bars depict the likelihood that an existing private firm goes public. As one can see in the chart, the regions with the highest entry rates have the slowest times to IPO; in fact, the lowest likelihood of an IPO occurs among firms located in South San Francisco, one of the densest concentrations of dedicated biotech firms.

---

[7] Few biotech firms had failed by the end of the observation period in 1996, so failure rates offer little statistical power in this sample; however, several studies of the industry have used time-to-IPO and IPO valuation as measures of performance since access to public markets represents an important stage in the lifecycle of these firms.

[8] Using negative binomial regression, the full analyses estimated the zip code-level count of the number of biotech companies founded each year, controlling for the demand for public biotech stocks, the national count of biotech firms, the local population, the age of the local biotech industry, as well as time-invariant state-level effects (using fixed effects). The time-to-IPO models estimated the continuous hazard of going public while accounting for the following factors: firm age, firm financing rounds, firm capital raised, firm patents, the demand for public biotech stocks, the number of biotech firms nationally, and the age of the local biotech industry.

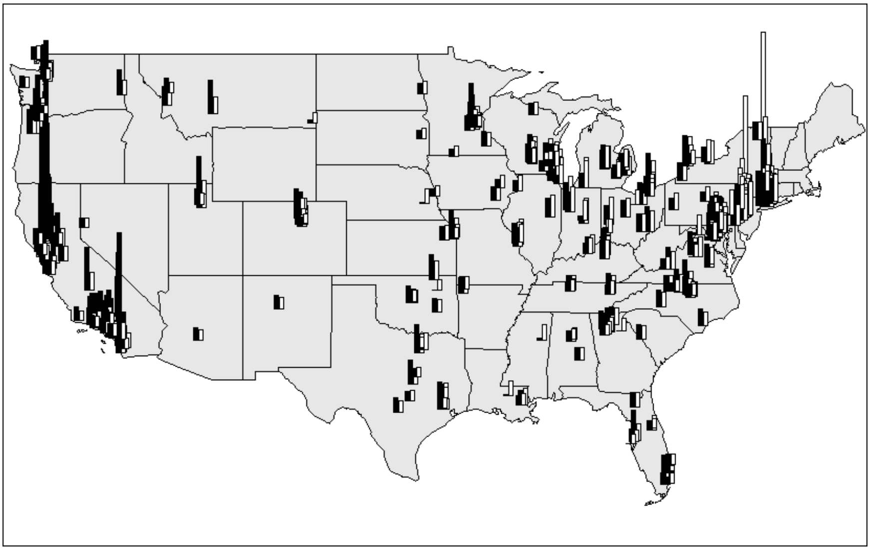


Fig. 4. Predicted entry (dark) and IPO rates (light) of biotech firms in 1995

6 Conclusion

Social networks play an important role in the entrepreneurial process, and in doing so also significantly influence the dynamics of the geographic dispersion of industry over time. Traditional explanations for geographic concentration have centered on the production efficiency of such spatial distributions, through either the minimization of transportation costs or economies of co-location. These accounts fail, however, to explain why entrepreneurs would continue to locate their firms in close proximity to rivals even when these locations promise lower average performance – the pattern found in both footwear manufacturing and biotechnology. These dynamics do accord with an understanding of how social networks influence the entrepreneurial process. Entrepreneurs primarily arise from within an industry because few outsiders have the deep social networks necessary to recognize new opportunities in the industry and to mobilize the intellectual, financial and human capital to realize their vision. These networks also bind entrepreneurs to the locations in which they reside because only there do they have the access to the resources and social support required to sustain their entrepreneurial ventures. Even though better locations exist, most entrepreneurs cannot exploit them due to these constraints.

Integration of these ideas with a closely related line of research, spin-offs, may further improve our understanding of industry spatial dynamics. Like the social networks based account, a growing literature on spin-offs also points to the importance of access to information in determining who founds new firms (for a review, see Klepper, 2001). Empirical work in this tradition tends to focus on how both environmental and organizational factors influence the likelihood that an employee will exit the firm to found their own venture. For example, Brittain and Freeman (1986) found that three factors increased the likelihood of a spin-off: a change in CEO, an acquisition, and slowed firm growth. Hence, this research importantly alerts us to the fact that not all firms in an industry equally stimulate the founding of new ventures, though it has left unstudied the ecological dynamics of regions. Combining these two lines of research offers the opportunity to enhance our understanding of the spatial dynamics of industries, as well as their evolution on other dimensions.

In addition to its contribution to theory, understanding the dynamics underlying industrial clusters can also inform public policy. Regional planners have shown interest in replicating the successes of established industrial districts in new locations. As manufacturing jobs shift to countries with lower wages, policymakers increasingly look to high tech districts as a stimulus for regional growth; for many, Silicon Valley represents the archetype for stimulating employment and economic growth. Regional and local governments around the world have launched initiatives in an attempt to duplicate this success in their own backyards. Despite this enormous investment, relatively few systematic empirical studies analyze the effect of geographic location on organizational viability. As a result, although urban planners have been informed by several detailed case studies of industrial districts, they lack systematic empirical evidence upon which to ground policy. The results of these studies suggest that firms, at least, do not benefit from agglomeration. Thus, one might question the wisdom of promoting agglomeration. Regardless, additional research must determine whether the benefits of co-location accrue to other stakeholders (e.g., to employees through higher wages).

Though the desirability of encouraging industrial concentration remains uncertain, these results offer clearer implications for how to pursue such a course of action. Traditional approaches to regional development point to the building of infrastructures, such as technical schools, efficient transportation routes, etc. as key steps favoring the formation of agglomerations. These results suggest instead that policymakers might enjoy greater success by seeding the pollination process through the recruitment of one or more successful companies to the region to 'fertilize' the area. Once the entrepreneurial process has started, it may become self-sustaining. Employees will leave the new organizations to create a second generation of ventures, and so on. Interestingly, although this process might benefit the community, these benefits probably come at the expense of any given firm that gets caught in these waves of creative destruction.

Although the geographic distribution of industry has received limited attention from sociologists, my colleagues and I find ample evidence to suggest that social networks play an important role in determining this distribution. Indeed, we believe that the availability of entrepreneurial opportunities largely drives the geographic distribution of industry. Although this research represents an early investigation along these lines, I believe this perspective can bring fresh insight to the analysis of industrial geography.

References

Akerlof GA (1970) The market for 'lemons': qualitative uncertainty and the market mechanism. Quarterly Journal of Economics 84: 488–500

Allen TJ (1977) Managing the flow of technology. MIT Press, Cambridge, MA

Arrow KJ (1962) The economic implications of learning by doing. Review of Economic Studies 29: 155–173

Baker W (1984) The social structure of a national securities market. American Journal of Sociology 89: 775–811

Blau PM (1977) Inequality and heterogeneity. Free Press, New York, NY

Bossard JS (1932) Residential propinquity as a factor in marriage selection. American Journal of Sociology 38: 219–224

Brittain JW, Freeman J (1986) Entrepreneurship in the semiconductor industry. Mimeo

Coleman JS (1990) Foundations of social theory. Harvard University Press, Cambridge, MA

Coleman JS, Katz E, Menzel H (1966) Medical innovation: a diffusion study. Bobbs-Merrill, New York, NY

Coval J, Moskowitz TJ (2001) The geography of investment: informed trading and asset prices. Journal of Political Economy 109: 811–841

Diamond C, Simon C (1990) Industrial specialization and increasing returns to labor. Journal of Labor Economics 8: 175–201

Festinger L, Schacter S, Back KW (1950) Social pressures in informal groups. Harper, New York, NY

Fox CR, Tversky A (1995) Ambiguity aversion and comparative ignorance. Quarterly Journal of Economics 110: 585–603

Fried VH, Hisrich RD (1994) Toward a model of venture capital investment decision making. Financial Management 23: 28–37

Garmaise MJ, Moskowitz TJ (2003) Confronting information asymmetries: evidence from real estate markets. Revew of Financial Studies: forthcoming

Granovetter MS (1973) The strength of weak ties. American Journal of Sociology 78: 1360–1380

Klepper S, Sleeper S (2000) Entry by spinoffs. Mimeo

Klepper S (2001) Employee startups in high-tech industries. Industrial and Corporate Change 10: 639–674

Kono C, Palmer D, Friedland R, Zafonte M (1998) Lost in space: the geography of corporate interlocking directorates. American Journal of Sociology 103: 863–911

Latour B (1989) Science in action. Harvard University Press, Cambridge, MA

Lazarsfeld PF, Merton RK (1954) Friendship as a social process: a substantive and methodological analysis. In: Berger M, Abel T, Page CH (eds) Freedom and control in modern society. Van Nostrand, New York, NY

Løvås B, Sorenson O (2003) Social capital, competition and resource mobilization. Mimeo

Marsden PV (1989) Homogeneity in confiding relations. Social Networks 10: 57–76

Marshall A (1890) Principles of economics. MacMillan, London

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annual Review of Sociology 27: 415–444

Nelson RR (1959) The simple economics of basic scientific research. Journal of Political Economy 67: 297–306

Porter ME (1990) The competitive advantage of nations. Free Press, New York, NY

Raehse SL, Sharkley MF (1991) Footwear. Fairchild Fashion and Merchandising Group, New York, NY

Rogers EM (1995) Diffusion of innovations. Free Press, New York, NY

Romer P (1987) Growth based on increasing returns to specialization. American Economic Review 77: 56–62

Rotemberg JJ, Saloner G (1990) Competition and human capital accumulation: a theory of interregional specialization and trade. Mimeo

Schumpeter JA (1942) Capitalism, socialism and democracy. Harper, New York, NY

Sorenson O, Audia PG (2000) The social structure of entrepreneurial activity: geographic concentration of footwear production in the United States, 1940–1989. American Journal of Sociology 106: 424–462

Sorenson O, Stuart TE (2001) Syndication networks and the spatial distribution of venture capital investments. American Journal of Sociology 106: 1546–1588

Stouffer SA (1940) Intervening opportunities: a theory relating mobility and distance. American Sociological Review 5: 845–867

Stuart TE, Sorenson O (2003) The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. Research Policy 32: 229–253

Szenberg M, Lombardi JW, Lee EY (1977) Welfare effects of trade restrictions: a case study of the U.S. footwear industry. Academic Press, New York, NY

von Thünen JH (1826). Der isolierte staat in beziehung auf landtschaft und nationalokonomie. G Fischer, Jena

Weber A (1909). Ueber den standort der industrien. J Verlag, Tübingen

Zipf GK (1949) Human behavior and the principle of least effort. Addison-Wesley, Reading, MA

Growing Silicon Valley on a landscape: an agent-based approach to high-tech industrial clusters[⋆]

Junfu Zhang

Public Policy Institute of California, 500 Washington Street, San Francisco, CA 94111, USA
(e-mail: zhang@ppic.org)

Abstract. We propose a Nelson-Winter model with an explicitly defined landscape to study the formation of high-tech industrial clusters such as those in Silicon Valley. The existing literature treats clusters as the result of location choices and focuses on how firms may benefit from locating in a cluster. We deviate from this tradition by emphasizing that high-tech industrial clusters are characterized by concentrated entrepreneurship. We argue that the emergence of clusters can be explained by the social effect through which the appearance of one or a few entrepreneurs inspire many followers locally. Agent-based simulation is employed to show the dynamics of the model. Data from the simulation and the properties of the model are discussed in light of empirical regularities. Variations of the model are simulated to study policies that are favorable to the high-tech economy.

Keywords: Silicon Valley – Agent-based simulation – Industiral clusters

JEL Classification: L11, R12

> *Do not regard Silicon Valley as some sort of economic machine, where various raw materials are poured in at one end and firms such as Apple and Cisco roll out at the other, but rather as a form of ecosystem that breeds companies: without the right soil and the right climate, nothing will grow.*
>
> – The Economist, March 29, 1997

[⋆] This paper has been presented at the 9th International Schumpeter Society Conference in Gainesville, Florida, the Western Economic Association's 77th Annual Conference in Seattle, Washington, the 24th Annual Research Conference of the Association for Public Policy Analysis and Management in Dallas, Texas, and the Workshop on Industry and Labor Dynamics: The Agent-Based Computational Economics Approach in Turin, Italy. I would like to thank Rob Axtell, Giovanni Dosi, Olav Sorenson, and an anonymous referee for their comments, suggestions, and encouragement. I am grateful to Nikesh Patel for his superb assistance.

1 Introduction

Silicon Valley is the most salient example of high-tech industrial clusters. Public policy makers throughout the world would like to learn the secrets of Silicon Valley in order to build their own high-tech economies. The existing literature on industrial clusters, which traces back to Marshall (1920), focuses on the way in which firms benefit from locating in a cluster; it suggests that once a cluster comes into existence, it tends to reinforce itself by attracting more firms. However, a more important question is how to reach this critical mass in the first place. In contrast to the literature, evidence suggests that entrepreneurs rarely move when they establish high-tech start-ups (Cooper and Folta, 2000). This contradicts the notion that location choice analyses lead entrepreneurs to a high-tech cluster.

A high-tech industrial cluster such as Silicon Valley is characterized by concentrated entrepreneurship. Following Schumpeter, we emphasize the fact that "the appearance of one or a few entrepreneurs facilitates the appearance of others" (Schumpeter, 1934). We propose an agent-based computational model to show how high-tech industrial clusters could emerge in a landscape in which no firms existed originally. The model is essentially a spatial version of the Nelson-Winter model: Boundedly rational agents are scattered over an explicitly defined landscape. Each agent is endowed with some technology, which determines his firm's productivity (if he has one). During each period of time, an agent with no firm would make a decision as to whether he wants to start one. This decision is mostly affected by the behavior of his social contacts, who are all his neighbors. If an agent's neighbors are successful in their entrepreneurial activities, the agent is more likely to found a firm himself. An entrepreneur makes business decisions according to some rules of thumb. When an agent does start a firm and begin to make a profit, he spends part of his profit on R&D in order to improve his productivity, part on imitating other firms' technology, and the rest on capital accumulation. Entrepreneurs who lag behind in the Schumpeterian competition lose money and eventually fail; however, it is possible that they learn from their failures and try again.

We use an agent-based simulation to show that Silicon Valley-type industrial clusters will emerge spontaneously on the landscape. In addition, the model exhibits the following properties: 1) First mover's advantage: the first firm has a better chance to survive and grow; a region in which firms enter the market early tends to capture a large piece of the industry. 2) Path dependence: the more firms a region has, the more it tends to have; once a cluster is formed, it can hardly be toppled. 3) Clustering of entrepreneurship: firms are continuously forming and dying within clusters. 4) Clustering of innovations: the productivity in clusters is much higher than elsewhere because of the collective learning within clusters through innovation and imitation.

Data from the simulation and the properties of the model are discussed in light of empirical regularities. We also explore variations of the model in order to study the factors that determine the location of emerging clusters. We learn two lessons from the model. First, the conventional knowledge-spillover literature may only tell part of the story; the contagion of entrepreneurship through peer effects seems to be an important mechanism through which high-tech industrial clusters emerge and grow. Second, while many scholars have recognized the importance of "seed capital" for a budding high-tech regional economy, our model suggests that "seed entrepreneurs" may be even more important because they serve as local role models and inspire new entrepreneurs.

The main contribution of this paper is the application of a novel methodological approach to study the formation of industrial clusters and related policy issues. While agent-based computational economics has introduced new tools for economic analysis, we have yet to see applications of this approach in policy analysis. This paper intends to fill in the blank. As our model shows, the agent-based approach is particularly useful for dealing with dynamic economic systems. It is also flexible for testing the effects of alternative assumptions.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 presents the model. The agent-based simulation of the model is described in Section 4. The last section concludes with some remarks.

2 Related literature

Our model builds upon the intersection of several strands of the literature.

2.1 The Nelson-Winter paradigm

The term evolutionary economics has been used in different contexts and by various groups of economists, including institutional economists, evolutionary game theorists, and those who follow Nelson and Winter (1982). The Nelson-Winter paradigm of evolutionary economics is a synthesis that integrates three sources of work: Simon's concept of "bounded rationality," Nelson's and others' work on invention and innovation (following Schumpeter), and Alchian's and Winter's work on "natural selection" in economic evolution.

A typical Nelson-Winter evolutionary model defines the state of the industry by a list of firm level state variables such as physical capital and productivity. Minimum environmental characteristics are specified, which may include input and output conditions, the space of innovations, and the way innovative search takes place. Based on these, the activities of the industry in the current period are calculated and in turn generate new values of state variables for the next period. New technology or new rules may be adopted if they increase expected profitability. Calculations are conducted for a series of periods and are used to study the evolution of technology, the application of rules, and other characteristics of the industry (Anderson et al., 1996). The Nelson-Winter paradigm provides a powerful and general theoretical framework for studying variety-creation and variety-selection within a given economic sector. Up until now, researchers within this tradition have worked with very simple examples; the potential of the general schema is far from fully exploited.[1]

---

[1] Nelson and Winter (1982) is a classic reference of their paradigm. Nelson (1995) gives a review of more recent developments in this area and related fields.

2.2 Agent-based computational economics

Agent-based computational economics uses computer simulation to study the economy as an evolving system of autonomous interacting agents. Researchers in this field try to understand why certain macro-level regularities emerge and persist in decentralized market economies, despite the absence of any forms of centralized coordination. For example, the agent-based computational approach has been applied to the study of business cycles, trade networks, market protocols, the formation of firms and cities, and the diffusion of technological innovations. Computer programs are used to demonstrate constructively how those macro-level regularities might arise from the bottom up through repeated local interactions of autonomous agents. A methodological advantage of agent-based computational economics is that it enables social scientists to do “laboratory experiments” to test a theory, because computational models usually can be modified quite easily to study alternative socioeconomic structures and examine their effects on economic outcomes (Tesfatsion, 2001).

Using poker chips on a checkerboard, Schelling (1971) simulates the dynamics of racial housing segregation, which is generally recognized as the pioneering application of this approach in the social sciences. In a ground-breaking work, Epstein and Axtell (1996) investigate how social structures and group behaviors arise from the interaction of individuals. With agent-based simulations, they show how fundamental collective behaviors such as group formation, cultural transmission, combat, and trade can emerge from the interaction of individual agents following simple local rules. Axtell (1999) presents a model in which heterogeneous agents form firms. Agents join firms or start new firms when it is advantageous for them to do so. As firms grow, agents have less incentive to supply their efforts and tend to become free riders, which causes large firms to decline. At the micro level, firms grow and perish; at the aggregate level, the model produces data about firm sizes, growth rates, and related aggregate regularities that parallel empirical findings.

Epstein (1999) characterizes the agent-based computational approach with the following features: heterogeneous agents in terms of preferences, culture, social networks, etc.; decentralized autonomous behavior of agents; explicitly defined space; local interactions among agents; and bounded rationality. Because of these features, agent-based modeling is particularly useful when the population is heterogenous, when interactions among agents are complex and nonlinear, and when the space is crucial.

2.3 Industrial clusters

An industrial cluster is a geographic area in which many firms in an industry are located and interact with each other through competition and cooperation. Economists’ interest in industrial clusters traces back to Marshall (1920), but we have seen a revival of this interest recently (see, e.g., Arthur, 1990; Krugman, 1991; Porter, 1998). This line of research emphasizes the net benefits to firms located in a cluster, which are determined by the benefits and the costs of agglomeration. Sources of benefits include pooled labor forces, specialized suppliers, access to capital, proximity to customers, and knowledge spillovers, while diseconomies of agglomeration stem from increased competition, congestion costs, and knowledge expropriation. If positive net benefits are expected from an industrial cluster, new entrants tend to arise, further enhancing the geographic concentration.

Those works that focus on net benefits from agglomeration treat clusters as the result of firms' locational choices. Yet it is not clear whether firm owners or entrepreneurs engage in such searching and comparing exercises. Moreover, high-tech start-ups might have concerns different from those of manufacturing firms. Cooper and Folta (2000) point out that the primary determinant of a high-tech start-up is the prior location of its founder. In fact, entrepreneurs seldom move once they decide to start their new firms. This is understandable because, by staying where they are, entrepreneurs can utilize their existing network to seek investors, employees, customers, suppliers, and advisors; they can start on a part-time basis and defer full commitment until the start-up becomes more promising; and they may want their spouses to keep their current jobs. Given that entrepreneurs do not move, one doubts whether they intentionally take advantage of the benefits of geographic clusters.

Using data on U.S. manufacturing employment in the period of 1860-1987, Kim (1995) shows that industry localization patterns are negatively correlated with characteristics associated with external economies. In particular, high-tech sectors, which are believed to have more positive externalities than other sectors, are less agglomerated. This is inconsistent with the location choice literature.

Social scientists have long noticed that clusters, of individuals or firms, are the result of two types of behavior. One is a sorting process. For example, individuals' racial preferences could lead to housing segregation in which clusters of black or white residents are formed. The other is a behavior-adapting process. For example, smokers can convert their non-smoking friends into smokers, resulting in clusters of smoking behaviors. The existing literature on industrial clusters has studied the sorting process in which firms choose to locate close to other firms, but has neglected the other process. We argue that entrepreneurship may be contagious and that a person surrounded by entrepreneurs is more likely to start a firm himself. This provides an alternative theory of the formation of industrial clusters.

2.4 The Schumpeterian entrepreneur

The "Irish banker in Paris," Richard Cantillon, acknowledged by many historians as the first great economic theorist, first recognized the important role of entrepreneurs in economic life in the 18th century. The concept of the entrepreneur then appeared in the writings of many French economists, including Quesnay, Turgot, and Say. Yet within the British tradition, the dominant classical school made no distinction between capitalists –who provide the means for investment in production – and entrepreneurs – who explore possibilities of innovation, seek profitable opportunities, and assume risks. Thus the followers of Smith and Ricardo excluded the entrepreneur from economic analysis.

Today's economists learn the theory of entrepreneurs mainly from Schumpeter (1934). Schumpeter starts by describing the economic system as a circular flow within a Walrasian-like general equilibrium. To him, economic development is driven by the activities of a class of entrepreneurs who take it upon themselves to disrupt the circular flow by introducing new products, reorganizing labor forces and capital, and rearranging the processes of business life in the hope of making a profit from the disequilibria they create. To address the question of what drives entrepreneurs to exercise their talents, Schumpeter might have given the most romantic reasoning in economics: he states that entrepreneurs choose their way of life because of the dream and the will to found a private kingdom, the will to conquer, the impulse to fight, to prove oneself superior to others, to succeed for the sake not of the fruits of successes but of success itself, and finally the joy of creating, getting things done, or simply of exercising one's energy and ingenuity. Schumpeter uses his concept of the entrepreneur to explain business cycles. The introduction of new and untried products and processes causes "disturbances"; these disturbances that appear "in groups or swarms" constitute business cycles. Entrepreneurial activities appear in clusters because "*the appearance of one or a few entrepreneurs facilitates the appearance of others, and those the appearance of more, in ever-increasing numbers* [Schumpeter's italics]".

Schumpeter's theory of entrepreneurs has been renowned and influential. However, it is fair to say that its influence has remained outside of neoclassical economics. Schumpeter's entrepreneur is by definition an equilibrium-disturbing figure; his entrepreneurial activities constantly interrupt the tendency toward equilibrium in the economic system. Therefore, since neoclassical economics focuses on equilibrium analysis, there is no room for Schumpeter's entrepreneur. By contrast, the Schumpeterian entrepreneur plays a crucial role in our model.

3 The model

Consider an $N \times N$ lattice graph, $\Lambda_N = (V, E)$, with periodic boundary conditions. V and E are the sets of vertexes and edges, respectively. Each vertex $i \in V$ represents an agent. An agent i is endowed with some human capital (technology) h_i.

At time 0, all agents are born and each agent's endowment of human capital is determined by a random draw: h_i^0.

In each period of time, an agent with no firm has to make a decision as to whether he wants to be an entrepreneur and start a firm. If he wants to do so at time t, he will raise some money to buy capital K_i^t. If agent i has a firm, his production function is

$$Y_i^t = h_i^t (K_i^t)^\alpha, \qquad \alpha < 1, \tag{1}$$

otherwise, he produces nothing: $Y_i^t = 0$. Capital is always obtainable at a fixed unit cost c. For simplicity, we deal with the single-factor production and do not bother with labor. This simplification may be understood in this way: each unit of capital is attached with a certain amount of labor according to a fixed capital-labor ratio and abundant labor is supplied at a constant price which is already included in c.

Aggregate supply in this industry is

$$S^t = \sum_i Y_i^t. \tag{2}$$

Aggregate demand D^t is given exogenously. Market price at time t is decided by

$$P^t = \frac{D^t}{S^t}. \tag{3}$$

If agent i produces, his profit is

$$\pi_i^t = P^t Y_i^t - cK_i^t. \tag{4}$$

Agents are boundedly rational; they act according to some rules of thumb. When an agent makes some profit, he will put part of it into R&D and spend the rest on capital accumulation. The R&D fund will be split again, with part of it being spent on technological innovation and the remainder on technological imitation. Each agent is born with two uniformly distributed random numbers $\gamma_i, \lambda_i \in (0, 1)$, which he takes as rules that govern his spending on technological innovation and imitation. If agent i makes profit $\pi_i^t > 0$, he puts aside $\gamma_i \pi_i^t$ for R&D. Among that amount, $\lambda_i \gamma_i \pi_i^t$ goes to technological innovation. Let IN_i and IM_i denote i's spendings on innovation and imitation, respectively. Then,

if $\pi_i^t > 0$, then

$$\begin{aligned} IN_i^{t+1} &= IN^t + \lambda_i \gamma_i \pi_i^t, \\ IM_i^{t+1} &= IM^t + (1 - \lambda_i)\gamma_i \pi_i^t, \\ K_i^{t+1} &= K^t(1 - d) + (1 - \gamma_i)\pi_i^t; \end{aligned}$$

if $\pi_i^t \leq 0$, then

$$K_i^{t+1} = K^t(1 - d) + \pi_i^t. \tag{5}$$

Here $d > 0$ represents the rate of capital depreciation.

Technological innovation and imitation are costly. In addition, the larger a firm is, the more costly it is to improve its technology. Whenever agent i's spending on innovation exceeds $f(K_i)$, he gets a chance to draw a new h_i from the distribution of technological opportunities $F(h, t)$. This distribution function is independent of i's current technology, but its mean increases with time. If the new draw is greater than the old one, he will adopt the new one. Whenever agent i's spending on imitation exceeds $g(K_i)$, he gets a chance to copy the best technology from neighboring firms. It is more costly for large firms to upgrade technology, so $f'(\cdot) > 0$ and $g'(\cdot) > 0$. i's neighboring firms are those started by surrounding agents:

$$B_i = \{j | d(i, j) \leq 2 \text{ and } K_j > 0\}, \tag{6}$$

where $d(i, j)$ is the distance between i and j, which is defined as the number of edges that constitute the shortest path between i and j. Therefore, the mechanism of innovation and imitation can be summarized as follows:

$$\begin{aligned}
&\text{If } IN_i^t \geq f(K_i^t), \text{ then } h_i^{t+1} = \max\{h_i^t, h_i'\}, \text{ where } h_i' \sim F(h,t),\\
&\qquad \text{and } IN^{t+1} = IN_i^t - f(K_i^t);\\
&\text{if } IM_i^t \geq g(K_i^t), \text{then } h_i^{t+1} = \max\{h_i^t, \max_{i' \in B_i} h_{i'}^t\},\\
&\qquad \text{and } IM^{t+1} = IM_i^t - g(K_i^t). \qquad (7)
\end{aligned}$$

If the firm simultaneously gets a random draw and a copy of the best technology in the neighborhood, the better one is adopted.

In addition, entrepreneurs learn from failures. Each time an entrepreneur fails, which means he keeps losing money and eventually does not have enough capital to operate, he earns a chance $\rho > 0$ to copy the best technology from neighboring firms. For the sake of parsimony, we use a single parameter, h, to represent technology, which should be understood as a combination of both management skills and production technology. A failed entrepreneur is likely to learn some management skills from the practices of nearby successful entrepreneurs. Similarly, he will likely recognize the better production technology used by neighboring entrepreneurs. This opportunity for a failed entrepreneur to copy a better technology from surviving firms may be interpreted as a chance for "zero-cost imitation." In this sense, we are assuming that imitation is easier for a re-starter than for an incumbent firm. Previous studies have shown that incumbent firms are less likely to adopt radical innovations because it is more costly for them to shift to a different technology standard (Foster, 1986; Christensen, 1997). But a failed entrepreneur who starts up a new firm faces no such costs.[2]

An agent's decision on firm-founding reflects his perception of risk and his evaluation of profitability. In turn, his attitude is affected by other agents in the society. We assume that social distance is proportional to physical distance and an agent's behavior is largely influenced by close neighbors. If many of his neighbors are entrepreneurs who make a lot of money, he will see the profitable opportunity and also get a psychological boost from their success. Hence, he is likely to choose to be an entrepreneur himself; otherwise, it is less likely that he will do so. A distant successful entrepreneur has smaller effects on an agent's decision. Specifically, the probability that an agent chooses to start a firm is defined as follows:

$$\begin{aligned}
&\Pr(K_i^{t+1} > 0 \mid K_i^t = 0) = \phi(K_{j_1}^t, K_{j_2}^t, ...),\\
&\text{and } \frac{\partial \phi}{\partial K_j^t} \geq 0, \forall j \neq i; \;\; \frac{\partial \phi}{\partial K_{j_x}^t} \geq \frac{\partial \phi}{\partial K_{j_y}^t} \text{ if } d(i, j_x) \leq d(i, j_y). \qquad (8)
\end{aligned}$$

For simplicity, we have assumed that a failed entrepreneur has no negative effect on another agent's decision. Casual observation of the real world helps justify

---

[2] For example, it is quite easy for a fresh starter to imitate Amazon.com, but not that easy for a conventional bookstore because its on-line service could hurt the position of its physical store. In this sense, it costs less for a failed bookstore owner to imitate Amazon's technology.

this asymmetry between the social effects of success and failure. For example, the limited liability corporation system creates an asymmetry between success and failure: a successful entrepreneur is usually worth millions but a failed one almost never loses much. Also, the media always gives more attention to successes than to failures, which further magnifies the relative psychological impact of successes. Research shows that even in the peak years of the Internet revolution, a large number of high-tech firms went out of business in Silicon Valley (Zhang, 2003). However, even the local media rarely covered such failures.

4 Agent-based simulation

Our dynamic model represents a complex Markov Process. To analyze it rigorously is a formidable task. We will proceed with an agent-based simulation to learn the properties of the model.

4.1 Parameterization of the model

One way to calibrate our model is to search for a set of parameter values, using methods such as the Genetic Algorithm, so that the outcomes of the model replicate some pre-selected empirical regularities. Since we will focus on the qualitative properties of the model, such a sophisticated method seems unwarranted. Instead, we take a simpler approach, that of picking a set of "reasonable" parameters through a few trials on the computer program. As you will see, the model works fine. This partly proves the robustness of our basic setup. The model is parameterized as follows:

- $N = 100$. That is, we have a population of 10,000 agents.
- Technology h_i^t is drawn in the way such that $\sqrt{h_i^t}$ follows $U(0, 1 + \frac{t}{2{,}500})$, a uniform distribution on $(0, 1 + \frac{t}{2{,}500})$. That is,

$$F(h,t) = \left(\frac{2{,}500h}{2{,}500 + t}\right)^2. \tag{9}$$

  This looks like a truncated normal distribution, which makes it difficult to attain a very efficient technology. Over time, the distribution expands to the right. This implies that a draw today is expected to yield a better technology than a draw yesterday. Therefore, the investment in technology, just as with the investment in capital, also depreciates.
- When agent i decides to start a firm, he simply takes money out of his savings and acquires capital $K_i = 5k$, where $k \in U(0, 1)$. We impose an arbitrary minimum capital requirement such that, if $K_i < 0.1$, the firm has to be shut down. The diminishing return to capital is captured by $\alpha = 0.995$.
- Aggregate demand is given exogenously to replicate the evolution of an industry that grows rapidly at an early stage and loses its momentum over time. We start

with $D^0 = 10$. Its growth rate is declining over the life cycle of the industry. Specifically, we define

$$g^t = \begin{cases} 0.03 & \text{if } D^t < 2{,}000; \\ 0.02 & \text{if } 2{,}000 \le D^t < 10{,}000; \\ 0.01 & \text{if } 10{,}000 \le D^t < 20{,}000; \\ 0.005 & \text{if } D^t \ge 20{,}000. \end{cases} \tag{10}$$

We also define a cyclical parameter as $b^t = \frac{1}{100} \sin\left(\frac{2\pi t}{40}\right)$, which simulates business cycles that span 40 periods (quarters).

In addition, we introduce random demand shocks in the form $\epsilon^t \in U(-0.01, 0.01)$.

The dynamics of aggregate demand follows

$$D^{t+1} = D^t(1 + g^t + b^t + \epsilon^t). \tag{11}$$

- If agent i is not producing at time t, the probability that he will choose to do so is

$$\begin{aligned} &\Pr(i \text{ starts a firm} \mid K_i^t = 0) \qquad (12) \\ &= \frac{1}{25} \sum_{j|d(i,j)=1} \frac{K_j^t}{a_j^t} + \frac{1}{50} \sum_{j|d(i,j)=2} \frac{K_j^t}{a_j^t} + \frac{1}{50{,}000}, \end{aligned}$$

  where a_j^t is the age of firm j at time t. This says that i may choose to be an entrepreneur independently with a low probability; if his neighbors accumulated a great deal of capital in a short time, he is more likely to found a firm himself. Notice that closer neighbors have larger effects on an agent's choice and distant entrepreneurs have no effects.
- Profitable firms spend money on R&D and try to improve their technology through innovation and imitation. The parameters that affect the costs of R&D activities are $f(K_i) = 0.1(K_i)^3$ and $g(K_i) = 0.3(K_i)^3$. The chance of learning from failure is $\rho = 0.1$.

4.2 Main results

We start our simulation with a blank landscape with no firm. In this case, the emergence of the first entrepreneur is a pure chance event. He may not be an agent endowed with superior technology, but one thing is certain, he will make a large profit for his entrepreneurial move. Once the first entrepreneur emerges on the landscape, many of his neighbors will recognize the opportunity and follow suit; at the same time, others may start firms by chance. Those firms that make profits will upgrade their technologies. As more firms are founded and technology is improved, the market price for the product decreases sharply. Many new firms are born around profitable firms. Before long, one or more clusters form in certain regions. Some firms are forced to exit because they cannot keep up with others in technological progress, which may result from lower R&D expenditure or continuous unlucky

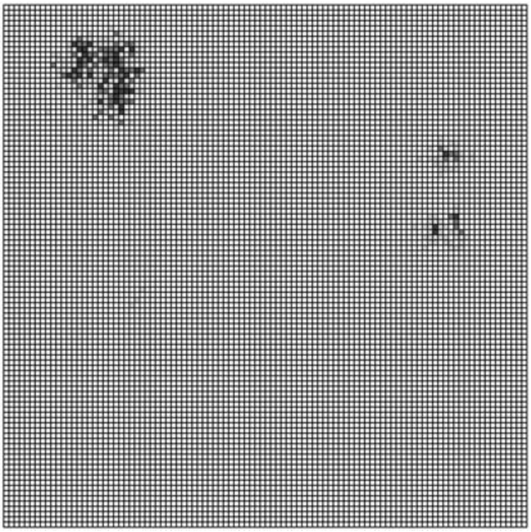


Fig. 1. A snapshot of industrial clusters

draws from the distribution of technology. In the long run, we see a spatial pattern of industrial clusters as shown in Figure 1.[3] In Figure 1, a green cell represents a small firm ($K_i \leq 10$); a red cell represents a medium firm ($10 < K_i \leq 100$); and a blue cell is a large firm ($K_i > 100$). A cluster generally hosts firms of all types.

In this model, firms in a cluster do benefit from knowledge spillovers as they imitate better technologies possessed by nearby firms. However, we see that entrepreneurs do not move and they do not intentionally seek the benefits from knowledge spillovers. In fact, if it is cheap to improve the technology through independent research, industrial clusters still tend to emerge even if we shut down the channels for the inter-firm transmission of technology.

Figure 2 shows the dynamics of market price. When the first firm enters the market, price is high. As capital is accumulated and production is expanded, market price is driven down quickly. Competition through entry of new firms continuously pushes the product price down to the cost of production. There is a cyclical pattern in the price series, which reflects the cyclical movement we have built in the demand.

Figure 3 shows the firm size distribution. Large firms are rare; small and medium firms dominate the industry. (Here we use output level to measure firm size. An alternative measure, capital stock, gives similar qualitative results.)

Since Gibrat's work in the early 1930s, it has been common practice to fit firm sizes with lognormal distributions (Sutton, 1997). A standard justification for the distribution is the so-called "law of proportional effect," which postulates that firms grow at random rates independent of firm size. This has now become well known as "Gibrat's law." A lognormal distribution is skewed to the right, meaning that firm sizes are concentrated on smaller values; in particular, the mean firm size is

---

[3] Interested readers may want to try the simulation by themselves. A Java Applet is available from the author upon request.

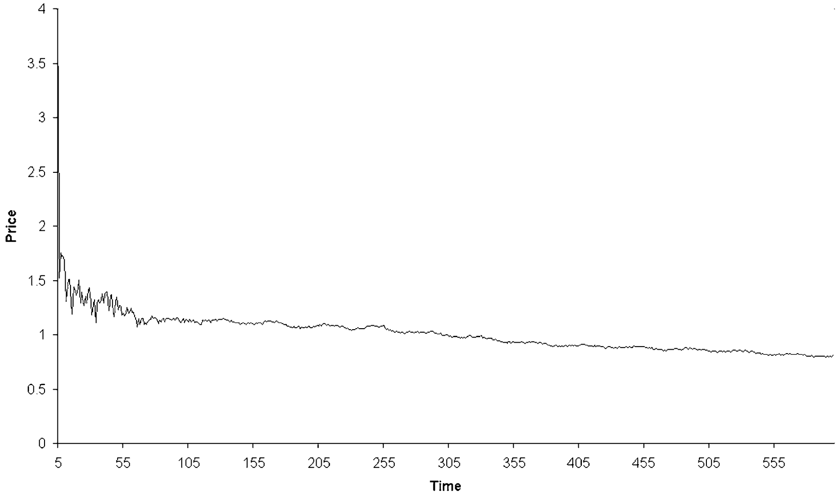


Fig. 2. Price series

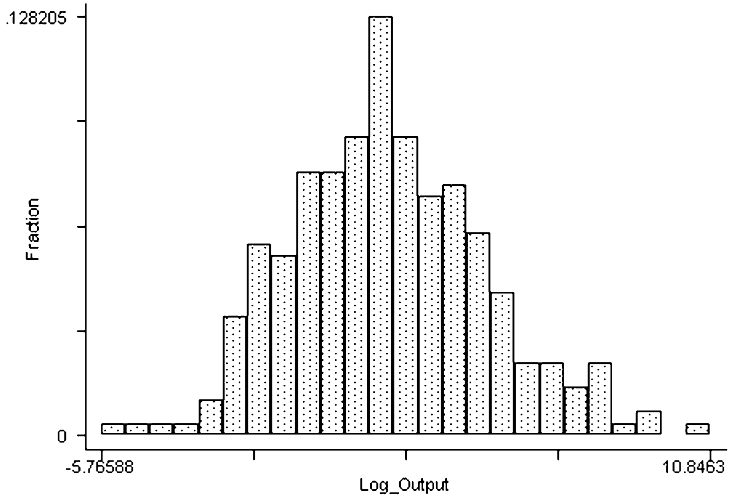


Fig. 3. Firm size distribution

larger than the median firm size, and both are larger than the modal firm size. By definition, a lognormal distribution of firm size implies a normal distribution of log firm size. Figure 3 roughly corresponds to a normal distribution.

The firm size distribution, especially its upper tail, has often been described by the Pareto law (Ijiri and Simon, 1977; Axtell, 2001):

$$sr^{\beta} = M, \tag{13}$$

where s is the size of a firm, r is its rank in an industry (or an economy) with the largest firm ranked 1, and β and M are constants. The power law implies a linear

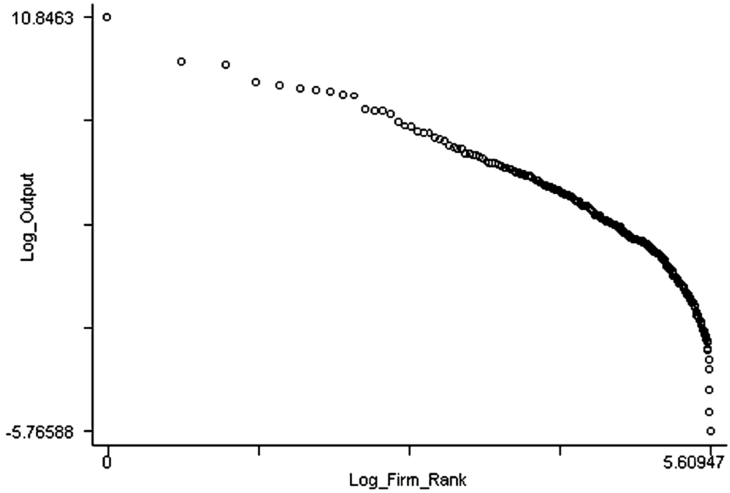


Fig. 4. Firm size-rank plot

relationship between log firm size and log firm rank:

$$\log s = \log M - \beta \log r. \tag{14}$$

Figure 4 plots firm size over firm rank. Cutting off the lower quartile of the sample, we fit a straight line to the remaining data and obtain:

$$\begin{aligned}\log(firm\_size) &= \underset{(0.125)}{12.33} - \underset{(0.028)}{2.11}\log(firm\_rank), \quad R^2 = 0.97.\end{aligned} \tag{15}$$

It is almost a perfect fit. To compare this with reality, we do the same exercise for the 211 U.S. high-tech firms on the list of Fortune 1000 largest firms. The firm size-rank plot is presented in Figure 5. Since this data is already truncated from below, we fit a straight line to the whole sample. The results are:

$$\begin{aligned}\log(firm\_size) &= \underset{(0.086)}{13.14} - \underset{(0.019)}{1.09}\log(firm\_rank), \quad R^2 = 0.94.\end{aligned} \tag{16}$$

We see that the real data also fits a straight line very well, although its slope is smaller.

The curvature in Figure 4 corresponds to a feature that is repeatedly observed in real data. Ijiri and Simon (1977) propose two possible interpretations for the "departure" from the Pareto distribution: autocorrelation in firm growth rates and the effects of mergers and acquisitions. Our model does not allow for mergers and acquisitions, but we do have autocorrelated firm growth. Figure 5 only plots the upper tail of the real data, which gives no indication as to how the lower tail behaves. However, based on what we know, we are able to get a rough idea about the profile of the complete sample. Assume the smallest high-tech firm has an annual revenue of $0.01 million. By equation 16, we predict its rank is higher than 11.7 million.

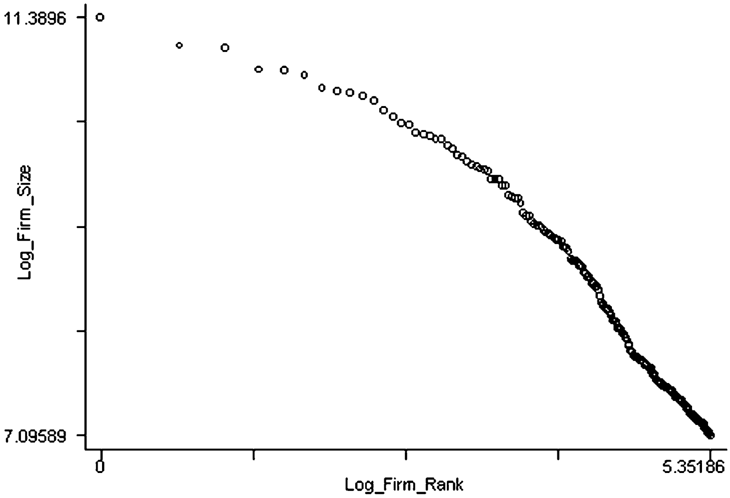


Fig. 5. Size-rank plot for fortune 1000 high-tech firms

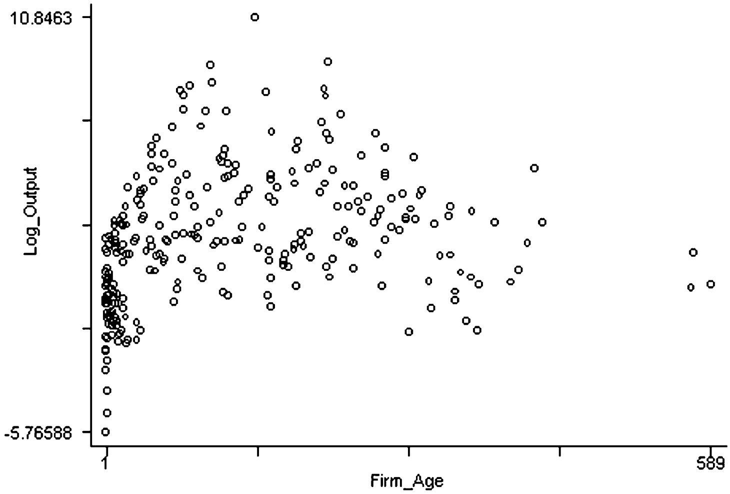


Fig. 6. Firm size-age plot

That rank is even larger than the total number of U.S. firms.[4] Therefore, the lower end of the real data must bend down as do the simulated data.

By construction, a firm's growth rate is related to its size in our model. In particular, a small firm is assumed to be able to upgrade its technology more easily. This is a violation of Gibrat's law. However, our model is able to generate firm size data that is qualitatively similar to empirical findings. It reminds us that Gibrat's law is only one of the parsimonious interpretations of empirical data.

---

[4] According to the Census Bureau, the total number of U.S. establishments in 1999 was 7,008,444.

Figure 6 plots firm size over firm age. Interestingly, there is not a strong positive correlation between firm size and age. In fact, the largest firms are all relatively young. This is because once an old firm reaches a certain size, it slows down in upgrading technology due to the high cost. But, remember, the distribution of technology moves to the right. A newcomer draws technology from a better space, and so it is more efficient and will outgrow older firms under the same market condition. This reflects what happens in the high-tech sector in the real world. For example, in Silicon Valley, more than half of the top 40 technology firms in 2002 were not even founded two decades ago; only four out of 40 largest firms in 2002 were survivors from the top-40 list in 1982 (Zhang, 2003).

The model also exhibits the following properties:

The first mover's advantage

At the firm level, the first few entrepreneurs tend to make a lot of money and have good chances to grow into large firms. However, their survival is not guaranteed. Some late comers may be endowed with better technology that can drive the pioneers out of the market; the first mover may follow a rule that spends very little profit on R&D and eventually lags behind in the Schumpeterian competition; the first mover may be so unlucky that his research efforts fail to generate superior technology before his followers do. Therefore, the first mover does enjoy some advantage, but only in a probabilistic sense.

At the economic region level, an area that enters the market early (with a few firms already operating at the early stage) tends to capture a large piece of the whole industry. In a variation of the model, we differentiate regions by assigning to them different innovation spaces. In one region, firms search new technology in $(0, 1)$, but in another, firms are only allowed to innovate in $(0, 0.99)$. We find that if the disadvantaged region first occupies the market, its first mover's advantage can overcome the technological disadvantage and sustain the regional economy for a long time. The reason is that firms in the disadvantaged region can innovate and learn from each other and approach their potential quickly. They drive the product price to a low level and leave a slim profit margin for new firms. Not all agents in the advantaged region are endowed with superior technology. Even if an agent has a very advanced technology to start with, he may make little profit, and an economic downturn may drive him out of business. Although the bigger firms in the disadvantaged region also lose money in the downturn, they will not go bankrupt and will recover during the next upturn.

Path dependence

A region with many firms, on the one hand, will have more agents thinking of starting up new firms, and on the other hand, will become more advanced in technology because of R&D. This property of "increasing return" tends to lock the development of the industry into certain regions. Once clusters are established, other regions have little chance to catch up. In the real word, for example, it is very unlikely that other regions can surpass Silicon Valley in the semiconductor industry.

Clustering of entrepreneurship

Within the clusters, firms enter and exit the industry constantly. People in clusters try new ideas and found new firms. They may not succeed the first time, but they will try again and learn from their failures. This is the reason why a cluster differs from other regions, has so many firms, and becomes so technologically advanced. However, any snapshot of the industry tends to ignore the fact that cluster status is achieved through continuous learning by trial and error.

Clustering of innovation

As entrepreneurship is clustered, innovative activities are also unevenly distributed over the landscape. Firms in clusters spend a lot of money on R&D. They make technological progress through both innovative research and imitation. In the long run, almost all firms in clusters have mastered very advanced technology, which leaves any firm outside clusters little chance of survival.

4.3 Location of clusters

The way we start our simulation implies that industrial clusters can emerge in any area. However, it is particularly interesting to know what factors may determine the location of clusters. To study that, we try simulations with different initial conditions.

Technological advantage

Our simulation shows that a region that is quicker at finding, learning, and imitating better technologies (the distribution of technology is further to the right, and/or lower values of f' and g') is more likely to develop into an industrial cluster. In reality, different regions do have different capacities in terms of research and innovation. For example, California and Massachusetts together house 14.3 percent of the U.S. population, yet 43.3 percent of the National Academy of Science members and 34.6 percent of the National Academy of Engineering members are based in these two states. Not surprisingly, California and Massachusetts lead the U.S. high-tech economy. Universities, research institutes and labs have always been a major source of technological advancement. The recent development of the biotech industry further proved the importance of academic research for a regional high-tech economy. Almost all biotech firms either were founded by academic researchers or received advice from them. At the same time, universities continuously provide high-quality laborers to the high-tech sector. It is safe to say that high-quality research universities are a necessary condition for a vibrant high-tech center, if not a sufficient condition.

Knowledge spillovers

Innovation such as superior technology is always first acquired by a lucky few. Other firms have to keep up with the pace of innovation through imitation. We find that regions in which firms may easily "copy" advanced technologies (smaller g' and/or imitation allowed for more distant firms) tend to develop into an industrial cluster. In reality, a local culture that tolerates inter-firm knowledge and labor transfers allows firms to learn collectively, which is favorable for the development of a cluster (Saxenian, 1994). A legal infrastructure, such as the enforceability of "not to compete" covenants, also has big effects on technology transfers (Gilson, 1999). Thus the way we see knowledge spillovers is different from the way in which the existing literature sees them. In our model, spillovers do benefit firms within clusters, but do not attract firms into clusters, in contrast to what the new economic geography literature suggests.

Seed capital and seed entrepreneurs

In a variation of our model, we assume that in some regions entrepreneurs have difficulties raising capital. In such regions, when agent i decides to start a firm, he acquires capital $K_i = 2k$ instead of $5k$ as in other regions, where $k \in U(0,1)$. Those capital-scarce regions are less likely to develop into industrial clusters. The availability of capital is important for fostering entrepreneurs, which is well recognized. Many scholars even suggest that local governments set up public funds to provide "seed capital" to potential entrepreneurs when the objective is to develop a high-tech regional economy.

In another variation of the model, we start by putting four "seed entrepreneurs" in four different regions and see whether that brings substantial advantage to those regions. Our simulation shows that, with a very high probability, one or more of the four regions will grow up into industrial clusters. In a different way, this proves first mover's advantage and path dependence. On the other hand, it also shows the importance of entrepreneurial leadership to a regional economy.

It is widely believed that the history of Silicon Valley traces back to the garage where Hewlett and Packard started their business in Palo Alto. Another frequently heard story is the departure of the "Traitorous Eight" from Shockley's Semiconductor Lab to found Fairchild. Those successes have inspired generations of entrepreneurs in the Valley. Almost every other high-tech center's history began with legendary entrepreneurs, who served as local heroes and role models who motivated others to pursue success in the same way. Famous examples include Ken Olson in Boston, Bill Gates in Seattle, and Robert Dell in Austin. It seems that the key to replicating the Silicon Valley model is to incubate such a heroic entrepreneur. Providing seed capital is certainly an important part of that game, but it is not sufficient. Although we know heroes in most cases spontaneously emerge, some local policies may facilitate their emergence. For example, local government may provide training program for those scientists and engineers who consider starting their own businesses. Favorable policies such as tax credits also help pioneers.

Trying, and learning by failing

A high-tech industrial cluster, by definition, is characterized by many successful firms. Our dynamic model allows us to see the other side of the story: clusters emerge on failures. Most successes are achieved through constant learning by trial and error. In fact, a region that does not tolerate failures (failed entrepreneurs not allowed to start over again or do not learn from failures ($\rho = 0$)) has a slim chance of success. Therefore, a cluster will most probably appear in a region where entrepreneurship is encouraged and failed experiences are valued (Saxenian, 1994).

5 Concluding remarks

We have proposed a simplified Nelson-Winter model with an explicit space dimension to study the way in which high-tech clusters emerge on a landscape in which no firm exists originally. We use agent-based simulation to show the dynamics of the model.

Social scientists have long been interested in clustering behaviors, such as racial housing segregation, the concentration of poverty and unemployment in certain neighborhoods, the exceedingly high crime rates in certain areas, the extreme dropout rates in certain schools, etc. Mainly two types of explanations are raised for clustering behaviors. One contends that clusters result from a sorting process in which individuals alike choose to associate with one another; for example, the residential segregation phenomenon can be explained in this way. The other argues that peer effects cause individuals to conform to norms in a social group. Obviously, the two arguments are not mutually exclusive. In many cases, including all other examples mentioned above, the two arguments may work simultaneously. The existing literature on industrial clusters recognizes the sorting process in which firms choose to locate close to other firms in order to exploit the benefits from a cluster. However, it neglects the possibility that entrepreneurial spirit can spread among the people in a region through social effects. We believe this kind of social contagion story is close to the reality of high-tech industrial clusters that are characterized by concentrated entrepreneurship. Our model shows that one does not have to invoke the benefits of industrial clusters such as knowledge spillovers in order to explain the formation of clusters; the contagion of entrepreneurship through peer effects alone is able to account for the emergence of clusters.

Another point our model has highlighted is the importance of pioneering entrepreneurs for an emerging industrial cluster. Entrepreneurial life style is by definition creative and disruptive. It takes at least one charismatic, successful role model to demonstrate the profits of taking risks and the joy of "changing the world" through innovation. Such "seed entrepreneurs" that generate a swarm of followers locally can be identified in every major high-tech industrial cluster in the United States. We are aware that such pioneers are not picked beforehand; in fact, in most cases, it seems as though those leaders had appeared by chance. However, we can certainly increase the chance of seeing such leaders by creating a favorable environment for entrepreneurial activities. Providing "seed capital" is one measure that may work, especially when entrepreneurs face binding financial constraints. Yet that is not enough. Given that high-tech firms are often founded by scientists and engineers, who do not necessarily have the impulse or knowledge to start as entrepreneurs, policies that help convert those people into entrepreneurs are useful.

Despite the simplicity of our theoretical model, it is beyond our capability to analyze it mathematically. For this reason, we resort to an agent-based simulation to show the evolutionary dynamics of the model. We have built a prototype for studying the emergence of high-tech industrial clusters. The primary advantage of the simulation approach is that we are free to try many variations of the model. For example, some authors have recently shown that large incumbent firms are likely to spin off new businesses, which provide an alternative mechanism through which industrial clusters emerge and grow (e.g., Klepper, 2001; Klepper and Sleeper, 2002; Lazerson and Lorenzoni, 1999; Zhang, 2003). Although our model completely shuts off the spin-off channel, one can easily modify our simulation to incorporate such spin-off activities.[5] Our model can also be modified to allow firms to move into or out of clusters, to have more sophisticated agents, to introduce product innovation in addition to new technology, or to test the consequences of different social network structures. We leave these for future work.

References

Anderson ES, Jenson AK, Madsen L, Jorgensen M (1996) The Nelson and Winter models revisited: prototypes for computer-based reconstruction of Schumpeterian competition. DRUID Working Paper No. 96-2

Arthur M (1990) Silicon Valley locational clusters: when do increasing returns imply monopoly? Mathematical Social Sciences 19: 235–273

Axtell R (1999) The emergence of firms in a population of agents: local increasing returns, unstable Nash equilibria, and power law size distributions. Working Paper, Brookings Institution

Axtell R (2001) Zipf distribution of U.S. firm sizes. Science 293: 1818–1820

Christensen CM (1997) The innovator's dilemma. Harvard Business School Press, Boston, MA

Cooper A, Folta T (2000) Entrepreneurship and high-technology clusters. In: Sexton DL, Landstrom H (eds) The Blackwell handbook of entrepreneurship. Blackwell, Malden, MA

Epstein JM (1999) Agent-based computational models and generative social science. Complexity 4: 41–60

Epstein JM, Axtell R (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press, Washington, D.C.

Foster RN (1986) Innovation: the attacker's advantage. Summit Books, New York, NY

Gilson RJ (1999) The legal infrastructure of high technology industrial districts: Silicon Valley, Route 128, and covenants not to compete. New York University Law Review 74: 575–629

Grebel T, Pyka A, Hanusch H (2001) An evolutionary approach to the theory of entrepreneurship. Mimeo, Department of Economics, University of Augsburg

Ijiri Y, Simon HA (1977) Skew distributions and the size of business firms. North-Holland, New York, NY

Kim S (1995) Expansion of markets and the geographic distribution of economic activities: the trends in U.S. regional manufacturing structure, 1860–1987. Quarterly Journal of Economics 110: 881–908

---

[5] Indeed, early spin-offs in Silicon Valley, such as those from the Fairchild Semiconductor, inspired many employees at incumbent firms to follow suit and start their own businesses, which has been happening over many generations. This is an important thread of Silicon Valley's history (Saxenian, 1994).

Klepper S (2001) Employee startups in high-tech industries. Industrial and Corporate Change 10: 639–674

Klepper S, Sleeper S (2002) Entry by spinoffs. Papers on Econonmics and Evolution 2002-07, Max Planck Institute for Research into Economic Systems

Krugman P (1991) Geography and trade. MIT Press, Cambridge, MA

Lazerson MH, Lorenzoni G (1999) The firms that feed industrial districts: a return to the Italian source. Industrial and Corporate Change 8: 235–266

Marshall A (1920) Principles of economics, 8th edn., Macmillan, London

Nelson RR, Winter S (1982) An evolutionary theory of economic change. Harvard University Press, Cambridge, MA

Nelson RR (1995) Recent evolutionary theorizing about economic change. Journal of Economic Literature 33: 48–90

Porter ME (1998) Clusters and the new economies of competition. Harvard Business Review 76: 77–90

Sexenian A (1994) Regional advantage: culture and competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA

Schelling TC (1971) Dynamic models of segregation. Journal of Mathematical Sociology 1: 143–186

Schumpeter JA (1934) The theory of economic development. Harvard University Press, Cambridge, MA

Sutton J (1997) Gibrat's legacy. Journal of Economic Literature 35: 40–59

Tesfatsion L (2001) Introduction to the special issue on agent-based computational economics. Journal of Economic Dynamics and Control 25: 281–293

Zhang J (2003) High-tech start-ups and industrial dynamics in Silicon Valley. Public Policy Institute of California, San Francisco, CA

The theory of the firm and the markets for strategic acquisitions*

Gunnar Eliasson[1] **and Åsa Eliasson**[2]

[1] Royal Institute of Technology (KTH), 100, S-100 44 Stockholm
(e-mail: gunnar.elias@telia.com)
[2] VitiGen AG, Siebeldingen, Germany

Abstract. Five problems are addressed: (1) the role of competent actors in the venture capital and exit markets supporting the industrialization of winning technologies in small innovative firms, (2) the competence of the large firm to integrate large-scale operational efficiency with small-scale innovative capability through distributed development work and integrated production and (3) the importance of viable markets for strategic acquisitions, both in making this possible and in allowing a flexible choice for the small firm between growing aggressively on its own through own acquisitions, or being acquired strategically itself. We (4) find that the less developed markets in continental Europe may be a disadvantage compared to the US in ushering in a future New Economy. We finally (5) discuss what becomes of the Coasian theory of the firm when production is constantly outsourced in, or insourced from the market as the relative efficiency of coordination through management and over the market changes. One logical consequence is that the costs of business mistakes will have to be included in transaction costs.

Keywords: competence bloc – experimentally organized economy – heterogeneity – Marshallian industrial district receiver competence – strategic acquisitions

JEL Classification: G24, G34, L16, L23, O33

---

* An earlier version of this paper was presented at the 9th Congress of the International Joseph A. Schumpeter Society (ISS) March 28-30, 2002, Gainesville, Florida, USA. This is a companion paper to Eliasson and Wihlborg (2003) which is also in this volume. Both papers are part of the Nödfor Project on Schumpeterian Creative Destruction, notably the exit and bankruptcy process, based at the Ratio Institute, Stockholm.

1 The problems

The last couple of decades have seen an increase in the fragmentation over markets of firms as centrally coordinated hierarchies (Eliasson, 1986b, 1996b, 2001b,a). Products have been modularized and the production of components or entire systems outsourced in the market, only to be insourced again at some later stage. Coase (1937) outlined the principles behind such organizational change when explaining the rationale for the existence of the firm as a hierarchy in those instances in which management has a transaction cost advantage over production coordination. Holmstroem and Tirole (1989) extend that notion of a firm to a "contract between a multitude of parties" imposed over the market to "minimize transaction costs between specialized factors of production". Since this is an authoritative statement in the Handbook of Industrial Organization on the theory of the firm, we begin there. Empirically integrated development work and production distributed over markets are becoming an increasingly important productivity factor in the emerging industrial technology (Eliasson, 1996b; Jovanovic and Rosseau, 2002; Lerner and Merges, 1997) moved, notably, by modern computer and communications technology. Firms are reorganizing through acquisitions and divestments to gain competitive advantage. Both the notion of a firm and the structure of markets for control, hence, are experiencing radical change. To integrate the organizational problem of firms that have to rely on the external market for innovations with the theory of the firm, the notion of the reorganization of production structures over markets has to be endogenized. This means not only accepting complexity, ignorance and business failure, but also making *business mistakes part of transaction costs*. This disrupts the exogenous equilibrium properties of the neoclassical model. Transaction costs can no longer be minimized independently of the production organization. Further, the story to follow argues that it is empirically unacceptable to structure the theory of the firm such that this is possible.

We use competence bloc theory (Eliasson and Eliasson, 1996, 2002a) to (1) model the firm as an endogenously changing organization distributed over markets, the extended firm (Eliasson, 1996b), and (2) to demonstrate that an endogenous hierarchy makes it possible simultaneously to achieve both the narrow focus needed for operations efficiency and the broad exposure to a maximum of varied competence needed to be dynamically[1] efficient in the Austrian-Schumpeterian environment of the Experimentally Organized Economy (EOE, Eliasson, 1987, 1992). More precisely, we address the existence of a market for strategic acquisitions as a source of systemic productivity gains ("economies of scale") and a mover of industrial dynamics. The problem addressed is elucidated by the fact that the large firm, normally oriented towards large-scale operational efficiency (Eliasson, 1976, 1984, 1996a, 2001a; Acs and Audretsch, 1988), has problems with its innovative capabilities. Small firms, on the other hand, are less formally organized and more flexible and, therefore, thought to be more capable of innovative achievement. The small firms, oriented towards innovative performance and pursuing radically new innovations,

---

[1] The reader should observe already here that this notion of dynamics takes us far beyond the neoclassical notion of dynamics, that often uses the attribute as soon as a time variable figures in the equations

by contrast, suffer from the ignorance of the financial community when it comes to understanding what the firm is doing and the high financial risks for the innovators/ entrepreneurs associated with taking a winning innovation on to industrial scale production. The *industrial competence of the actors* in the financial markets intermediating trade in intangible knowledge assets, therefore, is a key concern. Here we draw directly on the property rights analysis explored in the companion paper (Eliasson and Wihlborg, 2003, see also this volume).

Venture capitalists, as we define them (Eliasson and Eliasson, 1996; Eliasson, 2003), are characterized by their competence and capacity to understand radically new business ideas and provide reasonably priced financing. But the little firm doing it on its own faces the additional problem of being too slow in reaching industrial scale, and, therefore, risks being overtaken or imitated by a larger firm with ample financial resources. The second part of the same problem is the increasing inability of large business firms to do it all and to efficiently incorporate or internalize all the needed competences within one hierarchy. The large firms normally have the financial capacity to buy and possibly also the competence to discover and to access new technology at fairly advanced stages of development and close to their core business, but have problems with their indigenous capacity to create the same technology. In addition, the large firm normally has great difficulties introducing radically new technology in its operations-oriented organization because of the lack of *receiver competence* and a consequent skeptical attitude among the staff to the introduction of novel and organizationally disrupting ideas (Eliasson, 1976, 1984, 1990a; Eliasson and Granstrand, 1985). In the Holmstroem (1993) model, this is explained in terms of a bureaucratization that arises because of the higher transaction costs associated with "mixing hard to measure activities (innovations) with easy to measure activities (routine)". For the large firm, there is, however, the additional problem of competence supply, notably of innovative, technological variety that has become critical for survival in the new economy. Since the single firm normally lacks the capacity of internally supplying all the needed variety of innovative services, the solution has increasingly become to acquire complementary services externally. For this to be possible, the advanced manufacturing firm has to access the broad and deep markets of subcontractors. The more advanced and the more dependent on R&D, the more important it is that technological supplies can be outsourced. Outsourcing of technological development is a difficult part of advanced production that has become necessary and has been seriously learned only in the last decade or so as new computing and communication technologies have made the integration of globally distributed production feasible (Eliasson, 2001a, 2002b). We look especially at the existence and the role of viable markets for strategic acquisitions[2], and how the incentives needed to support such markets depend on competition for their innovative services from a varied set of large corporate customers (Eliasson, 1986b).

This paper, hence, focuses on three problems:

---

[2] The implicit assumption of Arrow (1962) that technological services can be outsourced to technical universities and government-run laboratories is based on the assumptions of the static general equilibrium model with zero transaction costs, and is simply a misconception in this context, even though it has been extensively used in the theoretical innovation literature. See further below.

1. The venture capital competence needed to discover and to commercialize radically new technology in, and support expansion of, the small firms,
2. the competence of the large firm to integrate large-scale operational efficiency with small-scale innovative capacity through distributed development work and production integrated over the market, and
3. the conditions for the existence of viable markets for strategic acquisitions that offer profitable choices for the small innovative firms to pursue their own growth plans and/or (second best) to aim at being strategically acquired. This choice will be seen as a determining incentive for a rich supply of innovative firms.

The critical role of appropriately designed contractual rights to knowledge to establish *the tradability in intellectual capital* needed to support knowledge creation and allocation over markets for strategic acquisitions has already been addressed in the companion paper (Eliasson and Wihlborg, 2003).

2 Background theory

The classical representation of a firm is that of a monolithic hierarchy controlled from the top. Before Coase (1937), and even decades after the publication of his article, most economists bothered neither about the firm nor about the organization of the economy. They were concerned with analyzing industries in which live firms disappeared in aggregates.

Marshall (1890, 1919) wanted to change this situation and is credited by Schumpeter (1954) with having been the first to bring business economics into economic theory. His "representative firm" was an attempt to deal with the aggregation problem, although his "industrial district" analysis is more innovative and to the point in this context. This analysis featured a network of subcontractors – an organization of production within which systemic productivity gains could be captured – that allowed him to make increasing returns compatible with the then-dominant Walrasian model. Marshall's industrial district included already in 1890 a micro-based formulation of what later (in the 1980s) came to be called "new (macro) growth theory".

Coase (1937) recognized that the outer limits of the firm were determined by the relative costs (transaction costs) to coordinate the business through a hierarchy and through the market. The "hierarchy" or firm became endogenized and changed in response to market forces, drawing significant transaction costs. Arrow (1965) emphasized the role of the organization in bearing risks where the market failed. But *organization is a more general instrument to cope with competitive change* . This makes it natural to extend the Coasian (1937) model to handle also the dynamics of the new, loosely structured extended firms (Eliasson, 1996a,b, 1998b) that constantly reconfigure their internal structure and trade in parts over the M&A market. In neoclassical R&D based innovation functions, the roles of the innovator, the entrepreneur and the venture capitalist are collapsed into one. Technology becomes a linear driver of growth. Also Joseph Schumpeter (1942) superior scientifically-based firm that would eventually dominate its market is based on a linear technology growth relationship. The organization of innovative activity and of *"The Markets* *for Innovation, Ownership and Control"* (Day et al., 1993) may, however, *matter not only for innovative output* but also for the link between innovative output and economic growth. Once that possibility has been recognized, the intersection between hierarchies and markets (the organization of the firm) becomes endogenized and tradability in technology assets becomes a determining factor. Control rights to assets is the signum of a firm as a hierarchy and the optimal assignment of assets is one way to understand the boundaries of the firm (Hart and Moore, 1990). But a firm is more than a contractual arrangement to allocate ownership, control and responsibilities of the parties involved (Holmstroem and Tirole, 1989). The financial structure is not independent of the underlying production organization; for instance, the choice between outsourcing and internalizing through vertical integration also depends on the control of production desired (Lewis and Sappington, 1991)[3].

The competence embodied in the hierarchy can be improved by reassigning control rights to the actors with maximum competence to run the business (Aghion and Tirole, 1994), who might in turn change the production organization. A different authority (hierarchy) can thus be superimposed on, and exceed the limits of, the Coasian firm. The stronger the property rights, the more tradable technology assets and the stronger the influence on production organization (Eliasson and Wihlborg, 2003).This reassignment has a precise meaning in the knowledge-based information economy (Eliasson, 1990b), featuring large information and communications costs and a virtually unlimited set of business opportunities. Firms, *defined as competent teams* (Eliasson, 1990a), are normally grossly ignorant about circumstances relevant to their business and long-run survival, not least about what competitors are up to. They, therefore, set up business experiments to the best of their knowledge, which sometimes succeed, but often fail. Business mistakes, therefore, become a normal cost for economic development and part of the transaction costs incurred when doing business. Hence the term the Experimentally Organized Economy (EOE). The central firm (management) problem in the EOE now becomes to minimize the economic costs of two types of errors, namely (Table 2A) to keep business mistakes on the books for too long and to lose the winners. We identify the scope of the organization called a firm within which management can do this. The key problem is to avoid losing the winners, which for a competent management is perhaps the largest item in transaction costs. Part of the competence involved in achieving this is the art of delimiting the scope of the firm (the span of management, Simon, 1957)[4]. Minimizing transaction costs cannot be done independently of the exercising of this art. As we will see, this is no trivial problem in the theory of the firm. *Competence bloc theory* also deals with this problem of dynamic efficiency in the EOE. In competence bloc theory, the creation and selection of projects can be distributed over competent actors in the market, or be internalized within the firm.

---

[3] Here Desai et al. (2002) observe that US multinationals have, over the last 20 years or so, gone from loosely structured alliances to 100 percent ownership control. They explain that by a desire to exercise more control in coordinating production and in technology transfer, a development also induced by a liberalization of ownership restrictions in host countries and by trouble with new US tax reforms when it came to the free use of rational internal transfer prices across borders.

[4] The loss of winners is no problem in the WAD model, since it cannot occur there by assumption (Eliasson, 1992)

von Hayek (1937) formulated this as a parallel to Adam Smith's dictum of decentralized production, when he discussed the "division of knowledge". Knowledge dominates all other physical forms of capital in determining the productivities of other factors of production. But knowledge capital is not well defined and cannot be understood and managed analytically under the assumption of full information economics . Knowledge is largely tacit and incommunicable, and can only be allocated by knowledge (cf. Demsetz 1969 and Pelikan 1986,1988 on economic selforganization). Markets in tacit knowledge are often characterized by infinite regresses and the non-existence of an external equilibrium. As a consequence, dynamic efficiency in the sense of minimizing the economic consequences of the two types of errors in Table 2A can only be achieved by exposing each project to a maximum competent evaluation. That, in turn, can only be achieved in low transaction costs markets with well developed property (control) rights that support trade in intellectual assets (see Eliasson and Wihlborg, 2003). Attempts to centralize the decision will make the decision/selection more narrow and raise the risk of losing winners. We, therefore, introduce competence bloc theory as an organizing device for the distributed (over the market) creation, identification and selection of projects in the experimentally organized economy (EOE). Competence bloc theory will allow us to identify the markets for competence that are critical for the project selection that is the key to the efficient solution of all three empirical problems of this paper. For this to be possible, however, tradability in the competence/control rights or intellectual assets has to be established in the non-equilibrium setting of the EOE.

3 The commercialization of winners in the experimentally organized economy – the first problem

The theory of the EOE features growth through experimental project creation and selection. Competence bloc theory explains the nature of that selection, which, in turn, allows us to understand the roles of a venture capital market with industrially competent actors and the market for strategic acquisitions in reconfiguring firms into new business combines. These two markets exist in, and integrate, the activities of actors in the competence bloc.

3.1 The experimentally organized economy

The notion of a knowledge-based information economy (Eliasson, 1990b) is used to establish the basic assumption of a business opportunities set of such complexity that practically all actors become grossly ignorant even of (for them) very relevant circumstances. This means that business decisions will have to be seen as more or less well prepared (business) experiments that often fail. In this experimentally organized economy (EOE), growth occurs through competitive project creation and selection.

The EOE offers an alternative to the Walras-Arrow-Debreu (WAD) model (Eliasson, 1992), the main difference being the assumed dimensions of the (business) opportunities space, or state space, and the appearance of significant information and communications (transaction) costs in the form of business mistakes. The latter removes the property of an exogeneous equilibrium of the WAD model. The WAD model assumes the state space to be extremely small and sufficiently transparent for all options to be identified. In the EOE the state space is extremely large and non-transparent. The theory of the EOE thus embodies the experimental nature of dynamic markets and allows ignorance and business mistakes natural roles to play. It has its roots in the Austrian economics of Menger (1872) and in Schumpeter (1911), before Schumpeter turned "linear" in 1942. The EOE features economic growth through experimental creation and selection of innovative projects. A policymaker in the EOE would constantly face the problem of efficient exit, i.e. of forcing badly managed incumbents or new entrants to exit without exiting winners. This is the dynamic efficiency problem of the theory, demanding great and varied competence on the part of actors participating in the selection process, *including* the policy maker if "it" feels a need to get involved (Eliasson, 2000). Salter curve analysis (Salter, 1960) then allows the *Schumpeterian creative destruction process* of Table 1 to be derived, and relates it to macroeconomic growth (see Eliasson, 1996a, Section II.7).

Table 1. The four mechanisms of Schumpeterian creative destruction and economic growth

1	Innovative entry
	enforces (through competition)
2	Reorganization
3	Rationalization
	or
4	Exit (shut down)

Source: "Företagens, institutionernas och marknadernas roll i Sverige", Appendix 6 in A. Lindbeck (ed.), Nya villkor för ekonomi och politik (SOU 1993:16) and Eliasson (1996a, p.45).

The performance characteristics of an agent can be ranked in each market. The Salter curves of Fig. 1 exhibit such rankings of rates of return or temporary knowledge rents[5] for two years in Swedish industry. Superior firms to the left can outbid lower down firms in hiring people, buying components, lowering prices or acquiring firms. But the challenged firms know this and have to act to improve their situation, thus challenging the (temporarily) superior firm. All incumbents are challenged by new entrants, and challenged firms that cannot cope with the situation are forced down the curve, eventually to exit at the low right hand corner. Competition is endogenous, forcing organizational innovative behavior as represented

---

[5] The rates of return shown in Fig. 1 minus an appropriate interest rate can be said to measure temporary knowledge rents and the incidence of random factors or "luck". Expected such returns to capital over the interest rate drive firm investments in the MOSES micro-to-macro model to be referred to in the next footnote. The rents so defined have been estimated for the real firms in the so called planning survey of the Federation of Swedish Industries that make up the population of MOSES firms since the mid 1970s (MOSES Data Base). Those rents exhibit considerable volatility over the firm population and time (see Albrecht et al., 1992).

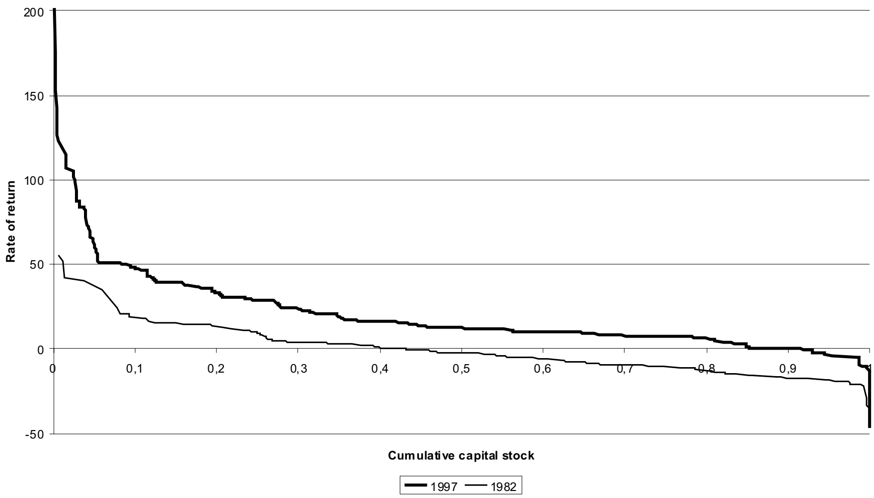


Fig. 1. Rates of return (per cent), 1982 and 1997
Note: Swedish manufacturing industry
Source: MOSES database.

by the four categories of Table 1. Only if society is "efficiently" organized and equipped with the right institutions and incentives will this dynamically competitive process of experimental selection lead to macroeconomic growth through the outward shifting of the Salter curves. But endogenous competition could also lead to contraction and exit[6]. The problem for the policy maker is to organize institutions such that winners are helped to move on and losers are forced to release resources, notably competent people, for the growing firms. By increasing factor supply factor prices are held down. This dynamic turns exiting losers into growth contributors. *Dynamic efficiency* in the EOE can thus be characterized by the capacity of the economic system to "minimize" the economic consequences of two types of errors in the Schumpeterian creative destruction process, shown in Table 1, not to keep losers for too long and (most importantly) not to lose the winners (see Table 2). Competence bloc theory organizes tacit knowledge distributed over markets and hierarchies to achieve that outcome.

Table 2. The dominant selection problem

Error Type I: Losers kept too long
Error Type II: Winners rejected

Source: Eliasson and Eliasson (1996).

---

[6] This Schumpeterian creative destruction process endogenizes economic growth in the Swedish micro-to-macro model MOSES (Ballot and Taymaz, 1998; Eliasson, 1991, 1996a, 2000). Johansson (2001) has econometrically tested for the characteristics of firms and their environment that make firms expand rather than contract under competitive pressure, i.e. for the circumstances that make the Schumpeterian creative destruction process of Table 1 lead to industry growth rather than contraction. Also see Eliasson (2000) and Eliasson et al. (2002, 2004)

Table 3. Actors in the competence bloc

1	Competent and active *customers*
2	*Innovators* who integrate technologies in new ways
3	*Entrepreneurs* who identify profitable innovations
4	*Competent venture capitalists* who recognize and finance the entrepreneurs
5	*Exit markets* that facilitate ownership change
6	*Industrialists* who take successful innovations to industrial scale production

Source: Eliasson and Eliasson (1996). The Biotechnological Competence Bloc, Revue d'Economie Industrielle, 78 – 4^0, Trimestre.

3.2 Competence bloc theory

The competence bloc (Eliasson and Eliasson, 1996, 2002a) lists the minimum number of actors with competence needed to minimize the economic consequences of the two kinds of business errors. It is a theoretical design that allows an organization of decentralized tacit knowledge without specifying the content of knowledge, except by *function and carrier*. The solution is to organize diverse and distributed competences in the economy such that each project is exposed to a maximum of competent and varied evaluation. The competent *customer* (item 1; Table 3) defines the maximum degree of "sophistication" of the product for which the most advanced customers are willing to pay. *Without competent customers there will be no markets for sophisticated products*. The competence bloc incorporates the Burenstam-Linder (1961) idea that advanced customers constitute a comparative advantage for the rich industrial countries. During the development of advanced products, such as aircraft, technologically knowledgeable customers often contribute directly to product technology (Eliasson, 1996b). The normal situation, however, is that the customer chooses between different product offers. Heterogeneity in the supply of innovative new products and the supply of competent customers with varying tastes, therefore, set limits to technological advance. The *innovator* (item 2) is defined as the actor who combines new and old technologies into new composite technologies, to be selected by economic (profitability) criteria by the *entrepreneurs* (item 3)[7]. The entrepreneur, in turn, normally needs external financing to move expected winners on, but that financing has to be associated with a competence on the part of the financial contributor to understand the entrepreneurial selection. Otherwise (Eliasson and Eliasson, 1996), the conditions will be so tough as to leave little or nothing for the innovator and the entrepreneur. The *competent venture capitalist* (item 4) selects the winning entrepreneurs. The venture capitalist, however, needs large and deep *exit markets* to unload his stake with a large profit at the, for him, appropriate time. If a real winner is moving through the competence bloc, the next step is for a competent industrialist to take over and move the project on to industrial scale production and distribution. The industrialist now acts as a customer in the exit market, or rather in the intersection of the venture capital and exit mar-

---

[7] For theoretical reasons we want to give the innovator a technological definition. The innovation is selected and transformed into a business proposition by the entrepreneur. This is von Mises (1949) rather than Schumpeter , who does not distinguish clearly between the two concepts.

kets, that we call the market for *strategic acquisitions*. By analogy with the earlier customer analysis, without a broad range of sophisticated industrialists/customers, an active market for strategic acquisitions would not exist, and there would be no real incentives for sophisticated entrepreneurial firms to enter the *market*. Apparently, all later stages (in the competence bloc) are important for incentives to be effective at the earlier stages. If the competence bloc is not *vertically complete*, the risks are large for the earlier stage innovators and entrepreneurs[8]. But vertical completeness is not sufficient. One actor of each does not guarantee a varied and competent project evaluation. Many of each with very different competences are needed. Only when *vertical completeness* and *horizontal variety* are in place can *critical mass* be reached and *potential winners* confidently pursue their search. *Increasing returns to continued search* then prevail and the risks that winners may get lost are minimized. The competence bloc then functions as an *attractor* for advanced firms that both *benefit* from localizing there, and *contribute* to the further development of the competence bloc. In that sense, the advanced firm of the competence bloc also functions as a technological *spillover source*, or as a technical university (Eliasson, 1996b). The competence bloc transfers valuable and more or less tacit knowledge ("technology") between actors with competence capable of adding value. This transformation takes place in the internal markets for innovation within firms or through trade over external markets between firms. For this tradability in technology assets to be achieved, the problem of establishing property rights to intangible knowledge assets has to be solved (Eliasson and Wihlborg, 2003). Lamoreaux and Sokoloff (2002) in fact argue that institutional support, notably the patent system that "created secure and tradable property rights in invention" was instrumental in commercializing technology and initiating the rapid productivity advance in the US economy 1870-1920. Some large firms internalize almost entire competence blocs to solve the property rights problem, as IBM did in its heyday in the 1970s. It was even an advanced customer of its own products, such as advanced microchips. Most large firms internalize significant elements of the competence bloc, notably the venture capital function. An internalization of the link between the innovator and the industrialist in the competence bloc as shown in Table 3, however, suggests a narrowing of the competence a project is exposed to in the internal firm evaluation (in the hierarchy). Distributing the same evaluation over a competence bloc with many competent actors means a broader and more varied project evaluation, but produces larger transaction costs and introduces an extra element of uncertainty about the distribution of rents. That innovative variety may disturb the operational efficiency of a large hierarchy is illustrated by a quote from a well known Swedish business leader: "We would get very irritated at an entrepreneur at the postal office that delays the morning mail delivery, even though this person has interesting ideas about how to improve postal service...We need some creativity —

---

[8] The results of Darby and Zucker (2002) illustrate the difficulties of effective selection and the nature of industrial knowledge in the financial community. They find that the quality of biotech firms' science base measured by the number of articles published by academic stars associated with the firm signals economic performance potential of the firm "making it easier to find capital and to obtain it in large amounts"

but not much"[9]. The standard way of attempting to solve that problem is to keep the innovative and operations responsibilities organizationally separate within the firm (Eliasson, 1976).

Thus competence bloc theory is sufficient to demonstrate our first proposition of the critical role played by competent venture capitalists and exit markets to identify and move winners in new technologies on to industrial scale production. Selection has to be decentralized over markets to be truly dynamically efficient[10]. Internalizing the competence bloc into one hierarchy reduces variety and hence the innovative capabilities of the economic system.

3.3 Redefining transaction costs in the EOE

The value of an asset depends on the ability of the owner to control its use (management), to capture its rents (access) and to trade in the asset. Hence, the value can be calculated as the present value of the rent flow net of transaction costs, and discounted by a market interest rate plus an appropriately scaled risk premium (Eliasson, 1998a). Restrictions on the control rights and on tradability are factored into the risk premium. A particularly important matter for the valuation of assets in the EOE is the definition of transaction costs. The mainstream model does not recognize business mistakes. The theory of the EOE does. There, the costs of business mistakes, in terms of Table 2, figure as a cost for learning and economic development contributing to the creation and commercialization of "winners". If incurred within one hierarchy, it belongs to its cost structure, with the important distinction that lost winners are not charged to any cost account. The business mistakes may have been made in other firms, making it possible for a particular firm to learn from these mistakes. The costs are then carried by others or society at large. Costs associated with the commitment of business mistakes thus have to be included in a correct definition of transaction costs in the EOE. This was first recognized by Dahlman (1979). Only then will a market allocation solution to the allocation of resources get a fair theoretical comparison with a centralized hierarchical solution to the same problem. Only then will it also be possible to understand theoretically that a distributed (over the market) reallocation of intellectual assets ("competence capital") often is dynamically more efficient than a narrow evaluation within a hierarchy. Even though more costly in terms of direct transaction costs, the more varied evaluation reduces the incidence of business mistakes and hence total transaction costs, appropriately defined for the EOE. Dynamic, or Schumpeterian efficiency is increased.

---

[9] PG Gyllenhammar, then CEO of Volvo, at the 90 years celebration in 1986 of the Swedish Engineering Industry Association. For the full quote, see Eliasson (2002b, p.97)

[10] Efficiency or opportunity costs in the EOE can, however, not (as in WAD theory) be measured by reference to a well defined benchmark, i.e. static equilibrium when all actors operate on the production frontiers. There will always be unknown better projects that cannot be "objectively identified". They are only known to exist (Eliasson, 1998b, 2001b). Dynamic or Schumpeterian efficiency (Eliasson, 1985, p.329 and Eliasson, 1991, p.165) is measured against a minimum of lost winners which is indeterminate in the EOE.

The gradual emergence of informed and dynamically efficient markets for corporate control increasingly offers distributed (over the market) solutions to the problem of internalizing the innovative and operational functions of production. Distributing sophisticated production based on tacit competence capital over the market, however, also requires that intellectual capital be competently and fairly valued in markets. For this to occur, property rights have to be competently assigned such that trade can be established in these values at low transaction costs. This particular problem is dealt with in a companion paper (Eliasson and Wihlborg, 2003). Again, competence bloc theory helps us understand and explain how.

4 Integrating innovative and operational efficiency over the market – the second problem

The innovator, the entrepreneur, the venture capitalist (financier) and the efficient large scale organizer of production are rarely embodied in the same person or hierarchy. New, winning ideas are often lost in an efficient manufacturing environment. Hence, the creation, diffusion and introduction of winning innovations in production, the incentives to innovate and to industrialize and the sharing of rents from winners are the critical problems of economic dynamics and growth. The internal economies of large firms are normally conservative and inclined to reject radically new (alien) project proposals, losers and winners alike. The small firm, with the radically new idea, on the other hand, does not have the financial resources of the large firm.

Large, successful industrial economies are often dominated by large firms in mature markets, excelling in efficient volume production. To strike the right balance between efficient volume production and the capacity to innovate, therefore, is as critical for the wealthy industrial economy as it is for the large firm. In the long run, a conservative attitude in the dominant part of the industrial establishment of a nation may be detrimental to the supply and absorption (*receiver competence*, Eliasson, 1986a, pp.47,57; Cohen and Levinthal, 1990; Eliasson, 1990a) of radically new technology. Hence, the organization of markets for innovation is a core economic design problem in an industrial economy.

4.1 On the existence of a market for strategic acquisitions

Outsourcing innovation over an entire competence bloc is one organizational solution to the problem of project selection. This requires the existence of a market for innovations, which is a matter of the existence of venture capital and exit markets (items 4 and 5 in Table 3). The exit market, then, becomes a market for strategic acquisitions, offering a supply of radically new innovations embodied in “small new firms”, the innovations having been moved beyond the entrepreneurial stage by venture capitalists, who now supply the exit market with strategic investment opportunities. The existence of such a market for strategic acquisitions cannot be taken for granted. First, rather than being created, selected and carried on to industrial scale production within one hierarchy, the same functions are now distributed over subcontractors. Functionality, hence (and first), requires the existence of a complete competence bloc that has reached the critical mass and variety to identify and move winners up to and into the exit market, where industrial buyers wait. *Second*, since these activities are now organized over the market, involving trade in intangible knowledge assets, the art of defining, assigning and valuing to make the requisite property rights tradable becomes critical (see accompanying paper Eliasson and Wihlborg, 2003). *Third*, the low internal firm cost of a narrow and often incompetent valuation and selection procedure and the loss of winners have to be weighed against the higher direct transaction costs over the market to achieve a more informed valuation and a better allocation of the total knowledge capital. With the loss of winners included as a transaction cost, the distributed market solution now may become the low-cost alternative.

The *incentives* that move the market evaluation process are not exogenous, but rest on the competence of industrial buyers to understand the projects. Hence, variety among industrial buyers (Eliasson, 1986b) raises competition for the winners and moves their price above the prices offered by incompetent industrial buyers. The high price is critical. Incompetent industrial buyers that have acquired a winner in a distressed situation cheaply can incur large losses by making business errors of type II without privately losing much money (see case below). For the economy at large, the loss of a better and/or a winning production organization may, however, be great.

On the one hand, we have the competent industrial buyers who can pay the right (higher) price for industrially valuable innovators because they know how to create value by integrating them into their business. But innovations are not supplied to order by entrepreneurs and venture capitalists in such markets. On the other hand, we have the industrially incompetent buyers who shop for cheap acquisitions that may, or may not, turn a profit but that often entail a loss of winners because of the industrial incompetence of the buyers. Hence, *many industrial buyers representing a varied competence are needed to support a viable market for strategic acquisitions* (Eliasson, 1986b). Since biotechnology and pharmaceuticals are industries where this market is critical, the current merger activity and concentration among the large pharmaceutical companies are not a good signal in this context. It reduces competition for innovative firms in the market for strategic acquisitions.

4.2 Large systemic productivity gains

The market for strategic acquisitions offers a way for large companies to integrate both innovative activity and economies of scale volume production within one distributed hierarchy, which we will call an extended Coasian firm or a Marshallian industrial district. The combination, if it can be organized, establishes *a positive sum game with systemic productivity gains*. These potential systemic effects also offer incentives for the competent and innovative industrial organizer. Competence bloc theory explains the principles for this in the EOE. Undeveloped markets for strategic acquisitions have been shown to be a handicap for the early stage actors, the innovators, who often have to part with a winner to a financially strong, later stage actor. But we can also conclude theoretically that if the big companies collude and/or squeeze the prices of strategic innovation offerings, the policy runs against their own long-term interests. On the other hand, a market that induces many large companies to compete for winning projects is the preferred situation for the small innovative company. In underdeveloped markets for strategic acquisitions, on the other hand winners easily go undiscovered and large incompetent companies can pick up a winner cheaply and scrap it at a small loss if it fails. In fact, innovators in a badly developed competence bloc with no market for strategic acquisitions have to be irrationally overoptimistic to go on innovating at large private risk.

4.3 The concept of dynamic or Schumpeterian efficiency

A reference or a bench mark to define and measure efficiency is needed. We need to know the *opportunity cost* of not doing something in a different and perhaps better way. Such a bench mark used frequently among firms is to compare the situation with a best competitor or the best plant in the own firm[11]. Ideally the reference should be the best possible or maximum performance. This is in principle easy in mainstream equilibrium modelling where the (full information, perfect) equilibrium is associated with the notion of maximum achievable performance: if it can be shown to exist, you have a benchmark for efficiency measurement. Standard economic analysis "attempts" to organize its assumptions such that the economic model can be solved for such equilibria or exogenous benchmarks. The problems with this model, and with such analysis, is how it relates to the underlying reality. We have no such principal problem with the theory of the EOE. On the other hand, the EOE has no stable exogenous equilibrium to be used as an efficiency reference, but we regard that as an advantage. Defining efficiency, however, becomes a problem, because the optimum reference or the opportunity cost has to incorporate the hypothetical economic performance of lost winners, had they not been lost. This reference is indeterminate since it depends on all factors ruling the growth process in the EOE, and the basic idea of the theory of the EOE is that far better solutions than the existing ones are possible for those economically motivated to search for them and capable of identifying them. Hence, the model of the EOE cannot be solved for an external equilibrium[12]. The indeterminacy of a reference tilts the policy focus away from the analytical ambition of the WAD model. Instead of using information to determine the best solutions the policy ambition is now to design institutions and instead boost incentives to search for the better solutions and to help build the institutional and human capital infrastructure embodied in the competence bloc to "maximize" the exposure of each project to a competent evaluation. The paradoxical coincidence is that the presence of large information and communications costs in production

---

[11] Such bench marking in large firms with multiple production facilities of the same kind was common already in the early 1970s (see Eliasson, 1976, pp.180, cases 13).

[12] We can simulate possible better trajectories involving fewer losses of winners, for comparison, using the Swedish micro-to-macro model MOSES, which approximates the EOE (Ballot and Taymaz, 1998; Eliasson, 1991). MOSES is an evolutionary model which develops differently, depending on initial circumstances and the discrete choices made by actors in the model during the simulation, and it never settles on an exogenous equilibrium path Eliasson and Taymaz (2000).

when incorporated in the theory of the EOE is what causes this redirection of theoretical attention away from information towards institutions and incentives.

4.4 Failure in the market for strategic acquisitions

Market failure in the form of lost winners easily occurs in the EOE, and always occurs to some extent if the competence bloc is not complete and/or not sufficiently varied horizontally. We have to watch our tongue, however. What looks like market failure often originates in policy or political failure. For instance, if the tax system makes it impossible for industrially experienced and competent rich individuals to develop into venture capitalists and/or if the wrong people become rich and enter venture capital financing, the critical venture capital competence input in the competence bloc will be lost – a political failure. Similarly, if policy creates a long depression of values in the stock market, making it easy for large and not very competent buyers to shop for bargains, often losing a winner here and there, we have again an instance of policy failure, not of market failure. The most common origin of business failure, however, is lack of competence to perceive the right combination of technologies through strategic acquisitions and divestments. For the acquiring company, the potential value may be much larger than the sum of values the acquisition objects can fetch individually in the market- if it has the competence to do it, not only about right, but exactly right.

Three cases will illustrate the latter aspect in particular.

Case I: Uppsala based firm in molecular diagnostics (Eurona Medicals).

This firm was spun off from Pharmacia in 1994, when Pharmacia decided not to pursue its molecular diagnostics venture, a then pioneering field aimed at making individual genetic diagnostics and personalized medicine possible. This market is now considered to be the promising area for new innovative health care (Eliasson and Eliasson, 2002b). Eurona had two mutually supporting specialties, *substance testing* (lab processes, data base analysis and data access) and *genetic diagnosis*, the second specialty being the by far more innovative and promising venture. Here Eurona was a pioneer, perhaps too early.

Today the average “hit rate” for a substance is some 20 percent, meaning that most patients will score no hits for a while, only suffering from cumulative side effects. Some unlucky patients score no hits and only suffer from the side effects. The business opportunity lies in the fact that the genetic variation between patients makes them react differently (for the same disease) when prescribed the same substance. The potential lies in genetically diagnosing each patient, and tailoring the substance to the patient. The potential of personal medicine is, therefore, considered to be enormous, with equally large life quality improvements to be gained. This possibility, however, clashes with the interest of Big Pharma, that prefers one standardized substance for each patient and illness. Big Pharma do not have the incentives to be pioneers in breaking their large scale producer advantage until challenged by small biomedical niche players that make successful inroads into their markets. Hence, niche players such as Eurona are also a socially valuable competition factor.

Apparently Eurona was too early and/or venture capitalists did not understand the business idea . Even though Eurona announced that its first diagnostics product capable of predicting which patients would respond positively to a particular blood pressure inhibitor would be on the market the same autumn (*Svenska Dagbladet*, June 7, 1999), the thin Swedish venture capital market went dead in 1999. UK Gemini picked up Eurona at a low price from its "supporting" venture capitalists.

Several experts interviewed within and around the company held widely different opinions about the time horizon for take off, from one half year to ten years (Eliasson, 2003). Gemini, a smaller company with money, was primarily interested in the testing competence of Eurona to support its analysis of twins, and shelved (at least temporarily) whatever was left of the personalized medicine project. Gemini was introduced on the Nasdaq in 2000. During an interview with Gemini in late 2000, it was indicated that Gemini probably would have to complement its technology through strategic acquisitions. Seventy percent of the sources of new technology, however, reside on the West coast of the US. In 2001, Gemini was acquired by US *Sequenom* on the West coast, one of the new players in personal medicine, a field now considered one of the most promising in medical businesses.

Case II: Perbio Science

Until recently, Perbio Science was a mostly US based, but Swedish owned, company in biotech supplies, headquartered in Sweden (Helsingborg). Earlier the company had been a division in the Perstorp chemical group, which had acquired Pierce Chemical and Athos Medical in the US during the 1990s. Perbio was spun off to Perstorp's owners in the late 1990s.

Perbio Science considers itself to be the supreme performer in protein cell culture, which accounts for more than 50 percent of sales, the world leader in voice protheses (the Swedish part of the company and 9 percent of sales), and a major player in bioresearch supplies of reagents, kits and services for protein studies to both industry and university laboratories.

Perbio management had long been on the lookout for a solution to its strategic problem of deciding whether to invest and grow organically, grow aggressively through complementary strategic acquisitions or wait to be acquired at a high price. Organic growth was considered too slow and too risky. To be acquired would mean a US buyer only. Europe and Sweden did not have the complementary receiver competence to commercialize the potential and be willing to pay a high price. Complementary strategic acquisition objects, in addition, could not be found in Europe. Lack of local Swedish management competence on which to build an acquisition program was also embarrassing. To grow from a technology base in Sweden, therefore, was no longer considered a viable solution.

There had been an opportunity to create a growth base in Sweden in the mid 1990s. Pharmacia had just merged with US UpJohn and was looking for a partner to Pharmacia Biotech. Discussions were conducted with Perstorp, which saw an opportunity to combine Perbio's world leading cell culture technology with the world leading protein separation technology of Pharmacia Biotech into a global cell-culture company. The management of Pharmacia Biotech, however, considered Perbio too small a player and balked at the plans of Perbio management immediately to unload the larger but less profitable instrument activity, considered alien to a cell culture company. The instrument activity was also considered a potential financial burden to the new company that would draw disproportionally large management attention and would require very large investments to become profitable , circumstances that would hold back growth in the market segments where the new company would have the best opportunities. Rapid technology development was one reason for the very large investments needed in instruments, an area that was easily overrun by competitors. Instruments, furthermore, did not generate the desired cash flow, but needed the later sales of consumption chemicals. The whole deal evaporated when UK Amersham acquired Pharmacia Biotech in 1997 and renamed the company Amersham Biosciencies, which was in turn acquired by General Electric's Medical Businesses in 2003. GE develops and manufactures instruments, such as medical scanners, that already use contrast chemicals produced by Amersham.

Perbio itself was acquired by US Fisher Scientific in 2003 for 155 kronor per share (*Dagens Industri*, June 27 and August 28, 2003). For Perstorp, the owners of which had acquired Perbio for 35 kronor when it was introduced as a separate company on the Stockholm stock exhange in 1999, this more than compensated for the bad stock market performance of Perstorp itself (*Dagens Industri* Nov. 8, 2000).

The choice menu for business combinatorics is great. Gothenbourg-based Nobel Biocare, a company formed from the diversification of the defence firm Bofors in the early 1980s, opted for the long and risky road of organic growth based on the Brånemark method of titanium dental implants (Fridh, 2002), only to find itself almost overrun in the 1990s by a Swiss imitator. Nobel Biocare sold out cheaply to a Swiss medical investor group when the stock market declined in 2001.

Case III: Karo Bio[13]

Karo Bio is a biotech firm that operates as an intermediary in the markets between large pharmaceutical firms and university research. Even though the company has not fared well in the market recently, it is principally interesting here as a hybrid of market and hierarchical organization. Karo Bio's business idea is to look for and discover business opportunities in academic research laboratories and to develop them commercially up to the stage of "routine" clinical testing, when the projects can be understood sufficiently well for a large pharmaceutical firm to be interested. Karo Bio contributes both entrepreneurial and venture capital competence (see Table 3) to upgrade the commercial value of promising academic research projects. Karo Bio thus represents an intermediate organizational solution to deal simultaneously with both operational focus and innovative variety, through outsourcing the innovative and entrepreneurial function. Karo Bio then lowers the risk of committing errors of both type I and II by exposing each project to a more competent commercial evaluation than would otherwise have occurred. The project is pulled out of a commercially incompetent academic environment and prevented from being narrowly inspected and rejected in a big company environment, and so winners are probably saved.

---

[13] A more detailed presentation can be found in Eliasson and Eliasson (1997, pp.151).

The complexity of this more varied evaluation is illustrated by the fact that Karo Bio (still) has had (1) to specialize in a few diseases that involve nuclear (hormone) receptors and (2) to form complementary partnerships with academic labs, specialized firms or even industrial customers (see item 6 in Table 3) to broaden its competence base. The problem has even been raised that KaroBio has opted for the wrong screening technology (*Dagens Industri* Aug.20. 2003). The business idea is to make drug screening and discovery more efficient through a more innovative and efficient pre-screening process than that of the big pharmaceutical companies. While the big pharmaceutical companies are excellent at clinical testing of *given* substances for *known* biological effects, this excellence is a foolproof method for missing radically new winners. So Karo Bio looks actively for winners and then applies its own, more efficient methods[14] to narrow the number of promising candidate substances.

Projects may be packaged as a company, but Karo Bio prefers to offer a license deal, thus illustrating the importance of competent customers (industrial buyers). KaroBio claims to understand the potential of a project better than the customer, so why sell it for the low price an "incompetent" customer is offering? If you can finance development yourself, wait and license. Then you can increase the price when the buyer finally apprehends the situation. Again, this also illustrates the importance of a competent venture capitalist, who understands better than the big industrial customer how promising the project really is. Such competence is rare externally, but it exists, and Karo Bio aims for reaching the level needed to be its own sustainable venture capital provider.

5 The existence of a market for strategic acquisitions – the third problem

The big firm has the money but not the capacity to create and to bring radically new ideas to the attention of its decision makers. The small firm/ entrepreneur has the ideas but not the money. In between the two there is the market for innovations (Day et al., 1993), in which radically new ideas are developed as far as is needed for an industrialist to understand the commercial potential. This development is intermediated by the actors of the competence bloc, notably the actors in the venture capital and exit markets. For the little firm to capture the rents of its own innovative capacity (competence capital), it is dependent on the efficient functioning of these two markets. Speedy access to venture capital often decides the outcome. The small innovative firm, therefore, depends more than other firms on the competence of the actors in the financial markets to understand what they are doing.

There are six principally different strategies for the small, innovative firm to pursue (Eliasson, 2002a and Table 4). It can (1) go slowly and organically, at the rate internal finance permits, (2) grow aggressively through external acquisitions, (3) opt for internal growth, based on external venture capital, (4) aim at being strategically acquired by a large firm, (5) develop technologies for licensing or (6) do contract work. The categories in Table 4 correspond to a different assignment of ownership and control rights and/or different contracts, each involving different

---

[14] For more information on these methods, see Eliasson and Eliasson (1997).

Table 4. Strategic choices for the small innovative firm

1.	Do it alone, slowly on internal funds and risk going bankrupt or being imitated early by a big company.
2.	Grow internally and share the risks and profits with an external venture capitalist, often unfavorably.
3.	Grow aggressively through strategic acquisitions or external venture capital and/or paying with own stock and dependent on the competence of actors in the stock market to value your company.
4.	Aim for being strategically acquired at a high price.
5.	License your technologies.
6.	Do contract work

Source: Eliasson (2002a).

risk. The risk level decreases as you go down Table 4, but potential profits increase as you go the other way. Each strategic choice, or each combination of choices, corresponds to a different definition of the hierarchy or the firm.

Strategies (2) and (3) can be combined. It is quite common among small biotechnology firms (also cf. US high tech firms in Eliasson, 2000, p.234) to aim for internal growth, but to sell out if a suitor offers a sufficiently high price. The most demanding and the most risky, but also the potentially most rewarding, approach of the small innovative firm with a potential inhouse winner is to grow through a combination of early venture capital and own strategic acquisitions to complement its own technologies to reach industrial scale production and distribution quickly. This was the early ambition of Perbio Science above. As we have concluded already, for the risks to be reasonable under this strategy, a vertically complete and a sufficiently varied competence bloc has to be in place. Only then can the potential winner confidently continue searching for new resources on its own. There are increasing returns to continued search. The objectives of a new start- up firm are not independent of its sources of finance and the agreements on risk sharing. There are three principally different ways of funding the commercialization of a radically new innovation (Eliasson, 2002a).

(a) *High Risk* venture (items 2 and 3 in Table 4).
Build the company to industry level on external venture financing.
(b) *Medium Risk* venture (item 4).
Aim for product being strategically acquired by large company.
(c) Low Risk venture (items 5 and 6).
License or do contract work.

The first high risk venture requires the support of a complete competence bloc. The second medium risk venture requires the existence of fully developed markets for strategic acquisitions, notably for bidding up prices of acquired objects sufficiently to establish incentives for innovators. The low risk ambition requires that there are sophisticated and large customers for technology in the market. While the US offers the whole range of options a, b and c, Europe offers c and only to some extent b, but not much of a.

The market for strategic acquisitions allows the small, advanced biotechnology firm to complement its competence and technologies through acquiring a firm or part of a firm. This is a way to increase its rate of growth, compared to internal development of the same technologies and to capture the market ahead of imitators. The large pharmaceutical company can acquire know-how it has been unable to develop internally through firm acquisitions. Small, innovative firms can supply their technologies in the same market at high prices if many big firms compete for their technology. If a profitable selling opportunity arises, the small, innovative firms growing internally through venture capital finance and/or through acquisitions might opt out of that strategy and sell out. The more options, the higher the probability that winners are identified and allocated to the right users.

A strategic acquisition is a means for a firm to solve a particular business problem. It is, however, also a matter of interest for the policy maker, since the non-availability of a dynamically efficient market for strategic acquisitions limits new firm formation both technologically and financially, and, hence, growth. We have shown (in terms of competence bloc theory) how the large firm can internalize "over the market" a dynamically efficient integration of innovative capability and large-scale operations efficiency. Strategic acquisitions, furthermore, are an increasingly important channel of technology diffusion. This time we see an innovative organizational solution that bridges the inability of large operations-oriented firms to be creative and the difficulties of the creative (innovative) small firm to capture the rents from its winning innovations. The market for strategic acquisitions creates value where technology might otherwise be wasted. The venture capital market and the market for strategic acquisitions embodies industrial competence within finance, thereby raising the competence each project is exposed to in the evaluation process of the competence bloc. It is in the interest of all parties in the game (seller, buyer and government) that a competitive market for strategic acquisitions develops where the value of the winner (firm) is bid up to the highest price the most competent acquiring firm is willing to pay. Cheap acquisitions will then be stopped. This market has to be global, and it appears to be the case (see Eliasson, 2001a) that high technology can be acquired at a distance, since only competent customers are in this market. The market for strategic acquisitions enhances the flexibility of choice for the small firm to commercialize winning technologies on their own, and for the large and less innovative companies to access new technologies through acquisitions. This was the third problem.

6 The diffuse notion of a hierarchy – the theory of the firm revisited

The fathers of economics were not really interested in industrial dynamics, but rather in the "higher level" policy problems of Government. The economists of those days were satisfied with discussing technologies and industries defined as aggregates, and possibly factories. The role of live firms in wealth creation was rarely addressed, except as in Smith (1776) referring both to the joint stock company as a socially negative privilege or monopoly, and to the importance of new firm formation for exposing these monopolies to competition.

Industrial monopoly formation and antitrust problems brought the firm into policy focus in the late 19^{th} century in the US, and then again in the 1930s. Industrial organization theory developed from this policy base (Scherer, 1980), but was rapidly (in the 1980s) integrated with the neoclassical tradition. Since there was no place for firm dynamics in static equilibrium, no distinction was made between the innovator, the entrepreneur , the venture capitalist and the industrialist. They were either bunched together in a firm or sector production function or assumed to be fully outsourceable (in so far as their knowledge mattered) in perfect markets (see for instance Fama, 1980).

Coase brought an end to this tradition in 1937, a contribution the importance of which was not realized until decades later. In practice, all the (competence) functions of the competence bloc are now coming apart in the markets in a truly Coasian fashion, and new C&C technology is playing a critical role in making such distributed, still integrated production both possible and profitable . Outsourcing, however, is not the same as the separability of Fisher (1930), which is one of the corner stones of modern financial economics. Fisherian separability is incompatible with dynamic or Schumpeterian efficiency, since striving to reduce transaction costs within the extended firm boundaries involves attempting to minimize the extended definition of transaction costs by reorganizing the limits of the firm. This also takes us outside the mainstream definition of the firm as formulated by Holmstroem and Tirole (1989) since that definition is based on transaction costs minimization over a given firm hierarchy. Integrated production based on modularization and outsourcing over the market can be organized very differently and some of all possible organizational designs exhibit very large, positive systemic productivity gains (Eliasson, 1996b). A systems responsible firm coordinates the whole, and the systems coordinating competence is one of the strong competitive advantages of the advanced western industrial firms. The dynamics of integrated production, however, still diffuse the notion of the firm as a well defined and centrally controlled hierarchy.

Integrated production defines the *extended firm* based in an increasingly sophisticated system of specialized subcontractors. Integrated production requires control rights that can be organized through the assignment of appropriately designed property rights (ownership) and contracts superimposed on physical manufacturing and distribution. The dynamics of contracting and recontracting of the extended firm is, however, moved by people with competence (Eliasson, 1990a), forging temporary configurations of property rights in the market for strategic acquisitions. It is an economically viable entity to the extent it can be configured to lower transaction (read information and communications) costs. This may be possible if (1) the competence bloc is vertically complete and sufficiently varied horizontally and (2) if transaction costs are understood to include the potential loss of winners. The large potential systemic productivity effects that a competent organizer can realize are an incentive for the formation of distributed and integrated production. Hence, there will be a demand for supporting markets for strategic acquisitions to develop. We can also conclude theoretically that the development of such markets for strategic acquisitions to support the free formation of extended firm arrangements will be a contributing factor behind the successful formation of a New Economy. We have already indicated that deficiencies on this score may be what keeps the US economy ahead of Europe.

7 Is Europe a bunch of laggards?

The markets for innovation, entrepreneurship and venture capital in their developed form (Day et al., 1993) are fairly new. One might safely say that the US is the only economy that features an advanced venture capital industry capable of evaluating and financing large scale, radically new innovative projects (Eliasson, 2003). There are several reasons for this. *First*, the US economy, notably California, has a larger concentration and diversity of wealthy people than any other country, people who have become rich through private industrial activities, notably in the new industries. *Second*, deregulation of the US insurance markets in the 1970s allowed the insurance industry to enter the venture capital market. The supply situation then was dramatically changed for the better, notably through the creation of very deep exit markets. *Third*, the early start of new high technology industries expanded the set of industrially experienced and rich individuals that now populate the venture capital industry. Together, this means that the formation of sophistical markets for strategic acquisitions began in the US. Overcoming the handicap of financial markets lacking industrial experience in Europe is no easy task, and it is not supported by the political ambitions in Europe of making the formation of private wealth through innovative industrial ventures difficult. The scarcity of competent venture capitalists who understand radically new technology might mean that small European start-up firms will have difficulties funding both their own internal expansion and an aggressive expansion through own acquisitions because “incompetent” local venture capitalists take too long to make decisions and/or take too large a share of the capital gains. There will be a bias towards selling out. For a small country, this will probably be to a foreign suitor as was the case with Eurona and Perbio Science. This problem is interesting because both Sweden and Europe are as advanced in both health care technology (Eliasson, 1997; Eliasson and Eliasson, 2002b) and in agricultural biotechnology (Eliasson, 2002a) as the US, but both lack commercial and industrial competence, including venture capital competence, compared to the US. Hence, there might be a bias in the flow of industrialized technology from Europe to the US.

Already in the early 1980s, US venture capital was eying Swedish Pharmacia and Danish Novo, at the time erroneously believed to be on the verge of a technology breakthrough in biotech (Eliasson, 2003). It was also suggested (Eliasson, 1997) that US venture capital be invited to invest in Sweden, to compensate for the lack of commercial and venture capital competence in Swedish health care. Foreign venture capital would help create a market for strategic acquisitions, raising the economic value of locally developed technology that might otherwise be wasted.

One could argue that the European economies at least feature a large number of potential industrial buyers in the market for strategic acquisitions. This positive factor is, however, diminished by the relative dominance in Europe of old, mature and most likely conservative firms. There is a reason for this. Legal rules and policy makers in Europe bias incentive systems in favor of large firms, meaning, by definition, a bias against the small, innovative firms. One reason for this negative bias in incentive systems has been the political ambition in some countries to control private industry, which can only be accomplished if the firms are few and large (Eliasson, 1998a, pp.64 ff.). Another reason has been a concern about unemployment and the assumed protective internal labor markets of large firms. In conclusion, then, the *receiver competence* at the economy-wide level, or the capacity of the economy at large to take on and build new businesses on new technologies, is not only deficient in Europe compared to the US because of incomplete and horizontally less varied competence blocs, but also because of a political reluctance to allow the markets to push freely for change. Europe, therefore, is more exposed to the risks of committing business mistakes of type II (i.e. losing the winners) than the US. The gestation period to correct the situation (build competence and change policies) and to see positive results may, however, be too long for political patience to survive. The policy catch is paradoxical. Austrian/European economists were the first to realize the nature of a dynamically growing economy. This understanding has been washed out in mainstream textbook economic theory and research in favor of an economic theory refined to perfection in the US that is more a theory of central planning than it is of a dynamic economy. But real industrial dynamics is to be found in the US, more so than in Europe.

References

Acs ZI, Audretsch DB (1988) Innovation in Small and Large Firms: An Experimental Analysis. American Economic Review, 78: 678–690

Aghion P, Tirole J (1994) On the Management of Innovation. Quarterly Journal of Economics, 109: 1185–1207

Albrecht JW, Braunerhjelm P, Eliasson G, Nilson J, Nordström T, Taymaz E (1992) Moses DataBase. IUI

Arrow KJ (1962) Economic Welfare and the Allocation of Resources for Invention. In: Nelson R (ed.) The rate and Direction of Inventive Activity: Economic and Social Factors, 609–626, Princeton University Press, Princeton

Arrow KJ (1965) Aspects of the Theory of Risk-Bearing. The Academic book Store, Yrjö Jansson Lectures, Helsinki

Ballot G, Taymaz E (1998) Human Capital, Technological Lock-in and Evolutionary Dynamics. University of Michigan Press, Ann Arbor

Burenstam-Linder S (1961) An Essay on Trade and Transformation. Uppsala

Coase RH (1937) The Nature of the Firm. Economica, 4, 13-16: 386–405

Cohen WM, Levinthal DA (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly, 35: 128–152

Dahlman CJ (1979) The Problem of Externality. Journal of Law & Economics, 2: 141–162

Darby M, Zucker L (2002) Going Public when you can in Biotechnology. NBER Working Paper Series, No. 8954

Day RH, Eliasson G, Wihlborg C (1993) The Markets for Innovation, Ownership and Control. IUI, Stockholm and Amsterdam

Demsetz H (1969) Information and Efficiency: Another Viewpoint. Journal of Law and Economics, 12: 1–22

Desai MA, Foley CF, Hines, Jr JR (2002) International Joint Ventures and the Boundaries of the Firm. NBER Working paper Series, No. 9115

Eliasson Å (2002a) A competence bloc analysis of the economic potential of biotechnology in agriculture and food production. In: Carlsson B (ed.) New Technological systems in the Bio Industries, Kluwer Academic Publishers, Boston, Dordrecht, London

Eliasson G (1976) Business Economic Planning. Theory, Practice and Comparison. John Wiley & Sons, London-New York-Sydney-Toronto

Eliasson G (1984) Informations- och styrsystem i stora foeretag. In: Eliasson G, Fries H, Jagren L, Oxelheim L (eds.) Hur styrs storfoeretag? En studie av informationshantering och organisation, IUI, Stockholm

Eliasson G (1985) The Firm and Financial Markets in the Swedish Micro-to-Macro Model – Theory, Model and Verification. IUI, Stockholm

Eliasson G (1986a) Företaget som skapare och förvaltare av mänskligt kapital (The firm as an innovator and manager of human capital). In: Eliasson G, Carlsson B, Deiaco E, Lindberg T, Pousette T (eds.) Kunskap, information och tjänster. En studie av svenska industriföretag, IUI and Liber and Statens Industriverk

Eliasson G (1986b) Innovative Change, Dynamic Market Allocation and Long-Term Stability of Economic Growth. Working Paper No. 156, IUI, Stockholm

Eliasson G (1987) Technological Competition and Trade in the Experimentally Organized Economy. Research Report No. 32, IUI, Stockholm

Eliasson G (1990a) The Firm as a Competent Team. Journal of Economic Behavior and Organization, 13, 3: 273–298

Eliasson G (1990b) The Knowledge Based Information Economy, chap. The Knowledge-Based Information Economy. IUI, Stockholm

Eliasson G (1991) Modeling the Experimentally Organized Economy – Complex Dynamics in an Empirical Micro-Macro Model of Endogenous Economic Growth. Journal of Economic Behavior and Organization, 16, 1-2: 153–182

Eliasson G (1992) Business Competence, Organizational Learning and Economic Growth: Establishing the Smith Schumpeter Wicksell Connection. In: Scherer FM, Perlman M (eds.) Entrepreneurship, Technological Innovation, and Economic Growth. Studies in the Schumpeterian Tradition., The University of Michigan Press, Ann Arbor

Eliasson G (1996a) The Firm, Its Objectives, Its Controls and Its Organization. Kluwer Academic Publishers, Boston, Dordrecht, London

Eliasson G (1996b) Spillovers, Integrated Production and the Theory of the Firm. Journal of Evolutionary Economics, 6: 125–140

Eliasson G (1997) Hälso- och sjukvårdsindustrin – ett kompetensblock med stor affärspotential. KTH, Stockholm

Eliasson G (1998a) From Plan to Market. Journal of Economic Behavior and Organization, 34, 1: 49–68

Eliasson G (1998b) The Nature of Economic Change and Management in the Knowledge-Based Information Economy. KTH, TRITA IEO-R 1998:19

Eliasson G (2000) Industrial Policy, Competence Blocs and the Role of Science in Economic Development. Journal of Evolutionary Economics, 10: 217–241

Eliasson G (2001a) Decentralizations of Decisions, Outsourcing and the CHQ Functions. KTH, TRITA

Eliasson G (2001b) The Role of Knowledge in Economic Growth. In: Helliwell J (ed.) The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being, HRDC

Eliasson G (2002b) Den Nya och Omedelbara Ekonomin- ett Internet perspektiv (The New and Immediate Economy – a Internet Perspective). Vinnova, Teldok, Stockholm

Eliasson G (2003) The Venture Capitalist as a Competent Outsider. In: Alho K, Lassila J, Ylä-Antilla P (eds.) Economic Research and Decision Making-Essays on Structural Change, Growth and Economic Policy, ETLA, Helsinki, an earlier version was published under the same title by Stockholm: KTH, TRITA-IEO R 1997:06

Eliasson G, Eliasson Å (1996) The Biotechnological Competence Bloc. Revue d'Economie Industrielle, 78, 4

Eliasson G, Eliasson Å (1997) The Pharmaceutical and Biotechnological Competence Bloc. In: Carlsson B (ed.) Technological Systems and Industrial Dynamics, Kluwer Academic Publishers, Boston, Dordrecht, London

Eliasson G, Eliasson Å (2002a) Competence Blocs in the Experimentally Organized Economy. MIMEO KTH and TRITA, Stockholm

Eliasson G, Eliasson Å (2002b) Competence in Health care. MIMEO KTH

Eliasson G, Granstrand O (1985) Venture Capital and Management A Study of Venture Development Units in Four Swedish Firms. Mimeo, Stockholm

Eliasson G, Johansson D, Taymaz E (2002) Firm Turnover and the Rate of Growth – Simulating the Macroeconomic Effects of Schumpeterian Creative Destruction. Paper presented at the 9th Congress of the International Joseph A. Schumpeter Society (ISS), march 28–30,2002,. Gainesville, Florida, USA

Eliasson G, Johansson D, Taymaz E (2004) Simulating the new Economy. Forthcoming in Structural Change and Economic Dynamics

Eliasson G, Taymaz E (2000) Institutions, Entrepreneurship, Economic Flexibility and Growth – experiments on an evolutionary model. In: Cantner U, Hanusch H, Klepper S (eds.) Economic Evolution, Learning and Complexity – Econometric, Experimental and Simulation Approaches, Physica Verlag, Heidelberg

Eliasson G, Wihlborg C (2003) On the macroeconomic effects of establishing tradability in weak property rights. Journal of Evolutionary Economics, 13: 607–632, see also this volume

Fama EF (1980) Agency Problems and the Theory of the Firm. Journal of Political Economy, 88, 2: 288–307

Fisher I (1930) The Theory of Interest. Macmillan, New York

Fridh A (2002) Dynamics and Growth: The Health Care Industry (dissertation. KTH, Stockholm

Hart O, Moore J (1990) Property Rights and the Nature of the Firm. Journal of Political Economy, 98, 6: 1119–1158

Holmstroem B (1993) Agency Costs and Innovation. In: Day RH, Eliasson G, Wihlborg C (eds.) The Markets for Innovation, Ownership and Control, IUI, Stockholm

Holmstroem B, Tirole J (1989) The Theory of the Firm. In: Schmalensee R, Willig RD (eds.) Handbook of Industrial Organization, Elsevier, Amsterdam

Johansson D (2001) The Dynamics of Firm and Industry Growth – the Swedish Computing and Communications industry, Doctoral thesis. KTH, Stockholm

Jovanovic B, Rosseau PL (2002) Mergers as Reallocation. NBER, Working Paper Series, No. 9279 (October)

Lamoreaux NR, Sokoloff KL (2002) Intermediaries in the U.S. Market for Technology, 1870–1920. NBER Working Paper Series, No. 9017 (June)

Lerner J, Merges RP (1997) The Control of Strategic Alliances: An Empirical Analysis of Biotechnology Collaborations. NBER, Working Paper Series, No. 6014 (April)

Lewis TR, Sappington DEM (1991) Technological change and the Boundaries of the Firm. American Economic Review, 81, 4: 887–900

Marshall A (1890) Principles of Economics. London

Marshall A (1919) Industry and Trade. London

Menger C (1872) Grundsaetze der Volkswirtschaftslehre. Vienna

Pelikan P (1986) Why Private Enterprise? Towards a Dynamic Analysis of Economic Institutions and Policies. IUI Yearbook 1986 1987, Stockholm

Pelikan P (1988) Can the Imperfect Innovation Systems of Capitalism Be Outperformed? In: Dosi G (ed.) Technical Change and Economic Theory, Pinter Publishers Ltd., London

Salter WEG (1960) Productivity and Technical Change. Cambridge University Press, Cambridge, MA

Scherer FM (1980) Industrial Market Structure and Economic Performance. Houghton Mifflin, Boston

Schumpeter J (1911) The Theory of Economic Development (English edition 1934), vol. XLVI of *Harvard Economic Studies*. Harvard University Press, Cambridge, MA

Schumpeter J (1942) Capitalism, Socialism and Democracy. Harper & Row, New York

Schumpeter J (1954) History of Economic Analyses. Allen & Unwin [1986], London

Simon H (1957) The Compensation of Executives. Sociometry, 20: 32–35

von Hayek FA (1937) Economics and Knowledge. Economica, 4: 33–54

von Mises L (1949) Human Action. Contemporary Books, Chicago

The growth of commercialization – facilitating organizations and practices: A Schumpeterian perspective

Sten Thore[1] **and Robert Ronstadt**[2]

[1] The IC^2 Institute, The University of Texas at Austin, 2815 San Gabriel, Austin, Tx 78705, USA
(e-mail: thore@mail.telepac.pt)
[2] The Technology Commercialization Institute, The University of Boston, Mass., USA
(e-mail: ronstadt@bu.edu)

Abstract. During the long economic upswing of the 1980s and 1990s, the successful commercialization of new technology went together with the appearance of new financial vehicles, new organizational forms, and new practices promoting and mediating the transfer of the new technology from lab to market. They included business incubators, technology parks, venture capitalist firms, operators specializing in mergers and acquisitions, venture funds, and initial public offerings in the stock market. Commercialization-facilitating organizations and practices like these (here called "commercialization facilitators") are themselves born by dynamic processes in a capitalist economy that can be analyzed in Schumpeterian terms. We discuss at some length a unique university-based institution that has an impressive track record of creating and operating new facilitating models: the IC^2 Institute (Innovation, Creativity and Capital) of the University of Texas. During a twenty-five year period, IC^2 came to be instrumental in the conversion of the local economy to the high tech age. The Institute's activities span the range of the technology transfer and commercialization process, from the development and dissemination of new knowledge to the actual running of business incubators. We identify the IC^2 Institute as a "second order" facilitator and discuss its possible global evolution into a "third order" facilitator.

Keywords: commercialization facilitators – commercialization infrastructure – research to wealth continuum

1 Introduction

During the long economic upswing of the 1980s and 1990s, the successful commercialization of new technology went together with the appearance of new financial

---

Correspondence to: Sten A. Thore, post box 1284, 8400 Carvoeiro, Portugal

vehicles, new organizational forms and new practices promoting and mediating the transfer of the new technology from lab to market. They included business incubators, technology parks, venture capital firms, operators specializing in mergers and acquisitions, venture funds, and initial public offerings (IPOs) on the stock market. Commercialization-facilitating organizations and practices such as these (here called "commercialization facilitators") are themselves born by the favorable commercial conditions during a prolonged business boom; at the same time, they sustain and reinforce the regional impacts of the boom. In the scramble to convert the progress of basic research and new knowledge into new products, commercialization facilitators promote and speed up the transfer of new technology from the (academic, government, or commercial) laboratory to its successful introduction in the market.

A crucial aspect of Schumpeter's understanding of the business boom was the accumulation of new and as yet unrealized technological opportunities during the preceding downturn, and the cascading of innovations during the upswing. During the boom years of the 1990s, this clustering of innovations was reinforced in most regions by the appearance and operation of powerful commercialization facilitators.

Section 2 below identifies a list of model types of commercialization facilitators and discusses their characteristic modes of operation. Each facilitator typically addresses the needs of some particular group of potential entrepreneurs (located in some geographical region, or in some particular industry, such as the software industry) at some particular time during the life cycle of a new company (during the initial research phase, during test marketing, after the first round of equity financing etc.). Some facilitators promote the creation of "technoports" or high tech regions that harness powerful positive externalities spurring the growth of new technology and new companies.

Commercialization facilitators are instances of economic institutions and as such would never attract the attention of main-line economic theory. Recently, however, several authors have argued for the need to bring institutions into evolutionary growth theory, among them R.R. Nelson (Nelson, 2002). Distinguishing between "physical technologies" and "social technologies," Nelson calls for an understanding of the growth of social technologies during the evolutionary process. Others have pointed at the role of the knowledge-intensive service industry in fuelling the high technology boom. Indeed, business incubators and venture capitalists all offer knowledge-intensive business services[1] to their tenants and clients. As we see things, a novel *commercialization infrastructure* is gradually being put in place in the modern high tech economy. This infrastructure features a new kind of economic agent – commercialization facilitators.

Wynarczyk and Wynarczyk (2002), discuss the operations of the European Business and Innovation Centers (BICs) in economically less favored regions and find their rationale in the efforts of Schumpeterian entrepreneurs to break the circular flow and inertia of established practice.[2] Our own work aims further, demonstrat-

---

[1] So-called "KIBS," see Lundvall and Borrás (1997).

[2] For further comments on the BIC program, see Sect. 2 below.

ing that new commercialization infrastructure can spearhead high-tech innovative growth and the construction of the future technopolis.

The commercialization facilitators, in their turn, are spawned, established and grown through dynamic processes in a capitalist economy. We argue in Sect. 3 that the growth of commercialization facilitators should be seen as endogenous to the business cycle. The founding of new facilitators and their operations responds, like all entrepreneurship, to opportunities and favorable conditions. Indeed, it is helpful to understand the emergence of new commercialization facilitators in terms of "institutional innovation" or the development of new "institutional technology". Just as Joseph Schumpeter saw industrial innovations wax and wane in an endogenous interplay over the business cycle, we shall propose a cyclic format where industrial evolution and institutional evolution are coupled together and reinforce each other. Successful commercialization facilitators spur the growth of industry; and also evolve in response to the needs of the market.

Section 4 turns to the empirical record. We discuss a unique university-based institution with an impressive track record of creating and operating new facilitators: the IC^2 Institute (Innovation, Creativity and Capital) of the University of Texas. During a twenty-five year period, IC^2 came to be instrumental in the conversion of the Austin economy to the high tech age. The Institute's activities span the range of the technology transfer and commercialization process, from the development and dissemination of new knowledge to the actual running of business incubators.

Section 5 synthesizes and concludes by returning to the long-run perspective and the evolution of capitalism and entrepreneurship itself. Rather than being dominated by gradually larger corporations run by large management teams relying on automated decision-making and making routine investment decisions (as Joseph Schumpeter presumed), the capitalist economy has become invigorated by large numbers of small-scale entrepreneurs generating and commercializing new advanced technology. We believe that the growth of commercialization facilitators is an important part of this revival of the age of the entrepreneur. As new commercialization-facilitating organizations and practices continue to evolve, the scope of entrepreneurship and the pace of economic growth will be speeded up even further.

2 A catalogue of commercialization facilitators

The commercialization of new technology converts basic and applied knowledge into economic values: sales, revenues, profit, and enhanced market capitalization of the innovating firm. Figure 1 illustrates the knowledge-to-wealth continuum extending from research via venture creation and/or licensing to the creation of new wealth.

The commercialization facilitator occupies an intermediate position in this continuum, promoting and assisting the development and commercialization of new products or processes at one or several subsequent steps.

The commercialization of new products and new technologies can be promoted, facilitated and accelerated in many ways. Some well-known *institutions* and some evolving *institutional developments* are listed below.

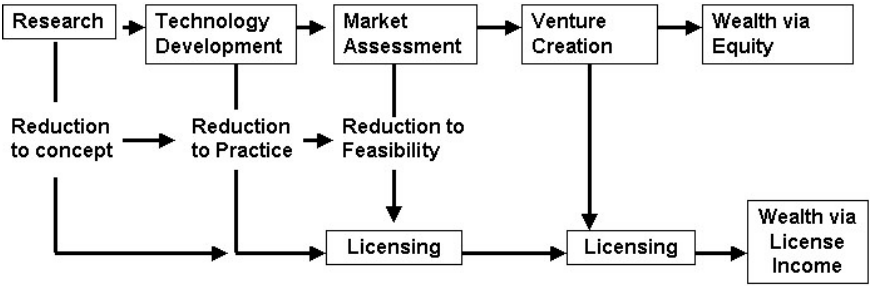


Fig. 1. The Knowledge-to-Wealth Continuum

- *The business incubator.* Incubators can be operated on a "for profit" or "non-profit" basis by local governments, academic institutions, or private businesses. To incubate means to maintain controlled conditions favorable for hatching or developing. The incubator houses and aids start-up companies. It provides manufacturing and office space and other amenities; it may offer secretarial support, computer services and management consulting services. It may physically house start-ups or only offer services (called "incubation without walls"). Hopefully, after a few years, the tenants at the incubator will "graduate," moving out from the incubator and making it on their own.
- *The venture capital firm.* The venture capital industry specializes in high-risk equity investments. It invests equity money in fledgling companies that it takes under its wings. Many venture capitalists (or venture capital companies, as the case many be) center their attention on emerging high technology corporations whose expansion is restricted by lack of equity capital. With few assets and without proven cash flows, such corporations are often unable to raise capital from conventional sources
- *Venture capital funds and limited venture capital partnerships* that private individuals can buy into. A partnership has a general partner, usually the venture capitalist's investment firm, which acts as manager. The limited partners – the investors – are passive. During the 1980s, investors poured more than $100 million into partnerships that put the money into investments such as apartment and office buildings, airplane leasing, oil wells and able television. But as many of these investments went sour, so did the partnerships.
- *The technology park and the "technopolis"* (the techno-port or the techno-city). The first technopolis was Silicon Valley. In the formative years, the proximity to Stanford University was important. The University had started a research park, – the Stanford Industrial Park – and Hewlett-Packard was among the first firms to locate there. There arose in the Valley at an early point the required infrastructure: suppliers, markets, and financiers. The technopolis feeds on the development of new technology – a continuous flow of new products and new designs that leave the laboratories and enter the marketplace.

We also list some important commercialization-facilitating *procedures or practices.*

- *Strategic alliances.* U.S. corporations have entered into thousands of research coalitions with other partners, both domestic and foreign. By pooling their resources of research and product development, two corporations can create a temporary match powerful enough to unlock the next step in the accelerating technological race.

- *Mergers and acquisitions*. One of the most important sources of venture funds is tapped when an existing company buys a minority stake in a start-up company. Large companies often have particular strengths in marketing, distributing and selling, as well as sources capital. Their weaknesses are apt to be innovation and speed in getting new products to market. The intuitive reaction of the manager of a large company, looking for specialized technologies that are not available in-house, would be to find a company in the marketplace that has what is needed, and to buy it. The wise move may not necessarily be to buy it outright – it may be smarter to acquire only a minority holding to make sure that the existing creative management team stays in place.
- *IPOs (Initial public offerings)* in the stock market offer an opportunity to the common investor to participate in the launching of new ventures and new technologies. The issuer of the stock may be a successful startup that has reached a critical point in its expansion, needing an additional injection of equity capital. Or, a large corporation may decide to *spin off* a separate entity, charged with developing existing or new technology under a new aegis. (Example: AT&T spinning off Lucent Technologies). Many recent IPOs in information technology and on the Internet (so-called dot-com companies) are startups still encumbered by heavy past and present development costs. Their sales and earnings may be growing, but there is still a long way to go before the IPO reaches positive profits. – Finally, the IPO may be a pure research venture, such as in biotechnology. It may yet be too early to say whether the research will ever result in a marketable process or product. The IPO has no current sales, no revenues, and certainly no profit.

The time lines in Fig. 2 gives an approximate dating of the emergence of some important categories of commercialization facilitators. Inspecting Fig. 2, perhaps the most important observation is the late date of most of these innovations. By 1980, there were only 12 business incubators in the US, but the number rose to 850 incubators by 2001 (see Wiggins and Gibson, 2003). The venture capital (VC) industry is of similar recent origin. The first venture capital company, ARD, was created in 1947; however most firms emerged in the 1980's and numbered approximately 500 by the early 1990's, a number that stabilized by the turn of the century. Figure 2 also illustrates the rapidly diversifying range of arrangements, such as business "angels" (wealthy individuals funding startup companies) and university technology licensing arrangements (TLO's).

Other arrangements were less successful. The landmark R&D consortium MCC (Microelectronics and Computer Technology Corporation) in Austin, pooling the research efforts of all major computer makers in the US except IBM, never blossomed the way the founders had hoped. Many for-profit incubators (often demanding up to a 50 percent equity fee from their tenants) found it difficult to generate the expected profit rates. Many small business development corporations (SBDCs), some of them operated under local government aegis, were never able to fulfill initial expectations.

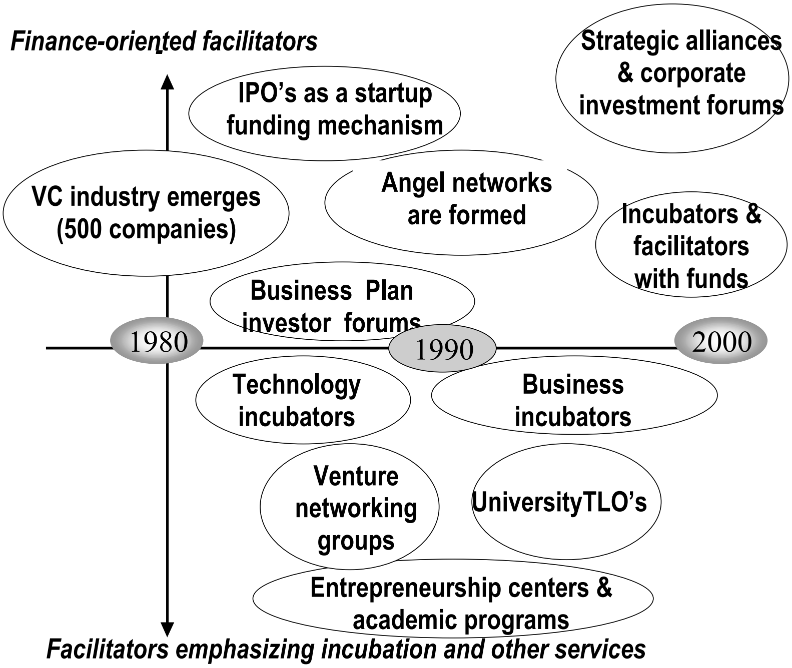


Fig. 2. Time lines illustrating the emergence of important categories of commercialization facilitators

Given the intense experimentation searching for new specialized facilitating forms and practices, it should be of no surprise that some of these attempts did not fare well. As in all entrepreneurship, some initiatives will succeed and others will be less favorably received in the marketplace. Failure is an essential part of institutional as well as industrial evolution.

Facilitators of the second and higher order. A commercialization facilitator of the *first* order promotes and mediates the commercialization of new technology. All the examples listed up to now are facilitators of the first order.

A facilitator of the *second* order promotes and contributes to the formation of facilitators of the first order. The IC^2 Institute of the University of Texas, to be described in some detail in Sect. 4 below, is a commercialization facilitator of the second order: it has created a number of other commercialization facilitators in Austin and elsewhere. IC^2 innovated one of the first and most successful business incubators, has helped start over a dozen other incubators in other parts of the world, created one of the first venture capital networks, and has promoted numerous other commercialization accelerating initiatives. It may be argued that the existence of this second order facilitator hastened the formation of the Austin-San Antonio corridor as a modern technopolis. See Fig. 3.

The European Business and Innovation Centers (BICs), set up by the EC Directorate-General for Regional Policy and the Directorate General for Enterprise, may also be viewed as 2$^{\text{nd}}$ order facilitators. Since 1984, over 150 BICs have

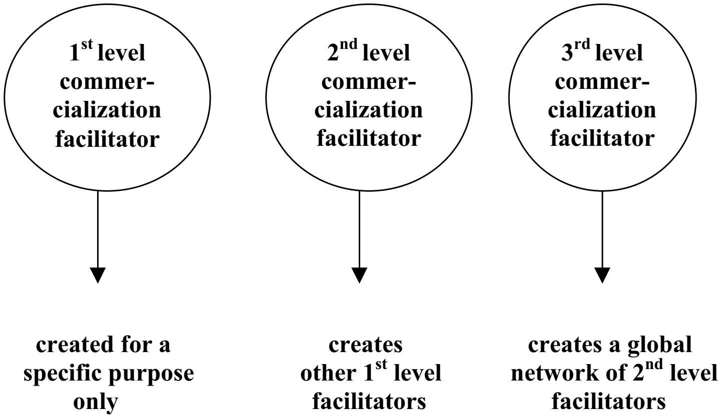


Fig. 3. A generic classification of commercialization facilitators

been set up in economically less favored regions of the EU. Each BIC features a partnership between operators in the local and regional economic development process, such as a local chamber of commerce, development agencies, university and research bodies, financial institutions, science and technology parks, as well as individual businesses. They contribute to the development of incubators and clusters.[3]

On the US scene, only a few 2nd order commercialization facilitators have ever been created. The IC^2 Institute is one. CONNECT of San Diego is another. A few others exist, but they are still relatively rare.[4]

Responding to market demand, the IC^2 Institute is currently engaged in a program to build "IC^2-like institutions" both domestically and abroad, thus possibly converting the Austin headquarters into a third order commercialization facilitator ... one that spawns other 2nd order facilitators and organizes them into a global network.

3 Toward a theory of commercialization facilitators

Commercialization mediation takes many forms, but the basic function is the same: promoting and facilitating the transfer of a new product or a new process technology from the laboratory to the market place. In order to describe this function, we shall imagine an abstract theory of the micro behavior of facilitators, just as there is a micro theory of the firm, of the financial institution, etc. What does this theory look like? We are entering un-chartered waters here. What follows are some elements of such theory – building blocks of an integrated theory yet to be written. It begins with the selection of research projects, extends through business startups, and concludes with the institutionalization of an organization as a going concern.

---

[3] For a recent study, see Wynarczyk and Wynarczyk (2002).

[4] Ronstadt et al. (2002) conducted a preliminary identification and analysis of 2nd order commercialization facilitators for five city regions in the United States.

3.1 Prioritization and selection of startup projects

Selection occurs, for example, when an incubator or other facilitating organization admits a prospective client startup company and takes it on as a tenant. Other applicants may be turned down. Selection also occurs when a large company decides to enter into a strategic alliance with a startup. Other prospective candidates may be passed over.

Facilitating organizations are always on the lookout for promising new candidates. The hallmark of the successful management of an incubator or a venture capital firm is the ability to spot early the winners of the future.

Quite generally, the selection problem involves (i) an identification of a number of prospective clients to be compared, (ii) the ranking of the performance and relative attractiveness of these various alternatives, and (iii) establishing criteria for acceptance or rejection.

The problem of selection and prioritization of R&D projects under a range of different institutional settings was recently dealt with in Thore (2002). A convenient tool for this purpose is *data envelopment analysis*, a mathematical technique for ranking projects or alternatives that are characterized by an entire vector of performance attributes (rather than a single figure of merit). The procedure can briefly be outlined as follows. Use the notation

$i = 1, 2, \ldots, n$ a list of R&D projects or prospective startup companies to be evaluated, ranked and prioritized;

$k = 1, 2, \ldots, s$ a list of inputs such as research hours and research costs of various categories;

$r = 1, 2, \ldots, t$ a list of performance criteria or outputs, such as expected sales volume, sales revenues, profit;

$X_{ik}, k = 1, 2, \ldots, s$ quantities of inputs used by project or startup company i;

$Y_{ir}, r = 1, 2, \ldots, t$ the amounts of outputs obtained by project or startup i.

Then the ranking of a given project, say project $i = 0$, is obtained by its so-called efficiency score θ_0, determined from the linear programming problem

$$\begin{aligned}
&\max \theta_0 \\
&\text{subject to} \\
&\textstyle\sum_i \lambda_i X_{ik} \leq \theta_0 X_{0k}, k = 1, 2, \ldots, s \\
&Y_{0r} - \textstyle\sum_i \lambda_i Y_{ir} \leq 0, r = 1, 2, \ldots, t \\
&\lambda_i \geq 0, i = 1, 2, \ldots, n
\end{aligned} \tag{1}$$

See further *ibid.*

Re-enumerating the projects $i = 1, 2, \ldots, n$ in the order of monotonically increasing efficiency scores, one has

$$\theta_1 \leq \theta_2 \leq \ldots \leq \theta_i \leq \ldots \leq \theta_n \tag{2}$$

Projects scoring below some low threshold value are slated for *rejection* or *termination*. Projects scoring above some high threshold are slated for immediate further development.

Ideally, prioritization and selection in this manner should occur intermittently during the life cycle of a project – when it is initiated, and as it moves along its path toward commercialization. All data and variables in program (1) are to be indexed by date. Accounting for uncertainty, they may even be indexed by several possible states of the world.[5]

Since there are many decision-points between project conception and project completion (successful commercialization, or termination), these decisions will combine retrospective assessments with forward-looking appraisals.

3.2 The concept of capture, and assisting startups to capture markets

The "law of capture" (Norton and Bass, 1992) describes the manner in which the latest technological generation of a product or a process takes over demand from earlier generations. As the sales of the startup venture expand, they slowly eat into the market for earlier product generations. Because of the exhaustion of diffusion of the earlier generations, they may continue to grow for some period of time, but will eventually start to decline. Geometrically, we may think of a series of sales cycles, where each new cycle eventually drives the sales of the earlier generation to approximately zero. The cyclical downturn of the preceding generation is the result of consumers substituting the new generation for the old one. The law of capture states that the manner in which the latest technological generation takes over demand from earlier generations is fundamentally the same across different product categories.

To accelerate the diffusion of a new technology is therefore also to accelerate the capture of existing markets. Commercialization facilitators can assist this process by identifying technological and marketing opportunities at hand.

A simple mathematical model of evolution by capture is presented below (we follow here the general evolutionary schema of Nicolis and Prigogine, 1977). We use the following notation

X_t = total sales of old generation of product or process at time t
Y_t = total sales of new generation of product or process by startup company at time t
A = saturation level of total sales (both new and old generation)

and assume that the growth of the two variables over time is determined by the two-coupled equations

$$X_{t+1} = aX_t(A - X_t - Y_t) \tag{3}$$

$$Y_{t+1} = bY_t(A - X_t - Y_t) \tag{4}$$

where a and b are known parameters. In other words, the two cycles are coupled together. The expansion of the old generation (Eq. 3) slows down as the sales of the new generation gradually build up; the expansion of the new generation (Eq. 4) is enhanced as it captures market share from the old one.

---

[5] *Ibid*, Chap.3 ("The Life Cycles of Sales and Profits: Dealing with the Uncertainties of the Commercialization Process").

As is well known, equations like (3)–(4) may lead to a smooth sequence of cycles, or to chaotic explosions.[6] The launching of a new product is not necessarily an orderly process, neither to the startup venture nor to the assisting facilitator. The new generation can be a wild success, or it can fizzle into nothing.

3.3 Valuation of the startup company, searching for suitable merger or buyout partners, or assisting the company with an IPO

The economic wealth created by the construction and installation of new production capital was by earlier economists referred to as "quasi-rent" (see e.g. Marshall (1890) and Schumpeter (1934)). Geometrically, quasi-rent would accrue as the area below a monotonically falling curve of the marginal product of capital. Similarly, the term quasi-rent may also be applied to the economic wealth created by the development and commercialization of new know-how and human capital – the surface below a monotonically descending curve of the marginal returns.

Eventually, the R&D costs of a successful new venture can be financed by its own quasi-rents. In addition, the quasi-rents of earlier generations of the new product or process may be drawn upon to cover the costs.

The market value of a new venture equals the present value of its future stream of quasi-rents. For a company not yet quoted on the stock exchange, this market value may seem purely hypothetical – until one day the company eyes the possibility of selling the entire venture, or a part of it, or to enter into some kind of merger or acquisition with a cash payout. Such a deal amounts to monetizing economic wealth already created. For a number of reasons, such a strategy may seem attractive to the current ownership – the owners may feel that their relative strength lies in R&D rather than the management of an ongoing company, they may hesitate bringing in new ownership and management interests, etc.

Commercialization facilitators will often want to encourage such developments. The mediating institution cares less about personal factors such as retaining management control to the initial startup group, and more about the timely equity financing of the new product or process. Seen from the point of view of the facilitator, it may be a definite advantage that ownership and management control of the new venture is partly or wholly transferred to an experienced company with its own management team.

The commercialization facilitator may therefore make it a point of assisting the startup company in searching for suitable merger or buyout partners. It is our experience that many startup owners welcome such opportunities to cash out with open arms. The facilitator may also draw upon its contacts to lay the groundwork for the flotation and public sale of new venture stock (an IPO on the stock exchange).

---

[6] The original work is Yorke and Li (1975). For a non-technical discussion of chaos theory, see Prigogine and Stengers (1984). The apparent randomness in many physical and social systems, and the relation between complexity and evolution in such systems is discussed in depth by Wolfram (2002).

3.4 Logistic growth of commercialization facilitators during a business cycle upswing

The growth of commercialization facilitators (measured, for instance, in terms of total dollar amounts of commercialization spending in a region) expresses the expectations and hopes of its tenants and clients: local startup entrepreneurs, launching new ventures. If successful, each one of those ventures will eventually embark upon its characteristic life cycle of growth. It is a cycle of an early upswing, and subsequent maturation as the product gains market share (and the eventual decline in competitors introduce new and more sophisticated products embodying more advanced technology). As is well known, the commercialization of new products is a "leading indicator" of the business cycle. In the same manner, the growth of the facilitators themselves presumably also is a leading indicator.

It is easy to find examples that document the crucial role of commercialization facilitators during the early upswing of the business cycle. One is close to our own experience: the startup of the Austin Technology Incubator in Austin, Texas, at the depth of the Texas economic downturn in 1989 and its subsequent role in electrifying a local economy that eventually was to become one of the fastest growth areas in computer and information technology in the nation. (See our discussion of ATI further below.)

For another example, of even wider scope, one may recollect the spectacular rise of the junk bond in the 1980s and the ascendancy (and subsequent infamy) of the investment-banking firm of Drexel Burnham Lambert. As it turned out, junk issues were critical to the US computer industry, providing between 1985 and 1989 some 80 % of all its finance. The crucial breakthrough for the fiber optics industry was MCI's fiber phone network. Drexel provided no less than $3 billion of capital. Seeing the advent of the "electronic superhighway" earlier than others, Michael Milken financed large building blocks of the electronic infrastructure that would become one of the major drivers of the US economy in the 1990s. Turner Broadcasting and TCI spearheaded the innovative drive in cable broadcasting, financed by multi-billion dollar junk issues. (See also Gilder, 1993 and Thore, 1995, Chapter 11.)

Commercialization facilitators provide synergy of stimulus to the local economy. The same principles apply here as in Joseph Schumpeter's well-known analysis of the clustering of innovations during the upswing of the business cycle. Not only do innovations cluster around the facilitators during the upswing - the facilitators themselves are powerful innovations, providing synergy to their tenants and clients. The synergy flows in several directions at once: from the local economy to a facilitating organization, from the facilitator to the local economy, and between its tenants and clients. Indeed, the totality of an organization and its present (and past) tenants and clients will ideally form a Schumpeterian growth cluster.

We now sketch a simple mathematical model of the growth of a commercialization facilitator (or a group of facilitators) and the snowballing effects transmitted back and forth between the facilitator and the local economy. This time, use the following notation

X_t = total sales of tenants or clients of the facilitator at time t
A = saturation level of sales of tenants or clients

Y_t = total sales of local economy at time t
B = saturation level of sales of local economy

and assume that the growth of the two variables over time is determined by the two-coupled equations

$$X_{t+1} = aX_t(A - X_t) + bY_t(B - Y_t) \tag{5}$$

$$Y_{t+1} = cX_t(A - X_t) + dY_t(B - Y_t) \tag{6}$$

where a,b,c and d are known parameters. In other words, rather than assuming that each sales statistic would follow its own separate cycle (in which case $b = c = 0$ in the two equations above), the two cycles are coupled together. The expansion of the facilitator over time depends not only on its own cycle but also on the local economy and its growth potential (the strength of the cross-effect is measured by the parameter b). Conversely, the growth of the local economy is stimulated by the growth of the facilitator (the impact is measured by the parameter c).

It is not necessary here to go into the details of the dynamic paths that can be generated by equations (5)–(6). The standard case is that the facilitator rides on the back of the growth curve of the local economy, benefiting from it and reinforcing it. There are also other possibilities. Systems of difference equations like (5)–(6) may produce chaotic behavior, just like the system (3)–(4) for a single product category.[7] Chaos will be associated with both explosive growth and collapse. Commercialization is always a risky business. Like a corporation, a facilitator may succeed beyond one's wildest dreams, or it may go bankrupt and have to fold.

4 A success story from Texas

We now turn to a striking example of the scope for creating synergy between entrepreneurs and facilitators that we have had a chance to observe first hand. The example actually involves synergy in a *triple* hierarchy:

- An academic institution devoted to the development of commercialization-facilitating initiatives, locally, nationally and internationally.
- A local business incubator and an affiliated venture capital network,
- Individual entrepreneurs and corporations commercializing new products and new technologies,

In this particular instance, the academic institution – an arm of the University of Texas – acted as the stimulus and primus motor of the ensuing development. Furthermore, this policy was quite deliberate and the goal of creating a Texas "technopolis" was spelled out clearly and at an early stage.[8] We describe events and institutional innovations below.[9]

---

[7] See Nicolis and Prigogine (1977) for an early account of multi-species diffusion systems like (5)–(6).

[8] For a quite prophetic statement, see Smilor et al. (1988).

[9] See also Porter (2001) and Ronstadt and Furino (2001).

4.1 The IC^2 institute

A powerful institutional innovation on the US scene occurred in 1977 with the founding of the IC^2 Institute (Innovation, Creativity and Capital), at the University of Texas at Austin. Focusing on technology, entrepreneurship and ideology, IC^2 gradually evolved into a hybrid organization that combined the research activities of a think tank with non-traditional education and outreach activities.

IC^2 was the brainchild of George Kozmetsky, who had co-founded Teledyne Inc., a company that grew into one of the largest conglomerates in the nation. In 1966 he accepted an invitation to become dean of the business school at the University of Texas at Austin. The IC^2 Institute developed a unique mission: to explore new organizational and management structures that can facilitate the commercialization of high technology. The philosophy behind the new creation was this: As the pace of commercialization of high technology accelerates, the nation's universities would have to assume a new role not only in the development of science and technology, but also in actually converting that technology into viable businesses[10].

4.2 Commercialization facilitating innovations: the IC^2 record

New institutions, just as corporations, need to respond to opportunities in the marketplace. As a result, their operations and their characteristics evolve over time. Certainly, this has been true of the commercialization-facilitating organizations and practices that are the subject of our present inquiry. In a fashion, this increasing diversification and specialization mirrors the general trend in the high tech economy toward an ever richer supply of "niches" of products to cater for ever more discerning consumers.

The following list of new commercialization-facilitating services is not exhaustive. It just catalogues some of the initiatives that grew out of the IC^2 operations over a 25 year time span that the present authors have witnessed first-hand.

- *Innovations of incubation operations:* Tenants admitted to the Austin Technology Incubator (see below) must be in possession of a workable and demonstrated technology or product. They are admitted for a period of three years, after which they "graduate," leaving the incubator premises. The network of past graduates ("exes") reaches out to new tenants. The tenants are offered the option to employ business school Ph.D. students as non-paid trainees (these trainee jobs are very much sought after). The incubator is available to the business school as a real-life laboratory, testing the commercialization process. Commercialization classes are taught in situ, on the incubator premises.[11]
- *Innovation of operations of venture capital firms:* The Texas Capital Network (operating under the IC^2 umbrella) created the idea of electronic matching of would-be entrepreneurs looking for venture funds and business "angels." The

---

[10] See Rogers (1998), Tanik (1999) and Zacks (2000).

[11] For a detailed discussion of ATI practices and various metrics of its success, see Wiggins and Gibson (2003).

initiative is currently being replicated in other states. The Texas Capital Network also has arranged a series of "venture fairs" designed to get entrepreneurs and venture capitalists together.

- *Innovations of academic activities:* The community of IC^2 fellows (currently more than 250 fellows) forms a worldwide resource of technology commercialization expertise. The fellows have interests that cross the traditional academic and professional boundaries. They come from business, government, academia, the media, and the non-profit world. They represent both the brain trust and the major implementing mechanism of the Institute. The fellows greatly extend IC^2's reach and capabilities as a "quasi-virtual" organization.
- *Innovations promoting emerging technopolis areas around the world.* Such missions operated by the IC^2 Institute are located in Curitiba, Brazil, in Guayaquil, Ecuador and in the Caribbean Basin. These efforts typically draw on the expertise of several institute fellows, and may include benchmarking, education, and incubation. Similar approaches are also being applied to transitional situations, such as along the Texas/Mexico border region, in Tblisi, Georgia, in Yerevan, Armenia, and in Medallin, Colombia.
- *Innovations of educational activities:* Drawing upon two decades of research into the nature of the commercialization process, the IC^2 Institute in the spring of launched its new Masters program in the Commercialization of Science and Technology. Originally taught in Austin and in Washington, D.C., the program has expanded into a multi-continental, multi-media classroom, with programs offered in Australia, Mexico, Poland, and Portugal. The same degree program is also offered on-line. The most recent item of this line of educational innovation is a Masters degree offered to employees of IBM, designed together with IBM management.

Overall, the genius of IC^2 has been to use Austin as its laboratory for learning how to effectively harness capitalism to innovation and creativity. In particular, IC^2 has worked closely with the city and county to nurture and grow a technology community and to build a sound basis for technology entrepreneurship and regional economic growth.

4.3 The incubator

The Austin Technology Incubator (ATI) was founded by the IC^2 Institute in 1989. It was formed by a coalition of university, government and business leaders as a three-year experiment to create wealth, generate jobs, and diversify the Austin economy, then in the grip of a severe recession. The experimental years were successful enough to be extended indefinitely and ATI became one of the model incubators in the nation.

ATI provided strategic, operational and infrastructure support to its tenants, including office space in a high-profile university locale. The office facilities included conference rooms, telecommunications, Internet access, receptionist, copy machines etc. The resident companies tapped into a variety of resources includ-

Table 1. Recent valuations of some successful ATI tenants

Year acquired	Company	Price
1999	Metrowerks	Purchased by Motorola for $95 million
1999–2000	Evity	Purchased for $100 million by BMC
1998–2001	Exterprise	$75 million to Commerce One
2001	DTM	Merged at $45 million valuation

ing in-house market research and public relations assistance, University of Texas faculty and students, and a network of professional service providers.[12]

Over the last dozen years, ATI has worked with approximately 110 high growth ventures. Of these, 65 graduated from ATI; 5 became public companies; another 13 were acquired or merged with other enterprises and several of these were acquired at high valuations. For instance, most recently:

To the best of our knowledge, only 10 of the 65 ventures can be classified as "failures."

The National Business Incubator Association named ATI "the incubator of the year" in 1994. Two years later, ATI won the Justin Morrill Award from the Technology Transfer Society. Several of its graduates have received similar distinctions. In 1996, Evolutionary Technologies, Inc. won national recognition as "Graduate of the Year" in Technology Start-ups. Applied Science Fiction and Infoglide were each recognized as "Client of the Year."

The achievements of ATI have prompted several mayors' councils nationwide to send representatives to Austin to learn more about how to generate business development and revitalize ailing municipal economies. ATI founded three additional incubators around the US and also the Austin Multimedia Incubator and the Clean Energy Incubator. ATI has established a network of international contacts by recruiting startup companies from Brazil, Israel, India, Canada, Australia and Japan.

4.4 Impact on the local economy

The operations of the IC^2 Institute, and the various initiatives operating under its umbrella, have resulted in significant impact on the local economy. In large measure, ATI served as a catalyst for Austin's economic recovery in the 1990s by developing an entrepreneurial support infrastructure.

Several measures compute the local economic impact. For instance, the average high tech wage in Texas is about $55,000 - approximately 75% higher than the average per capita job. ATI companies have created about 2,850 jobs yielding in year 2000 a total of $156,750,000 in wages. The 1% city tax on consumable spending would net the city $1,567,500 if a person spent all his or her income on consumable goods. Since only about 60% is so spent, the net to the city is approximately $940,500 this past year.

---

[12] See Wiggins and Gibson (2003).

The cumulative revenues of ATI companies have had a major impact on the city. By 2000, these revenues totaled \$1.25 billion. Since these sales stimulate other sales, the multiple impact is estimated at \$2.625 billion.

High tech jobs and sales are generally higher in value added than other sales. Assuming that the value added is 25% of sales, the total value added mediated or facilitated by ATI would equal approximately \$626 million. We want to stress the word "facilitated" because we are not claiming the ATI produced this value added, nor the other output measures. The companies themselves were primarily responsible for this output, although evidence exists that a number of companies would have likely failed or significantly underachieved without ATI's intervention prior to and during their residency at the incubator.[13]

While the success of ATI has been impressive, it is yet too early to evaluate whether this model can be generalized over either space or time.

5 The changing nature of capitalism

Joseph Schumpeter of course was the great protagonist of considering technology as an endogenous variable of the economic system He studied how new science and technology during the business upswing leads to waves of innovation and economic growth. To him, technology was the great engine of capitalism itself. In the same vein, financial economists later studied the endogenous evolution of new financial instruments and new forms of financial intermediation.

But up to now, the entrepreneurial function of the capitalist economy – the very driver of both industrial and financial innovation– has been taken mainly as given by economists, lying outside the scope of scientific inquiry. A common attitude by economists has been that entrepreneurship is an "art" that will always defy any attempts to categorize it.

In breaking this impasse, we believe that it is necessary to include the evolution of the entrepreneurial function as an endogenous variable of the economic system. To us, the basic function of the capitalist economy is entrepreneurship. The commercialization of new technology occurs in response to technological opportunities, market opportunities and financial opportunities. The high tech economy feeds on a rapid flow of new technology from the laboratory to the marketplace. Facilitators such as high tech incubators, spin-offs, and technology IPOs are new and evolving tools of the capitalist system, born by the system, directed by it, and directing it.

5.1 Schumpeter on entrepreneurship

As is well known, Schumpeter's was quite ambivalent as to the question of entrepreneurship. As a young man, he had acquired fame as the most eloquent interpreter of his time of the creative force of entrepreneurs and entrepreneurship (Schumpeter, 1934). Later, while still acknowledging that entrepreneurship is "a propelling force in the rationalization of human behavior" (Schumpeter, 1950,

---

[13] For the impact of the ATI on the local economy, see Harvard Business School, 1998.

p.125), he nevertheless took an extremely dim view on the future of entrepreneurship. In a section entitled *The Obsolescence of the Entrepreneurial Function*, he opined that

> "Technological progress is increasingly becoming the business of teams of trained specialists who turn out what is required and make it work in predictable ways . . . . Bureau and committee work tends to replace individual action . . . The leading man . . . is becoming just another office worker – and one who is not always difficult to replace"(*ibid*, p.133).

These lines were written in the early 1940s, during the war, when indeed much of the industrial might of the US was achieved through automated and bureaucratic organization - the manufacture of the Liberty ships, the Flying Fortresses and the Jeep.[14]

Today, in retrospect, economists are searching for reasons why Schumpeter's gloomy predictions failed completely to hit the mark. The revolutions in computing, information technology, and biotech that occurred in the 1980s and 1990s were all driven by extraordinary entrepreneurial vision, daring and risk-taking.

We believe that a major reason for Schumpeter's failure to appreciate the creative role of the entrepreneur in the modern economy was the defective understanding of competition and market formation that was the rule of his day. The entire body of received economic doctrine was based on the assumptions of "perfect competition." In 1936, Chamberlin and Robinson had added yet another market model: monopolistic competition. But the market form that was to rule in the developed economies in the late 20th century – the incessant evolution of high tech products with ever lengthening lists of desirable attributes – had not yet been discovered, neither theoretically nor empirically. To navigate a company in this storm, entrepreneurs with outstanding creative abilities are needed at the helm.

As an empirical fact, technology, production and market conditions in the Western world changed. Small scale, flexibility and customer proximity again and again led to superior performance in startups or small production units (Audretsch and Thurik, 2000). The new business opportunities could often be most suitably exploited by newly formed business organizations (Baldwin and Johnson, 1999; Acs and Audretsch, 1990). In particular, startup companies in the US computer industry enjoyed dramatically increasing returns to scale (Thore, 1995).

Briefly: Schumpeter's imagined world of hapless entrepreneurs never came into being. A new economy with new demands and new opportunities arose, requiring precisely the inventive entrepreneurial drive that Schumpeter had analyzed in brilliant terms in his youth but eventually thought would wither away.

---

[14] For a recent detailed study of Schumpeter's view on the role of the entrepreneur, see Brouwer (2002). According to Brouwer, Schumpeter's thoughts on the matter were not unique at the time, and that "the lack of interest in entrepreneurship in the second half of the past century can be attributed to the widespread idea that entrepreneurship would become more and more obsolete as capitalism developed" (*ibid.*, p.84). Brouwer also documents how Schumpeter's bleak outlook on the viability of capitalism can be traced back to his earlier thoughts on the matter as evident in the second edition of his *Theory of Economic Development*, in 1934.

5.2 Building an infrastructure of entrepreneurship

Returning now to our main theme and citing the rapid growth of various commercialization-promoting initiatives, we believe that entrepreneurship and commercialization can be permanently implanted in an economy as part of its institutional structure. This leads, irreversibly, to an economy where there is "more" entrepreneurship, both in terms of the number of entrepreneurs, number of companies, number of technologies commercialized, and sales of those technologies, and the number of commercialization facilitators. It also leads to a compression of the time lag between the development of new applied knowledge and its commercialization.[15]

This trend is all part of the arriving high tech society. Products and manufacturing processes are increasingly high tech; now, entrepreneurship is also becoming "high tech," drawing on sophisticated management and commercialization techniques.

This does not necessarily mean that entrepreneurship is becoming more "successful." Business failures and creative destruction are more than ever a part of the overall dynamic picture. Just as there is a rapid turnover of ever evolving technology, there is also a rapid turnover of firms. Successes and failures go hand in hand. Hicks (1991) calls this the "churn" of capitalism – high creation and high discontinuance rates. The stakes of entrepreneurship are getting higher. As a result, only the most insightful and advanced entrepreneurship survives and blossoms. Entrepreneurship itself evolves. And mediation and facilitation techniques evolve.

There are also attitudinal and political synergies at work, influencing the attitude towards entrepreneurship in society at large (Casson, 1995). There seems to evolve a growing appreciation of the opportunities of change (of the "American dream"). More individuals are involved in entrepreneurship, and the economic alienation of workers in the industrial age gives way to mobility, education, the cultivation of creative abilities, and founding of avalanches of new companies. Ours is rapidly becoming an "entrepreneurial society."

Hoping to emulate the successes of commercialization of high technology in the US and the spread of commercialization-promoting initiatives, many nations in the 21st century are embarking upon various programs of commercialization policy. Even former communist states these days place their bets on government-funded high tech incubators and the agglomeration of incubators into entire regional techno-cities.

5.3 Long term perspectives

Just as technology will continue to develop as long as man is creative, the development of new institutions and organizational forms of management designed to enhance the commercialization of new technology will also continue into the future. The present-day business incubator or IPO is just a step in a long chain of institutional developments yet to come. Looking into the future and into the next

---

[15] See Ramamoorthy (2000) on the so-called "Kozmetsky effect."

long-term economic upswing, we see the growth of a series of specialized institutions, many of them operating via the Internet, providing management advice, equity capital, and buyout opportunities for new startups.

Above all, we see future opportunities and need for a new breed of university-based institutions (commercialization facilitators of the second order) operating very much along the University of Texas model, benefiting regional development both nationally and internationally. The university has a unique impartiality and legacy of creative development that makes it natural to foster new kinds of institutions such as IC^2 that can show the way and act as a catalyst for the birth of new commercialization arrangements.[16]

References

Acs ZJ, Audretsch DB (1990) Innovation and small firms. MIT Press, Cambridge, Mass.

Audretsch DB, Thurik AR (2000) Capitalism and democracy in the 21st century: from the managed to the entrepreneurial economy. Journal of Evolutionary Economics, 10: 17–34

Baldwin JR, Johnson J (1999) Entry, innovation and firm growth. In: Acs ZJ (ed.) Are small firms important? Their role and impact, Kluwer, Dordrecht

Brouwer MT (2002) Weber, Schumpeter and Knight on entrepreneurship and economic development. Journal of Evolutionary Economics, 12: 83–105

Casson M (1995) Enterprise and competitiveness. Oxford University Press, Oxford

Gilder G (1993) America's best infrastructure program. The Wall Street Journal, March 2

Harvard Business School (1998) Austin, Texas: building a high-tech economy. Case 9-799-038, October

Hicks DA (1991) New dimensions of Dallas area economic development. The Survey of Regional Literature, 19, September: 11–24

Lundvall BA, Borrás S (1997) The globalising learning economy: implications for innovation policy. DG XII, Commission of the European Union

Marshall A (1890) Principles of economics. 8th ed.1946. Macmillan, London

MIT Media Laboratory (1999) Enabling technologies for learning and expression by people and machines. MIT, Cambridge, Mass. (March)

Nelson J (2002) Bringing institutions into evolutionary growth theory. Journal of Evolutionary Economics, 12: 17–28

Nicolis G, Prigogine I (1977) Self-organization in nonequilibrium systems: from dissipative structures to order through fluctuations. Wiley, New York

Norton JA, Bass FM (1992) Evolution of technological generations: the law of capture. Sloan Management Review, 33, 2: 66–77

Porter AW (2001) The knowledge seekers: how to turn your community into an engine for economic success. IC2 Institute, The University of Texas at Austin, Texas

Prigogine I, Stengers I (1984) Order out of chaos, Man's new dialogue with nature. Bantam Books, Toronto

Ramamoorthy CV (2000) A study of the service industry - functions, features and control. IEICE Transactions on Communications, E83-B(5): 885–902

Rogers RS (1998) Prickly research partnerships. C&EN (Sept. 21), 21–25

Ronstadt R, Furino A (2001) Creating IC2-like organizations: networking innovative organizations world-wide to design and launch new technologies, create jobs, and induce economic development. IC2 Institute, The University of Texas, Austin (working paper, October)

Ronstadt R, Furino A, Thore S (2002) Regional development strategies: creating a global network of commercialization accelerators. IC2 Institute, The University of Texas, Austin (working paper, August)

---

[16] For infant facilitators of the second order at other academic institutions, see Ronstadt and Paulin (1996), Ronstadt et al. (2002), 1996, and MIT Media Laboratory, 1999.

Ronstadt R, Paulin W (1996) Entrepreneurship facilitation for high technology businesses: forging a new theory. Pepperdine University, Malibu, Calif. (August)

Schumpeter J (1934) The theory of economic development: an inquiry into profits, capital, credit, interest and the business cycle, German original 1911, first English ed. 1934, 2^{nd} ed. 1978. Oxford University Press, London and New York

Schumpeter J (1950) Capitalism, socialism and democracy. 1^{st} ed. Harper and Brothers 1942, third edition, Harper Torchbooks 1950, New York

Smilor RW, Gibson DV, Kozmetsky G (1988) Creating the technopolis: high-technology development in Austin, Texas. Journal of Business Venturing, 4: 49–67

Tanik J (1999) University-based technology transfer: a state-of-the-art process based on evolutionary success. The IC2 Institute, University of Texas at Austin, Texas

Thore S (1995) The diversity, complexity, and evolution of high tech capitalism, Also published on the Internet, see http://www.ic2.org/pubs/sten.pdf. Kluwer Academic Publishers, Boston

Thore S (2002) Technology commercialization: DEA and related analytical methods for evaluating the use and implementation of technical innovation. Kluwer Academic Publishers, Boston

Wiggins J, Gibson DV (2003) Overview of US incubators and the case of the Austin Technology Incubator. International Journal of Entrepreneurship and Innovation Management, 3, 1/2: 56–66

Wolfram S (2002) A new kind of science. Wolfram Media

Wynarczyk P, Wynarczyk P (2002) Turning Schumpeterian theoretical insights into public policy action: fostering entrepreneurship and innovation through European Business and Innovation Centres (EuBICs) in less favoured regions of Europe. Paper presented at the 9th International Schumpeter Society Conference in Gainesville, Florida, March 2002

Yorke J, Li TY (1975) Period three implies chaos. American Mathematical Monthly, 82: 985–992

Zacks R (2000) University research scorecard. The Technology Review, 2000, July/August: 88–90

On the macroeconomic effects of establishing tradability in weak property rights*

Gunnar Eliasson[1] **and Clas Wihlborg**[2]

[1] Royal Technical University, 10044 Stockholm, Sweden
[2] Center for Law, Economics and Financial Institutions and Department of Finance, Copenhagen Business School, 2000 Frederiksberg, Denmark
(e-mail: cw.fi@cbs.dk)

Abstract. The New Economy is closely associated with computing & communications technology, notably the Internet. We discuss property rights to, and trade in, the difficult-to-define intangible assets increasingly dominating the New Economy, and the possibility of under-investment in these assets. For a realistic analysis we introduce a Schumpeterian market environment (the *experimentally organized economy*). *Weak property rights* prevail when the rights to *access, use,* and *trade in* intangible assets cannot be fully exercised. The trade-off between the benefits of open access on the Internet, and the incentive effects of strengthened property rights, depend both on the particular strategy a firm employs to secure property rights, and the protection offered by law. *Economic* property rights can be strengthened if the originator can find innovative ways to charge for the intangible assets. The extreme complexity of the New Economy and the large number of possible innovative private contract arrangements make it more important to facilitate the use and enforcement of private individualized contracts to protect intellectual property than to rely only on standard patent and copyright law. *Enabling law* is one proposed solution. Current patent legislation in the US has led to costly litigation processes weakening the position of small firms and individuals in patent disputes. The property rights of such firms and individuals could be strengthened with insurance or arbitration procedures.

Key words: Competence bloc theory – Enabling law – Experimentally organized economy – New economy – Weak property rights – Tradability – Underinvestment

JEL Classification: D21, D23, D52, D82, H54, K11, K22, K41, L11, L23, M13, O14, O33

---

* An earlier version of this paper was presented at the 9th Congress of the International Joseph A. Schumpeter Society (ISS), Gainesville, Florida, USA, March 28–30, 2002.
The paper is part of the *Nödfor* project on Schumpeterian Creative Destruction, notably the exit and bankruptcy process, based at the *Ratio Institute,* Stockholm.
Correspondence to: C. Wihlborg

I. Introduction

The existence of markets depends on the establishment of property rights. The reallocation of property rights through trade is central to the capacity of the economy to enhance economic welfare. Since at least Arrow (1962), we have had an academic debate on a presumed under-investment in knowledge and information assets with high social, relative to private, returns. Under-investment has been linked to a possible lack of incentives and badly-defined property rights (Kremer, 1997). Hence, the proper design of, and the support of, property rights should be a major concern of any central authority or Government. North and Thomas (1973) argue that not until a good institutional foundation (read legislation and conventions supporting property rights) had been laid in Western Europe did the industrial revolution begin, based on technology and knowledge that had accumulated during the previous centuries. Implicit in their argument was that nations that did not get their institutions right, could not benefit from the technology of the industrial revolution, despite a sufficient endowment of technology and production knowledge.

The evolving structures of the New Economy are, to an increasing extent, composed of digital abstractions of intangible assets with difficult-to-define property rights. Although the social value of these assets can be high, and may be enhanced by the Internet, the originators may not be able to charge a price for individuals' access to the assets. The Internet is a "double-edged sword": On the one hand, it enhances the social value of these assets, while on the other hand, it makes them increasingly accessible for potential users and, hence, difficult to charge for. The analysis of difficult-to-define property rights within endogenously-changing structures of intangible assets requires a dynamic, Schumpeterian-type model framework. To that end we introduce the Experimentally Organized Economy (Eliasson, 1992) populated by ignorant rather than fully, or marginally, uninformed actors.

The industrial policy debate has often focused on a presumed under-investment in knowledge creation due to a lack of incentives, notably in investments with a high social return. Incentives to do so are, in turn, closely related to the design and enforcement of property rights. Thus, for instance, investment rich in spillovers, but with a low private return, may not be made, even though the social return is very high, because the originator cannot appropriate the value of the spillovers privately (Eliasson, 2001b).

Property rights are more or less easily established depending on the type of assets. You can exercise your (property) right to an apple by holding it in your hand, and selling it to your neighbor who knows that it came from your garden. The (property) right to residential property and land is more complicated since it stretches over time. In the distant past this right was upheld by physical presence – living on the property and defending the lot with weapons. Today such property rights in civilized countries are enforced by elaborate registers and the legal system.

The difficulties of exercising property rights escalate with the degree of abstraction of the assets. The highly abstract rights in securities markets to the future profits from an investment commitment today took a long time to establish (Eliasson, 1993). In a very obvious way, the world political system was being overtaken during the last decade of the millennium by trading in financial abstractions (mathematical algorithms) representing property rights guaranteed by the legal systems of nations and, more importantly, by mutual trust between, and conventions among, financial institutions (Eliasson and Taymaz, 2002).

Perhaps the most difficult and perhaps also the most important property right for the New Economy is *intellectual property.* One aspect of the New Economy is the great share of asset values explained by intangible knowledge and information. A second aspect is the technology enabling transfers of information extremely cheaply and rapidly by means of digital abstractions on the Internet.

We argue in this paper that the capacity of nations to establish efficient institutions for property rights to intangible assets will be decisive for the ability of nations successfully to enter the New Economy.[1] Our focus is on the spectrum of more or less *weak property rights* between no right – and no tradability – and completely uncontested rights to intangible assets. Since none of the extremes exist, property rights cannot be regarded as either one or the other, as is conventional practice in the literature. Even when legal rights exist, the economic property rights can be contested and weak. The degree of contestability of *property rights* depends on law, precedent and convention, and especially on the private costs of protecting the rights by various means. Rights of entering contractual arrangements for the use of an intangible asset affect the contestability of property rights. We argue that the costs of a particular way of protecting property rights include lost opportunities as well as outright transaction costs. We question the efficacy of conventional legal approaches to protecting intellectual property rights for the important assets of the New Economy. Contractual arrangements should be supported, however, by enabling law.

In Section II, the concepts of the macroeconomic setting are described as in an "experimentally organized" economy, wherein "competence blocs" play an important role. The importance of intangible assets is emphasized. In Section III, we focus on the link between intellectual property and economic growth. Section IV reviews the literature on the efficiency of patent protection, and we argue that the "under-investment" literature does not capture important characteristics of intellectual property in the New Economy. The transactions costs approach to property rights is reviewed in Section V. We borrow the idea of redefining transaction costs to include the costs of business mistakes, which are typical of the experimentally organized economy, from the conference companion paper Eliasson – Eliasson (2002a), and derive the implications of this approach for property rights formation in the New Economy, and in competence blocs in particular. The implications of the new digital world and Internet technology for economic property rights and indirectly for economic growth are discussed in Section VI before concluding in Section VII.

---

[1] See also Eliasson, Johansson, and Taymaz 2002, on *Simulating the New Economy*

II. Intangible assets, their valuation and the experimentally organized economy

The valuation of, and trading in, intangible assets in an experimentally organized economy featuring frequent business mistakes are studied from a property rights and efficiency perspective, using competence bloc theory.

1. Valuing assets

It was, and still is, almost a dictum within the accounting profession that the measurement of intangible assets is a hopeless, arbitrary task. However, some three decades ago the economics profession was entrenched in an intellectual conflict called the capital controversy (see for instance Robinson, 1964; Solow, 1963). The winning side concluded that capital in general (physical or not) as a factor of production was theoretically *unmeasurable*, since its value could not be theoretically separated from its return. A more meaningful approach, however, would be to say: *yes* of course, but there are still methods of estimating capital or asset values that can serve as useful approximations in many contexts if one knows what one is doing. Our main argument, however, is that there is no principal difference between measuring physical and intangible capital, only a matter of degree in difficulty, and perhaps not even that.[2] For instance, financial derivates and similar financial instruments are highly intangible and abstract. They are constantly valued and traded in markets and are probably more concrete and definable than many physical capital items. Most of these instruments are not patented, although since the early 1970s new financial instruments ("mathematical algorithms") have been patentable in the US. This patentability was unambiguously established in 1998 when the new (since 1982) Court of Appeals for the Federal Circuit (CAFC) upheld a patent on a software system in a law suit between Signature Financial Group and State Street Bank.[3], [4]

For our future discussion, let us distinguish between hardware, intangible and entrepreneurial assets (capital)[5] as in Table 1 (see Eliasson, 2000a). Subtracting debt from the sum of the three capital items, we obtain an extended definition of net worth, as it is continuously valued and traded in the stock market.

The value of a firm is to a large extent embodied in people, or teams of people with tacit competencies representing "entrepreneurial competence". Entrepreneurial teams dominate the fate of large and small firms alike (Eliasson, 1990a), but the individuals are always free to leave the teams within the limits of their contractual obligations. This mobility creates a valuation problem, since the mobile entrepreneurial competence affects the value of most of the assets on the balance sheet of a firm.

---

[2] See Kingston (2002).

[3] The interesting thing is rather that the academics who had come up with the algorithms rarely patented them, signaling a great unawareness of the commercial opportunities of what they are doing in their research (see Lerner, 2000c).

[4] This decision explicitly rejected the notion that "business methods" were inherently non patentable.

[5] Entrepreneurial capital is intangible but also tacit and generally not codifiable.

Table 1. The complete balance sheet of a firm

+	(1)	Hard ware, financial/visible
+	(2)	Intangible
+	(3)	Entrepreneurial
–	(4)	Debt
−	(5)	Net worth $(= 1 + 2 + 3 - 4)$

Source: Eliasson, 2000a; Making Intangibles Visible. In Buiges et al. (2000), Table 3.3, p. 61

One problem with the mainstream neoclassical model is that capital is assumed to be computable by the conventional present value formula,[6] and no attention is being paid to the competence of actors in financial markets in pricing the assets they value. It is assumed that neither firm management nor stock market analysts and traders make mistakes that influence the long-term growth outcomes except for stochastic mistakes around the exogenous equilibrium trajectory. The producer, the innovator, the entrepreneur and the (venture) capitalist are all aggregated into one actor in the neoclassical model, and this one actor is assumed to be more or less predictable and calculable, barring a stochastic term.

We argue that the valuation of, and the capacity to, trade in claims on a firm depend on the actors in financial markets, who must *understand the role in production of the three categories of capital* (Eliasson, 1990a, 2003). For this we need an Austrian/Schumpeterian (1911) type model in which business mistakes become a normal, non-stochastic phenomenon and failure through exit a frequent consequence.

2. *The experimentally organized economy – the growth connection*

By far the most important prior assumption of economic analysis – often without prior comment – concerns the limited size and complexity of the models' state space. We prefer to call this space the investment *opportunities space*. This assumption on the size of the opportunity space is decisive for the state of information that can theoretically exist in the economy, and for understanding the existence and nature of markets and the dynamics of economic growth.

By expanding state space (by assumption) far beyond the limits allowed in the models of New Growth Theory and the models of asymmetrically informed markets of the Akerlof – Spence – Stiglitz type (ASS), it has been demonstrated that growth occurs through the four categories of *selection* or Schumpeterian creative destruction in Table 2 (see Eliasson, 1996a, 2001a). Among the four categories – entry, reorganization, rationalization and exit – the last one, involving the positive role of frequent business failure, will play a particularly important role in our growth analysis.

---

[6] Recent developments in corporate finance have undermined the discounted cash flow model by introducing "real options". In particular, "growth options" based on firms' intangible assets and entrepreneurial capital are viewed as a major share of firms' values in the New Economy.

Table 2. Schumpeterian creative destruction

1.	Innovative entry (through competition)
2.	Reorganization
3.	Rationalization
	or
4.	Exit (shut down)

Source: Företagens, institutionernas och marknadernas roll i Sverige, Appendix 6. In: A. Lindbeck (ed.) *Nya villkor för ekonomi och politik* (SOU 1993, p. 16) and Eliasson (1996a, p. 45)

Three circumstances of importance for our further discourse should now be observed. *First*, the creation and selection dynamics of the Schumpeterian creative destruction process of Table 2 does not automatically lead to growth. If competence and incentives are lacking, incumbent firms may contract and exit rather than invest and expand. *Second,* the Schumpeterian creative destruction process of Table 2 can be seen as a dynamic allocation scheme dominated by selection, i.e. creation through entry and failure through exit. The creation side must also be supported by competence, notably by the ability to create and to discover winners and carry them on through the competence bloc of Table 3B to industrial scale production. Losers must be driven out of business through competition and exit (item 4 in Table 2). *Third*, the smooth functioning of this creative destruction process, being supported by entrepreneurial and venture capital competence, requires the support of an appropriate infrastructure of institutions. This is where property rights enter to establish tradability in the knowledge or competence categories that rule the allocation process

3. Competence bloc theory – innovative creation and competitive selection

With economic dynamics and growth being dominated by experimental selection, the efficiency of that selection becomes important. One aspect of efficiency is defined by the minimization of the economic consequences of the two types of business mistakes of Table 3A; keeping losers on the budget for too long and losing the winners. Competence bloc theory explains the way in which this happens.

The competence bloc (Table 3B) lists the minimum number of actors/individuals with competence needed to create, to identify and to carry winners on to industrial scale production and distribution.

For such a successful and industrially dynamic outcome, the competence bloc has to be *vertically complete* and exhibit great *horizontal variety*, thereby reaching the *critical mass* needed to guarantee potential winners' increasing returns to the continued search for resources (see Eliasson and Eliasson, 1996; Eliasson, 2001a). A competence bloc typically includes several types of entrepreneurial capital. For didactic purposes, one can observe that a large firm may internalize most, or all the competence functions of the competence bloc in one hierarchy, thereby merging

Table 3A. The dominant selection problem

Error type I:	Losers kept too long
Error type II:	Winners rejected

Source: Eliasson and Eliasson 1996. The biotechnological competence bloc. Revue d'Economie Industrielle, 78-4^0, Trimestre

Table 3B. Actors in the competence bloc

1.	Competent and active *customers*
2.	*Innovators* who integrate technologies in new ways
3.	*Entrepreneurs* who identify profitable innovations
4.	*Competent venture capitalists* who recognize and finance the entrepreneurs
5.	*Exit markets* that facilitate ownership change
6.	*Industrialists* who take successful innovations to industrial scale production

Source: Eliasson and Eliasson 1996. The biotechnological competence bloc. Revue d'Economie Industrielle 78-4^0, Trimestre

the innovative, entrepreneurial, venture capital and industrialist functions into one aggregate, which is typical in neoclassical production analysis. IBM in fact did this in its heyday during the 1980s; it was even an advanced customer to itself (Eliasson, 1996a, pp. 175 ff). Alternatively, the competence bloc allows typical business functions to be distributed over the market where different types of competence are applied and allocated in a decentralized manner. It is argued in Eliasson and Eliasson (2002a) that decentralized competence blocs maximize the exposure of each project to a competent and varied evaluation and minimize the risk of losing a winner. The competence bloc solution has to be supported by economic property rights that allow trading in intangible competence capital over markets such that prices are correctly set for an efficient outcome. We return to this issue below.

III. Intellectual property and economic growth

Economic growth occurs through the innovative creation and competitive selection of projects at all levels of aggregation as projects are filtered and allocated through the competence bloc (Tables 3) and introduced in the economy as new entries of ideas, projects or entire firms. The process forces reorganization and change among incumbents, and it forces some actors to exit (see Table 2). Eliasson and Taymaz (2000, 2002) demonstrate in a micro to macro model framework that a healthy exit process is critical for sustainable growth.

The endogenous Schumpeterian creative destruction process could be more or less efficient in two ways. *First*, the creation and selection process in Table 2 can be more or less efficient in terms of minimizing the costs of the two business errors noted in Table 3A. In particular, losing winners as a result of deficient (vertically incomplete and/or horizontally narrow) competence blocs is costly. *Second*, even if the competence exists, incentives to innovate may be lacking due to a deficient property rights design and a weak capacity to reallocate industrially valuable knowledge or competence through trade. Both factors would contribute to an under-investment in innovative activity, even though the benchmark of what is best possible is realistically unclear in the experimentally organized economy, in contrast to being precisely and misleadingly well-defined in mainstream analysis.

A country's relative productivity depends on its capacity to absorb technology, observes Eaton and Kortum (1995). Eliasson (1990a) denotes this capacity *receiver competence*. Except for the US, continues Eaton and Kortum, OECD countries derive almost all of their productivity growth from abroad. However, they also argue that countries still earn most of their returns to innovation at home, while foreign countries are important sources of technology (Eaton and Kortum, 1994).

These references represent indirect evidence that (international) *trade in industrial knowledge* matters. The trade can take many forms, however, depending on the design and protection of property rights to knowledge. Since there is ample evidence that new technology stimulates growth, it is deficient tradability in knowledge that might lead to under-investment in innovative activity, less than efficient allocation of knowledge capital, and less growth than is potentially possible. Direct evidence in the economic literature on the growth effects of *legal* property rights systems is weak, however. Lanjouw, Pakes and Putnam (1996) point out that patent counts are imperfect measures of innovative output (read R&D). They then go on to adjust raw patent data for their value to the holder to derive a better measure of such output.[7] They also show that the economic value of the patent depends on the legal rules of patent protection. Aoki and Prusa (1995) observed that the Japanese patent system has allowed rival firms to look at the application. In the US the information is not available until the patent has been granted.[8] This implies, according to Aoki and Prusa, that Japanese rival firms have had an information advantage when they planned and coordinated their R&D efforts, which in turn should have led to smaller but more frequent quality improvements. Sakakibara and Branstetter (1999), however, found that the broadening of the scope of patent protection in Japan 1988 did not produce a significant positive change in innovative output in Japan.

The macro connection is invariably reached through an imposed equilibrium path, a method established innovatively in the Jorgenson and Griliches (1967) article, and returned to theory under the name of New Growth theory by *inter alia* Romer (1986). The positive macro growth effects in these models depend critically on strong positive spillovers in innovative production feeding into investment and output through conventional neoclassical production function analysis.

New Growth models predict that expansion in innovative outputs leads to a permanent increase in total factor productivity growth. This, however, may not be empirically correct, argues Porter and Stern (2000). Empirical evidence rather suggests "that most OECD economies have increased the size of their R&D workforce while experiencing (at best) constant total factor productivity growth rates". They go on to explain the role of "ideas production" in economic growth using patent

---

[7] Cf. The method used by Jorgenson and Griliches 1967, not quoted.

[8] Since November 2000 patent applications in the US are disclosed 18 months after having been filed (Johnson and Popp, 2001).

stocks to estimate the strength of spillovers from ideas–to–ideas, and find a small (but significant) effect of patent stocks on the level of total factor productivity, concluding that (op.cit., p. 27) "ideas-driven growth may be feasible. However, the size of such effects may be modest".[9]

Porter and Stern (2000) point to other factors such as "the national (or even regional) environment" that may be more important for understanding the dynamics of economic growth. These other dimensions are exactly what we emphasize, namely the dynamics of resource allocation on the micro level, notably of intellectual or competence capital embodied in human beings, systematically excluded by assumption in the macro new growth models. It is, however, explicitly present in the experimental economy/competence bloc approach that was presented above. The importance of tradability in intellectual production capital for efficient allocation of the same intellectual capital stands out. The importance of the micro-to-macro dynamics of such allocation (see Tables 2 and 3B) has been quantitatively demonstrated in Eliasson, Johansson, and Taymaz (2002). One insight from this paper is that the processes discussed in Porter and Stern (2000) are very drawn out in time, possibly explaining some of their weak econometric results based on considerably shorter time series data.

IV. Property rights and efficiency: the applicability of the literature on intellectual property in the new economy

There are two conflicting views on the role of patent rights in creating economic value. If patent rights, for instance to intangible assets, cannot be established, incentives to invest in such assets will be low and we have Arrow's (1962) *under-investment* problem. On the other hand, firms will invest resources in the protection of their property (legal or otherwise) and rival firms may engage in "wasteful" expenditures or reverse engineering to go around patents. Hence, there could be an *over-investment* problem (Hirschleifer, 1971).

Arrow (1962) concluded that competitive markets do not provide supporting incentives for innovative activity, but that patent rights, establishing temporary property (monopoly) rights to an intangible asset (a technology), may not be statically efficient. To make sure that innovative activities take place despite the lack of incentives, these activities should be conducted in publicly-funded laboratories making "innovations" freely available to everybody. The 1962 article has exercised considerable influence on the debate about patents and the organization of basic research, but Arrow (1962) did not incorporate a number of critical empirical circumstances in his analysis – the analysis was carried out in a zero transactions cost

---

[9] Kortum (1994), using a similar new growth theory model, observes that while R&D employment and TFP have both grown, the rate of productivity growth has remained flat. Eliasson, Johansson, and Taymaz (2002) observe, as one of their three paradoxes, that such small effects probably depend on the long gestation periods involved, effects being impossible to identify economically in short time series data. Instead E – J – T simulate the macroeconomic effects of ideas, or technology creation and diffusion on a micro-to-macro model of the Swedish economy using among other things genetic algorithms to model the learning and diffusion processes among firms.

environment in which all knowledge, once supported by legal rights, was treated as “tradable” information. In an economic theoretical environment where transactions costs are dominant and exceed 50 percent of all resource use in the economy (Eliasson, 1990b), the theoretical conclusions would be entirely different.

Arrow did not consider that the efficiency of innovative activity is critically dependent on its organization, and that the links between the academic laboratories and the industrial introduction of innovations are long, weak and costly.[10] His analysis has led to an overemphasis on academic research as a source of innovations. Basic technology development in academic or firm laboratories only draws a tiny fraction of the resources needed to take new technology to industrial scale production and distribution. Furthermore, technology is not merely information, and building a business on new technology requires considerable competence and resources. The consequence of Arrow’s proposal might even be that no innovations would reach industrial production. Hence, the main transactions cost associated with Arrow’s proposition has to be “lost winners”, a possibility excluded by assumption in Arrow’s analysis. Taken together, it is easy to reverse the conclusions of Arrow (1962) by one or two minor modifications of its underlying (empirical) assumptions.

Another serious objection to Arrow’s analysis is that it looks at (and this is unavoidable in static equilibrium analysis) the innovation as a well-defined optimum solution. In the experimentally organized economy (EOE), there are no well-defined innovations, and above all no determinate best innovations. Above all there is no fixed (fix point) reference for efficiency comparisons. Attempts to invent around patents in the EOE (called waste in the Arrow model) may be as innovative, and lead to as unique inventions as the original innovation. In the context of the EOE and Austrian/Schumpeterian analysis, waste becomes undefined and as much an act of learning and renewed innovation as the original invention. R&D subsidies would not improve upon the situation even though the social planner can now exercise his influence (Kremer, 1997). The evidence is overwhelming that subsidized industrial research, especially research carried out in special government-operated laboratories, is inefficient, does not turn out winners and leads to low social returns compared to private research which is rich in spillovers.[11]

There is another dimension to property rights, which is always disregarded in perfect competition analysis, namely the dynamics of the *allocation* of knowledge or new technology over the actors within the competence bloc. Private economic property rights confer *tradability* to the asset, which can now be sold to another owner who is more competent than the innovator to build a business on the technology. The stronger the property rights the more tradable the asset and the more dynamically efficient the allocation of knowledge or competence. Vice versa, if property rights are weakened, tradability is lowered and economic value destroyed.

---

[10] Under Arrow’s (1962) assumptions the business idea of Karo Bio (case in Eliasson – Eliasson, 2002a, this conference) would have no empirical foundation.

[11] (See Eliasson, 1996b, 1997a) In addition, big companies invested in large central corporate laboratories during the 1970s and 1980s, on the assumption that they would churn out new basic technology. The results were not positive (Eliasson and Granstrand, 1985) and firms, in contrast to Governments, have been fast to close down such facilities.

This aspect of tradability becomes particularly important when tacit knowledge is being considered. Tacit knowledge is embodied in human beings or groups of human beings and is typically reallocated through trade in the markets for executive competence and strategic acquisitions.

An alternative to the proposed research subsidies would be *patent buy-outs* (Kremer, 1997), especially when based on auction pricing. One advantage of *patent buy-outs* is that they can be more naturally placed in the environment of the EOE as long as research results can be codified. In order not to kill incentives, the buy-out should not be at a lower price than the private value. To elicit the private value, an *auction* could be used, and to avoid private rigging of the auction, sealed bids should be used. Kremer (1997) suggests that the Government should offer the private value (the auction price) plus a mark up to cover the social value and that the patent holder should always have the right to reject the offer. Since the social value of research is normally much higher than the private value, governments should offer much more to stimulate inventions with a high social value. Kremer (1997, p. 17) suggests that the latter is at least twice the estimated private value. If the inventor refuses to sell, for instance because of a strong information advantage, the system of patent buy-outs functions like the current patent system.[12] Shavell and van Ypersale (1999) in fact argue that "intellectual property rights do not possess a fundamental social advantage over reward systems".

It seems that the empirical literature on patent protection and efficiency comes out only hesitantly in favor of patents.[13] The reason appears to be a lingering negative attitude, or an ideological aversion, towards creating and protecting private (even though temporary) profits. Lerner (2000a) concludes from his 150 year survey of patent office practice that "Nations where information asymmetries between government officials and patentees are likely to be more prevalent – larger countries, wealthier economies, and those where international trade is more important – incorporate discretionary features into their patent systems more frequently" and "divide the responsibility for determining patentability between the patent officer and the courts when information problems are likely to be severe". "Wealthier countries", he continues (Lerner, 2000b) "are more likely to have patent systems", but "they are also likely to charge higher fees and limit patent protection". "The origin of a country's commercial law appears particularly important" in explaining how patent protection is decided, notably in terms of awarding privileges and providing for discriminatory provisions.

Lanjouw and Cockburn (2000) use the fact that much of the developing world has introduced patent protection for new substances developed by the pharmaceutical industry during the 1980s to investigate the incentive effects; and find "some, although limited, evidence" that more research has been allocated to products "specific to developing country markets". [14]

Hall, Jaffe, and Trajtenberg (2000) find that the stock market values of companies correlate strongly positively with their stock of patents, notably when they

---

[12] Kremer (1997) presents an interesting case from pharmaceutical industry.

[13] See Cohen, Nelson and Walsh (2000, p. 2) and also Kremer (1997).

[14] The results are based on "Indian Survey data, and interviews with industry".

use “weighted patent stocks[15]” as an explanatory variable. Many researchers have been surprised to observe that US manufacturing firms in most industries seem to rely more heavily on secrecy and lead time to recoup their R&D investments than on patents. Despite the increasingly “pro-patent” legal environment in the US since the beginning of the 1980s, patents as a means of appropriating R&D returns appear to have declined (see for instance Cohen, Nelson, and Walsh, 2000). Despite this reported decline in effectiveness of patent protection, an unprecedented surge in US patent applications occurred at the same time (Kortum and Lerner, 1997), notably in semiconductor technologies, where patent protection effectiveness has been reported to be particularly low. Hall and Ham (1999) conclude, from a study on the US semiconductor industry over the period 1980–1994, that we can understand this “patent paradox” if we give up thinking in terms of the simple “innovation” or “patent race” model, and instead reason in terms of complex “patent portfolios”, involving the many parties and many technologies and complex contracts needed both to design and produce modern complex products and to safeguard and appropriate their value. This resolves “the patent paradox”. US semiconductor manufacturing firms, indeed, patent aggressively, since the “pro-patent” legal environment was established in the US in 1982 – more to raise the positive signals to attract venture capital and to secure proprietary rights in niche product markets than to protect and be able to license particular technologies.

The pharmaceutical and biotech industries appears to be the exception everybody refers to where patent protection is needed (Kremer, 1997, pp. 46f.), because once the substance formula has been discovered, most of the innovation costs have been expended and replication is easy. Here, however, exceptions are found. Zucker, Darby, Brewer, and Peng (1995) observe that intellectual capital in biotech often rests on tacit hands-on-experience that cannot be commercialized as information to outsiders, except by knowledgeable people moving to a competitor. “Natural excludability” can be organized within closely-knit groups of collaborators who share the rents. There is an academic dilemma, however. If the important hands-on-experience cannot be communicated in coded form, academic control through repeat experiments based on published material becomes impossible. Thus, academic peers will have to do with checking the result (“the substance”) without understanding *how* the research team got it out of the test tubes. Alternatively, if the process can be completely and exactly coded for publication, it must be the case that the biotech industry and academics have come up with a rule system that allows the researcher to withhold temporarily critical process information from publication.

Protection of knowledge is more difficult to handle when the knowledge derives from the contractual cohesion of teams that make up the technological and entrepreneurial competence that is decisive for the market value of the firm. If the team breaks up, the value collapses. Darby, Liu, and Zucker (1999) use an option-pricing based technique to value the intangible assets defined by the ties to star scientists in biotech firms. They then compare values with the market valuation of the same firms. The underlying hypothesis is that the more ties among stars,

---

[15] “References” or “citations” in patents identify earlier inventions whose claims are close to the citing patent. “Citations received” are often used to measure the “generality” of the patent (Jaffe , Fogarty, and Banks, 1997).

the greater the probability that a firm makes a commercially valuable R&D breakthrough. This hypothesis is supported empirically. In fact, the value of a firm, they estimate, increases with 7.3 percent or 16 millions in 1984 dollars with one article written by a participating star scientist.

On the whole, the empirical results imply that when patent protection of property can be well defined it is effective and not very costly to obtain and to enforce. Firms then patent, and are significantly more protected from imitation than they would otherwise be. This situation, however, is not the general one and it changes radically when we consider intellectual property, notably entrepreneurial assets (see Table 1) and the technologies of the New Economy. The variability and complexity of the innovative technology to be protected increase dramatically in the New Economy, as do the possibilities to protect through innovative designs, contracts and organizations going beyond standard patent and copyright law.

V. Economic and legal property rights in economic dynamics

In this section, we review the transactions costs approach to property rights and apply it to competence blocs. The costs of business mistakes listed in Table 3A must be considered when assessing the efficiency of a system of property rights. We also argue that, in the New Economy, a legal system using standard, mandatory contracts for various contractual relations is not the most efficient way to protect the extremely varied assortment of entrepreneurial assets. Instead, the decentralized competence bloc must be based on the variety of contractual arrangements that the experimentally organized economy demands.

1. The transactions costs approach to property rights

To go further in the analysis of property rights to the assets of the New Economy, we follow Barzel (1997) and distinguish between *economic* and *legal* property rights. *Economic property rights* correspond to the ability of an agent to capture the present value of the cash flows the asset is expected to generate, while legal rights are those specified in law. Agents can be expected to maximize the value of economic rights by *appropriating cash flows net of transaction costs* one way or the other. Transactions costs can take a variety of forms associated with the protection of cash flows including direct costs, and the potential opportunity costs of the inability to enter into contracts and to initiate activities that could generate cash flows. Appropriation of cash flows, and thereby the creation of economic property rights, take many different forms, ranging from selling the asset, selling products incorporating the asset's services, selling products and services linked to the asset such as TV and Internet advertising, to the merging of firms controlling different assets that can be combined to create conditions for the appropriation of cash flows.

Imperfect or weak economic property rights imply that part of the potential cash flows from an asset are in the "public domain", in Barzel's words. Other agents then have an incentive to appropriate cash flows generated by the asset.

Transactions costs arise in this competition for cash flows in the public domain, as well as in attempts to retain appropriability.

The extent to which economic property rights can be appropriated and thereby remain in the "private domain" of the individual generating or holding an asset depends naturally on legal rights to property but also on the ability of the asset holder to utilize the resource as he sees fit. Restrictions on the rights to manage the resource in various activities affect the division of the economic value of the resource into the private and public domains. A great variety of legal rights and obligations affects this division.

If economic property rights were perfect, so that all cash flows potentially generated by an asset could be appropriated and placed in the private domain at zero transactions cost, then private incentives and social objectives would be aligned. The value of externalities would then be appropriated as well. This proposition follows from the Coase theorem. If there are costs associated with appropriation, then the economic policy problem is more complex. Benefits and costs associated with the creation of economic property rights to any asset with a social value must include the opportunity costs of not creating such rights. A common example is that it would be very costly to define rights to the air we breathe, while if we do not, then non-smokers face costs of drifting smoke, or vice versa. In a world with transaction costs to establish economic property rights, there is no clear reference point for economic policy but rather there is always a trade off between different kinds of transactions' costs and benefits. It is particularly important to consider that any particular way of establishing economic property rights implies that there will be missed opportunities that are hard to identify.

As noted, the appropriation of knowledge- and information assets is particularly costly for well-known reasons. Several types of costs are associated with appropriation or lack thereof. For example, the value of information cannot often be extracted in the market for the information per se, because a potential buyer cannot assess the value without obtaining the information. Another type of cost exists because the value of knowledge is often enhanced by so-called network effects creating scale economies and potential "natural monopolies". On the other hand, anti-trust legislation would then have opportunity costs in terms of lost economies of scale. Also, the value of a particular kind of knowledge depends on what other types of knowledge it can "join" in the production process for goods and services, and on the available expertise with an ability to utilize the knowledge.

An advantage of a competence bloc is that the varied competences within the bloc increase the likelihood that someone will be able to understand the potential economic value of a particular kind of knowledge or information. The bloc is also more likely to contain the entrepreneurial capital required to organize economic activities in such a way that the private domain of the value of the asset is maximized. Hence, in a complete and varied competence bloc, the risk of losing winners is "minimized" and losers are more likely to be weeded out rapidly.

The theory of the Experimentally Organized Economy (EOE) makes the economic value of business mistakes explicit. The mainstream economic model, focusing on the net present values of assets under the assumption that the opportunity sets for uses of assets are known, cannot account for such mistakes. In the EOE, the opportunity set is itself determined by, for example, the system of property rights and the organization of activities in hierarchies or in decentralized markets. The frequency and magnitude of business mistakes within a certain property rights system for asset should be recognized as a cost of that system. Eliasson and Eliasson (2002 a,b) argue that business mistakes are more costly in a hierarchically organized competence bloc than within a decentralized competence bloc where there are many independent, experimenting entrepreneurs controlling different assets. Also, learning from business mistakes is more rapid in a decentralized bloc. An additional efficiency issue to be discussed is how economic property rights to intangible assets are protected in competence blocs.

2. *Enhancing economic value through the legal system*

Economic rights to intellectual property can be enhanced either by the strengthening of legal property rights or by entrepreneurs and firms organizing their activities in such a way that a minimum of cash flows potentially generated by an asset goes into the public domain. The most well-known legal property rights are created by patent, copyright, and trade mark legislation. The strength of the rights created by this legislation depends on enforcement by the courts and also on the information contained in the creation of patents. Once a firm has applied for patent protection, the application is in the public domain. Competitors then may obtain potentially valuable information that may enable them to develop close substitutes or give them time strategically to reorganize their business. For this reason, certain types of knowledge and technology are not patented.

The empirical evidence reviewed in the previous section indicates that substantial benefits of patent protection are limited to specific industries: in particular, to those using knowledge to produce output with designs that can be easily imitated after observation of the product or service. Even such knowledge can be protected to some extent by the original producer's first comer advantage, and the protection can be prolonged by brand name reputation. Most knowledge and information being used as an input in the production of goods and services is not that easily accessible. Exclusivity in the supply of know-how may exist, because the know-how itself is not sufficient to put it to use. It is understood that observable product know-how must be combined with some unobservable knowledge or any privately held asset.

It is by no means obvious that strong legal property rights are superior to private methods for enhancing economic rights[16] to intangible intellectual property, even if codified. There are transactions costs associated with the enforcement of legal rights in the form of direct costs of enforcement, and a loss of value when a patent application is made public. Costs of monopoly power created by patent or copyright protection must also be considered. An important transactions cost of enforceable legal property rights can be the disincentive for potential competitors to invest in the development of competing technology. Because of this, the under-investment

---

[16] See Davis (2002a) and Jonasson (2001, 2002) on "innovative pricing" for a discussion of corporate strategies aimed at appropriating cash flows. Reichman (2001) discusses "Repackaging Rights" as a way of establishing economic property rights."

problem can actually be made more severe. Furthermore, if trade in the patented knowledge cannot occur, the monopoly owner of knowledge assets may not employ it as productively as a competitor would have.

Private contractual arrangements can contribute substantially to the creation of economic property rights even if legal property rights are not explicitly, or are only weakly, protected in law. The extent to which an intangible asset's value lies in the private versus public domain depends on the ability of the asset holder to retain exclusive control over cash flow generated by the asset. Contractual agreements between an asset holder and a buyer of the asset (or the services provided by it) can be used to appropriate the value of the asset when the information contained in the asset can be kept exclusive. For example, the originator of a particular kind of know-how can supply this know-how to someone else while specifying restrictions on the use and resale of the know-how. It is quite common with patented knowledge that license agreements are made specifying restrictions on the licensee's use of the knowledge, but contracts can be entered into without patents. The role of the formal patent in such cases is to provide proof of origin of a technology and the exact delineation of the property being licensed. The existence of the patent may reduce the complexity of the contract and serve a function similar to the registration of real estate. The advantage of the patent in this case must be weighed against the disadvantages mentioned above.

There are two factors limiting the contractual arrangement for the transfer of information and codified knowledge. One factor is that the value of the information may not be easily assessed by a potential buyer without it being revealed. The second factor is that the enforcement of a contract requires that leakage of information to a third party can be traced to its source. We argue in the next section that Internet technology weakens both arguments against private contracting for securing economic property rights to intangible assets.

3. Competence blocs and trade in intangibles

Returning to the benefits of competence blocs, it can be argued that a complete and varied bloc, with a variety of generators and users of intangible assets, raises the likelihood that winners of any particular kind generated within the bloc will be identified. If the intangible assets can be traded, then the competence bloc can be organized as a group of firms with different entrepreneurial capital, while if the asset cannot be traded, the competence bloc is more likely to be organized as a large firm with many kinds of entrepreneurial capital "under one hat". Within this firm, the trade in intangibles is "internalized". Most likely the selection of projects will now be more narrow (less innovative) and the risk of losing winners larger as noted above.

When it comes to more complex intangible assets, such as entrepreneurial know-how or other types of capital that are not only intangible but also tacit, and by definition not codified, contractual arrangements become even more important for the appropriation of economic value. In this case, the tacit, intangible, knowledge asset is typically embodied in a person or a group of persons who can appropriate cash flows only through contractual agreements with financiers, employees, and other potential stakeholders in a firm. The greater the contractual freedom of the entrepreneur, the greater is the possibility that the economic value of the know-how can be appropriated. Restrictions on the contractual agreement between an entrepreneur and suppliers of financing in the form of mandatory standard form formulations of contracts in law can reduce the value of the entrepreneurial capital. In other words, it is desirable that law be enabling with respect to the contractual arrangements among stakeholders (Wihlborg, 1998a,b).[17]

Competence blocs in the experimentally organized economy make it possible to combine two or more kinds of tacit, intangible entrepreneurial competences to create new ventures. The most productive form of such ventures may be joint ventures, strategic alliances or outright mergers of firms (Eliasson and Eliasson, 2002a). The greater the potential variety of possible contractual arrangements between those possessing the entrepreneurial assets, the greater the likelihood that the most productive forms of cooperation can be found through experimentation. Mandatory standard form contracts for alliances and joint ventures may reduce the ability of one or both entrepreneurs to appropriate values. Standard form contracts that are enabling in the sense that parties can deviate from the standard form by mutual agreement reduce the likelihood that winning combinations of knowledge will not become reality.

VI. Economic property rights and the social value of intellectual property on the internet

In this section, we discuss the manner in which Internet technology affects the costs associated with legal property rights and their enforcement. Thereafter, we turn to the protection of economic property rights on the Internet. Digital products and production in the New Economy are making the property rights to difficult to define intangible assets increasingly important. The Internet is a key technology shifting economic attention towards intangibles and business concerns towards strengthening weak property rights.

The private creator of a positive externality wants to be able to manage, to earn a profit from and to trade in the values he has created, i.e. to claim property rights to them. If the property right cannot be naturally asserted, for instance by holding the apple in your hand, there are two ways to achieve the desired end: (1) legal protection and (2) innovative contractual, organizational or strategic arrangements as discussed by Davis (2002a) and denoted innovative pricing (IP) by Jonasson, (2002). IP is a way to establish *weak property rights* by means of, for example, innovative product and marketing strategy, and is one method of internalizing an externality. Obviously, if IP can be made effective, the existence of external receiver competence raises the value of the spillover to the originator. It may even be in the interest of the latter to invest in raising receiver competence among customers to raise the value of its product to the customer, to be able to charge a higher price (see Eliasson, 2001b).

---

[17] See Kingston (2002) for a practical contribution to valuation of intellectual property. Valuation with some degree of precision is one aspect of tradeability.

1. Enforcement and litigation costs

The US legal system has expanded its recognition of legal property rights (Davis, 2002a), but Europe has not. Thus European firm must rely on IP to a greater extent than must American firms selling in Europe. To the extent European firms do not enter the US market, they could otherwise become free riders on concepts and ideas developed in the US. More likely the costs of expanding legal rights to concepts and ideas will create enormous costs of litigation, since ideas and concepts are often "in the air" and being formulated in many different places more or less simultaneously. Creating exclusive rights to one among many idea-originators may create substantial costs in terms of restrictions on the use of the ideas.

In assessing the role of legal intellectual property rights, Lanjouw and Lerner (1997), LL below, emphasize the costs of enforcement, which appear to be large and increasing in the US in particular. Litigation is a part of, and a significant part of, the expected transactions costs associated with establishing a tradable property right. Enforcement of legal property rights and the associated expected litigation costs are particularly important when it comes to establishing tradable intellectual property rights. In fact, new financial products addressing patent litigation costs have emerged in the market. It is possible to invest in part-ownership in patents solely for the purpose of litigating them. A patent enforcement insurance market also appears to be emerging (Hofman, 1995).

The empirical evidence presented by LL for the USA indicates that the broader the patent the more asymmetric is the information between patent holder and patent infringer, and the higher is the probability of litigation. The broader the patent the greater is also the stake of the plaintiff and the more valuable is the patent and the firm holding it. Furthermore, the more revolutionary the patent, the broader is often the patent coverage and the larger is the number of patent citations. The number of patent citations is "strongly" correlated with the probability of an infringement suit". Lanjouw and Schankerman (1997) furthermore find that more valuable patents, notably those in new technology areas, are "more likely to be involved in litigation".

There is also strong evidence that large firms use their financial leverage to take small firms to court (LL, 1997) and that larger and financially strong firms predate on less financially healthy firms through patent litigation (LL, 1996). Since small firms are significantly more innovative than large firms (see Acs and Audretsch, 1988; Eliasson and Eliasson, 2002a), and small firms seem to be the creative origin of the revolutionary new technology moving the New Economy, this bias in transactions costs may hold back the emergence of a new economy.

An interesting observation in this context is that the probability of being involved in a patent suit is much higher for valuable patents and/or for patents owned by individuals and smaller firms. Patentees with large portfolios of patents to trade, on the other hand, encourage "cooperative" interaction and avoid court action more successfully (Lanjouw and Schankerman, 2001). The bad side, again, is that the smaller the firm holding the patent, the more vulnerable it is. This encourages the large firms to predate on small firms to acquire patents cheaply. Thus, efficient incentives to innovate and to allocate intellectual assets in a property rights system with wide scope as in the US require that small firms and individuals be able to protect themselves against the possibility that they will face large litigation costs.

Kingston (2000) argues that insurance arrangements for the protection of the intellectual property of small firms are not likely to work for reasons of moral hazard. He favors instead arbitration procedures, since disputes are intrinsically technical in nature. Furthermore, the arbitration procedures should be compulsory because voluntary procedures would not be accepted by financially, relatively strong parties. Were this the case, the litigation costs associated with the legal property rights system could be reduced.

A legal system with more narrow legal protection of intellectual property rights would have to rely on contractual, organizational, or strategic approaches (IP) to secure protection of economic property rights. To the extent contractual approaches are used, legal enforcement costs would exist but to a lesser extent if contracts were specifically designed for the parties, and possibly combined with organizational and strategic innovation. Such innovations carry costs themselves, however. An issue is how these costs are affected by the new technology.

2. *Realizing the potential value of the Internet through economic property rights*

In 1837, Daguerre offered to sell his photography process to a single buyer for 200.000 francs or to 100 to 400 subscribers for 1000 francs each (see Kremer, 1997, pp. 11 ff.) Nobody was willing to buy, either because they did not understand the potential of photography and/or found the offer too expensive. Somebody close to the French Government, however, must have understood the economic potential of photography and convinced the French Government to purchase the patent from Daguerre in 1937 and to make the patent freely available to the world for a lifetime pension offer,[18] considerably larger than what Daguerre had tried to elicit from the private market. Privately this patent buy-out may have looked generous. Socially, however, the deal was a winner for the world. In terms of economic insight, it compares extremely favorably and fairly with the deal offered Swedish inventor Håkan Lans by the UN and the large IT-companies in 2001 for his patent on a GPS-based Global Positioning & Communications (GP&C) system. To become a global standard, Håkan Lans had to turn over his patent to the world for free.

These examples illustrate the potentially enormous social values that may flow from intellectual property. The Internet is itself an intellectual innovation with potential social value of magnitudes that cannot be imagined.[19] To realize the potential value, conditions for establishing economic property rights to information on the Internet must be understood and clarified.

---

[18] Of 6000 francs per year, corresponding to some $1.8 million in 1988.

[19] Timothy J. Berners-Lee passed up great wealth when he – in 1990 – decided not to patent the technology used to create the WWW. A similar problem is coming up with the project on a semantic web in which Berners – Lee is also involved. IBM and Microsoft have proposed installing toll booths on the information highway to allow patented software to be used. Berners-Lee is against such arrangements as are the *Free Software Foundation* and the *Open Source Initiative* (Business Week, March 4, 2002, pp. 83–87).

The open access of the Internet is the cause of an enormous increase in the general accessibility of information. Much information on the Internet is literally in the public domain by a simple click. From the point of view of property rights, this implies that costly arrangements must be introduced to establish exclusivity of information and thereby to place it in the private domain. Alternatively, appropriation of the economic value of information on the Internet can take place by indirect means –.for example, advertising on a web-site with valuable information, or enhancement of the value of a product supplied separately.

The technology related to the Internet is rapidly developing with potentially important consequences for economic property rights. First, electronic contracting can be expected to become widespread, secure, and enforceable in the near future with the support of courts' acceptance of electronic signatures (Hultmark, 1999). Secondly, all activities on the Internet leave an imprint which enables the flow of information to be traced. There are also ways of hiding or obscuring the source of information flows but technology is moving in the direction of greater ability to trace flows from computer to computer.[20] This characteristic of the Internet is already seen as a threat to privacy in many dimensions. The third characteristic of the Internet with potential consequences for property rights is that very tiny pieces of digital information can be identified and potentially transferred exclusively. In combination with electronic contracting and the ability to trace information flows, it should be possible to define economic property rights to very small pieces of information. The limitation to the appropriation of the economic value of tiny pieces of information lies not in the size or magnitude of the information itself, but in the cost of enforcement relative to the economic value of the information. Proposals exist on the Internet for organizing the information flows through a database in such a way that enforcement through tracing can be made very cheap. The database would contain many individuals' small and large pieces of information, and it would enable members of the database to trade in "information assets". Contracts restricting the use of information on the database could be enforced by the organizer of the database.[21]

If economic property rights to tiny pieces of information can be realized, the consequences for both economic activity and privacy can be great. For example, the tracing of flows can now be used by firms to identify preference patterns of consumers. If rights to tiny pieces of information can be enforced, then individuals could claim the economic rights to information about their preferences with respect to products and services, and their specification. If so, they could contract with providers of Internet services that information about their use of the Internet must not be sold. Instead, individuals' preferences would be available at a price, and even very small payments could be made nearly costless on the Internet. On the production side, the ability to trade in tiny pieces of information could have important consequences. Economic rights to a greater range of intangible assets could be defined. Tacit, entrepreneurial capital could be devoted to combine the pieces

---

[20] Windows 2000. Also see Jonasson (2001, ch. 4).

[21] See www.preference.tv.

in new productive ventures. The virtual entrepreneur and the virtual competence blocs may become reality.

VII. Conclusions on the role of patent and copyright protection in economic growth

Most or all analyses of patent and copyright protection have been carried out within the constraints of the mainstream imperfect information/asymmetric information model of I/O analysis. We believe that the intellectual constraints of that model bias both the theoretical and the empirical results. We have, therefore, based our analysis on the assumptions of the experimentally organized economy (EOE) in which actors are not marginally uninformed, but rather are grossly ignorant about circumstances critical to their long run survival, and constantly make more or less serious, often fatal business mistakes. Since this model has no fixed (external equilibrium) reference for efficiency or opportunity cost measurements, the definition of transactions costs must consider opportunities for experimentation. To come up with any firm conclusions on the role of patent and copyright protection in supporting the introduction of a possible New Economy we have, therefore, approached the problem in steps.

First, we have introduced the concept of weak property rights and imperfect tradability of (intangible) assets as a normal phenomenon in the Experimentally Organized Economy (Sect. II) and then (*second)* linked that tradability to the efficiency in allocating the same assets to generate economic growth (Sect. III). *Third,* we have assessed the role of patent protection in general in promoting economic growth. This is where the bulk of the literature is to be found (Sect. IV). *Fourth,* we have looked at the particular assets (intellectual property) associated with the Internet and other technologies of the New Economy (Sect. V).

With this, we have addressed the difficult and specific problem of less than perfect (weak) property rights and the less than perfect tradability in the assets that move growth in the New Economy, an aspect of reality that is incompatible with the mainstream model and, therefore, almost completely missing in the literature. The analysis has been conducted in terms of the theory of the Experimentally Organized Economy (EOE) and we are now ready to present the two main conclusions of this essay.

In the EOE, costs of business mistakes, notably lost winners, are potentially very large. This analytical outcome of the theory is consistent with a true endogenization of growth. We apply the theory of competence blocs within the EOE to understand how the costs of such business mistakes can be "minimized". As a consequence of this redefinition of the benchmark for opportunity cost "measurements", the implications of tradability of intangible production capital for dynamically efficient resource allocation change radically relative to the mainstream model. The direct transactions costs of protecting and handling innovations within a hierarchy are relatively low. The implicit costs of business mistakes, on the other hand, are large. The incidence of business mistakes increases when competence bloc selection is internalized within one hierarchy and opportunity costs escalate when the economic value of the loss of winners (loss in output) are taken into account. A decentralized market based competence bloc increases observable transactions costs compared to those in a narrowly controlled hierarchy (Eliasson-Eliasson, 2002a), but reduces the losses of winners. Since the opportunity cost of business mistakes are not recognized in the mainstream imperfect information/asymmetric information model, except as a minor stochastic error term, comparison of the two theoretical models produce radically different theoretical conclusions, the mainstream model favoring central planning by assumption.

Our analysis of the dynamic efficiency of allocation, on the other hand, favors a decentralized, market based system. The efficiency of that decentralized allocation depends on the possibilities of establishing tradability in intangible assets, such that a smooth and competent selection through the competence bloc can be achieved. This first conclusion implies that economic property rights are important, but it does not say much about legal property rights.

Practically all empirical studies on patent and copyright protection have been intellectually formed within the imperfect market/asymmetric information model and have neglected this tradability aspect. A survey of the literature is rather inconclusive with respect to the traditional legal patent protection, indicating that the complexity of reality requires more innovative and varied designs to protect the economic value of intellectual property.

Our analysis, placed in the context of the experimentally organized economy (EOE) and the Computing & Communications technology of the New Economy, adds a new dimension of complexity. The number of possible arrangements to claim economic property rights is very large, as are the number of opportunities of protecting intellectual properties through innovative private designs of contracts, organizations, and product- and marketing strategies. This variation is costly to maintain, but those costs should be seen in the context of the enormous gains that can be captured in the form of lower total transactions costs, and in the form of a smaller loss of winners. The computing and communications technology of the New Economy is adding to that complexity and raising the stakes of the game. Hence, our second conclusion is that economic property rights to, and tradability in tacit competencies are not enhanced through expanded protection by means of easy access to standardized patent and copyright legislation. Such methods may *lower the degree of variation in the selection process and raise the incidence of business mistakes*. The solution should rather be to facilitate flexible and, in one sense possibly more costly, contractual protection through the market in order to enhance the dynamics of competence blocs. The legal framework must be *enabling* with respect to a variety of contractual arrangements that support economic property rights. Mandatory laws for contractual arrangements should be avoided, and legislation should support arbitration with respect to disputes that are primarily technical in nature.

Obviously, the increased efficiency in the allocation of intangible assets discussed in this paper contributes to diminishing the underinvestment problem as it is discussed in literature. The same improvement in the allocation of intellectual capital also contributes to the solution of a different underinvestment problem not discussed in the literature, namely the capturing of winners that would otherwise have been lost because of weak property rights and low tradability. This is, however, a conclusion that can only be visualized within the domain of the Experimentally Organized Economy.

References

Acs ZI, Audretsch DB (1988) Innovation in small and large firms: an experimental analysis. AER 78: 678–690

Aoki R, Prusa TJ (1995) Product development and the timing of information disclosure under U.S. and Japanese patent systems. Working Paper 5063, NBER Working Paper Series, Cambridge, MA

Arrow KJ (1962b) Economic welfare and the allocation of resources for invention, Chapter 7. In: Nelson R (ed) The rate and direction of inventive activity: economic and social factors, pp 609–626. NBER, Princeton University Press, Princeton

Barzel Y (1997) Economic analysis of property rights. Cambridge University Press, Cambridge, MA

Buigues P, Jacquemin A, Marchipoints J-F (2000) Competitiveness and the value of intangible assets. Edvard Elgar, Aldershot

Carlsson B (ed) (1997) Technological systems; cases, analyses, comparisons. Kluwer, Boston Dordrecht London

Coase RH (1937) The nature of the firm. Economica IV(13–16): 386–405

Cohen WM, Nelson RR, Walsh JP (2000) Protecting their intellectual assets; appropriability conditions and why U.S. manufacturing firms patent (or not). Working Paper 7552, NBER Working Paper Series, Cambridge, MA

Darby ME, Liu Q, Zucker LG (1999) Stakes and stars: the effect of intellectual human capital on the level and variability of high-tech firms' market values. Working Paper 7201, NBER Working Paper Series, Cambridge, MA

Day RH, Eliasson G, Wihlborg C (eds) (1993) The markets for innovation, ownership and control. Elsevier, North-Holland, Amsterdam

Davis L (2002a) Patents and the internet. LEFIC Working Paper 2002-1, Copenhagen Businesss School, Copenhagen

Davis L (2002b) Is appropriability a problem for innovations in digital information goods. LEFIC Working Paper 2002-2, Copenhagen Business School, Copenagen and forthcoming in Eliasson G, Wihlborg C (eds) op.cit. (forthcoming)

Eaton J, Kortum S (1994) International patenting and technology diffusion. Working Paper 4931, NBER Working Paper Series, Cambridge, MA

Eaton J, Kortum S (1995) Trade in ideas: Patenting and productivity in the OECD. Working Paper 5049, NBER Working Paper Series, Cambridge, MA

Eliasson G (1990a) The firm as a competent team. Journal of Economic Behavior and Organization 13(3): 275–298

Eliasson G (1990b) The knowledge-based information economy, Chapter I. In: Eliasson G, Fölster S, et al. (eds) The knowledge based information economy. IUI, Stockholm

Eliasson G (1991a) Financial institutions in a European market for executive competence, Chapter 7. In: Wihlborg C, Fratianni M, Willet TA (eds) Financial regulation and monetary arrangements after 1992. Elsevier, Amsterdam

Eliasson G (1992) Business competence, organizational learning, and economic growth: establishing the Smith-Schumpeter-Wicksell (SSW) connection. In: Scherer FM, Perlman M (eds) [see below]

Eliasson G (1993) A note: on privatization, contract technology and economic growth. In: Day R, Eliasson G, Wihlborg L (eds) Markets for innovation, ownership and control. Elsevier, North-Holland, Amsterdam

Eliasson G (1996a) The firm, its objectives, its controls and its organization. Kluwer, Boston Dordrecht London

Eliasson G (1996b) Spillovers, integrated production and the theory of the firm. Journal of Evolutionary Economics 6: 125–140

Eliasson G (1997a) General purpose technologies, industrial competence blocs and economic growth. In: Carlsson B (ed) Technological systems: cases, analyses, comparisons. Kluwer, Boston Dordrecht London [see above]

Eliasson G (1998) From plan to market. Journal of Economic Behavior and Organization 34(1): 49–68

Eliasson G (2000a) Making intangibles visible – the value, the efficiency and the economic consequences of knowledge. In: Buigues P, Jacquemin A, Marchipon JF (eds) [see above]

Eliasson G (2001a) The role of knowledge in economic growth. In: Helliwell J (ed) [see below]

Eliasson G (2003) The venture capitalist as a competent outsider. In: Alho K-J, Lassila- Pekka Y-A (eds) Economic research and decision making-essays on structural change, growth and economic policy. ETLA, Helsinki

Eliasson G, Eliasson Å (1996) The biotechnological competence bloc. Revue d'Economie Industrielle 78-4^{0}”

Eliasson G, Eliasson Å (2002a) The theory of the firm and the market for strategic acquisitions. Paper presented to the 9th Congress of the Intrernational Joseph A. Schumpeteers Society (ISS), Gainesville, FL, USA, March 28–30, 2002

Eliasson G, Eliasson Å (2002b) Competence blocs in the experimentally organized economy. KTH, TRITA Series

Eliasson G, Granstrand O (1985) Venture capital and management – a study of venture development units in four Swedish firms (mimeo, Jan. 9). IUI, Stockholm, Chalmers University of Technology, Göteborg

Eliasson G, Green C (eds) (1998) The micro foundations of economic growth. University of Michigan Press, Ann Arbor

Eliasson G, Karlsson N (eds) (1998) The limits of government – on policy competence and economic growth. City University, Stockholm

Eliasson G, Taymaz E (2000) Institutions, entrepreneurship, economic flexibility and growth – experiments on an evolutionary model. In: Cantner U, Hanusch H, Klepper S (2000) Economic evolution, learning and complexity – econometric, experimental and simulation approaches. Physica, Berlin Heidelberg New York

Eliasson G, Taymaz E (2002) Modeling derivatives in a firm based evolutionary macro model (MOSES) – the integration of computing, communications and financial services. Mimeo, KTH, Stockholm

Eliasson G, Johansson D, Taymaz E (2002) Simulating the new economy. TRITA -IEO-R 2002:14, KTH, Stockholm. In: Structural change and economic dynamics. Forthcoming

Hall BH, Ham RM (1999) The patent paradox revisited: Determinants of patenting in the US semiconductor industry, 1980–1994. Working Paper 7062, NBER Working Paper Series, Cambridge, MA

Hall BH, Jaffe A, Trajtenberg M (2000) Market value and patent citations: a first look. Working Paper 7741, NBER Working Paper Series, Cambridge, MA

Helliwell J (ed) (2001) The contribution of human and social capital to sustained economic growth and well-being. HRDC, Canada

Hirschleifer J (1971) The private and social value of information and the reward of inventive activity. American Economic Review LXI(4): 561–574

Hultmark CR (1999) European and US perspectives on electronic documents and electronic signatures. Tulane European and Civil Law Forum 14: 123–153

Jaffe AB, Fogarty MS, Banks BA (1997) Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. Working Paper 6044, NBER Working Paper Series, Cambridge, MA

Jaffe AB, Trajtenberg M (1998) International knowledge flows: Evidence from patent citations. Working Paper 6507, NBER Working Paper Series, Cambridge, MA

Jaffe AB, Trajtenberg M, Fogarty MS (2000) The meaning of patent creations. Report on the NBER/Case Western Reserve Survey of Patentees. Working Paper 7631, NBER Working Paper Series, Cambridge, MA

Johnson DKN, Popp D (2001) Forced out of the closet: The impact of the American protection act on the timing of patent disclosure. Working Paper 8374, NBER Working Paper Series, Cambridge, MA

Jonasson A (2001) Innovative pricing. Doctorate thesis, Department of of Industrial Economics and Management (INDEK), Royal Institute of Technology (KTH), Stockholm

Jonasson A (2002) Innovative pricing in the experimentally organized economy. This conference

Jorgenson DW, Griliches Z (1967) The explanation of productivity change. Review of Economic Studies XXXIV(3): 249–282

Kingston W (2000) The case for compulsory arbitration: Empirical evidence. European Intellectual Property Review 9

Kingston W (2002) Intellectual property needs help from accounting. European Intellectual Property Review 11: 508–515

Kortum S (1994) A model of research patenting, and productivity growth. Working Paper 4646. NBER Working Paper Series, Cambridge, MA

Kortnum S, Lerner J (1997) Stronger protection or technological revolution: what is behind the recent surge in patenting? NBER Working Paper Nr 6204

Kremer M (1997) Patent buy-outs: A mechanism for encouraging innovation. Working Paper 6304, December, NBER Working Paper Series, Cambridge, MA

Lanjouw JO, Cockburn I (2000) Do patents matter?: Empirical evidence after GATT. Working Paper 7495, NBER Working Paper Series, Cambridge, MA

Lanjouw JO, Lerner J (1996) Preliminary injunctive relief: Theory and evidence from patent litigation. Working Paper 5689, NBER Working Paper Series, Cambridge, MA

Lanjouw JO, Lerner J (1997) The enforcement of intellectual property rights: A survey of the empirical literature. Working Paper 6296, NBER Working Paper Series, Cambridge, MA

Lanjouw JO, Pakes A, Putnam J (1996) How to count patents and value intellectual property; uses of patent renewal and application data. Working Paper 5741, NBER Working Paper Series, Cambridge, MA

Lanjouw JO, Schankerman M (1997) Stylized facts of patent litigation: Value, scope and ownership. NBER Working Paper Nr 6297

Lanjouw JO, Schankerman M (2001) Enforcing intellectual property rights. Working Paper 8656, NBER Working Paper Series, Cambridge, MA

Lerner J (2000a) 150 years of patent office practice. Working Paper 7477, NBER Working Paper Series, Cambridge, MA

Lerner J (2000b) 150 years of patent protection. Working Paper 7478, NBER Working Paper Series, Cambridge, MA

Lerner J (2000c) Where does state street lead? A first look at finance patents, 1971–2000. Working Paper 7918, NBER Working Paper Series, Cambridge, MA

North DC, Thomas RP (1973) The rise of the western world: A new economic history. Cambridge University Press, Cambridge

Porter ME, Stern S (2000) Measuring the “ideas” production function: Evidence from international patent output. Working Paper 7891, NBER Working Paper Series, Cambridge, MA

Reichman JH (2001) Of green tulips and legal kudzu: Repackaging righs in subpatentable innovation. In: Dreyfuss RC, Zimmerman DL, First H (eds) Expanding the boundaries of intellectual property. Oxford University Press, Oxford

Robinson J (1964) Solow on the rate of return. Economic Journal 74: 410–417

Romer PM (1986) Increasing returns and longrun growth. Journal of Political Economy 94(5): 1002–1037

Sakakibara M, Branstetter L (1999) Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms. Working Paper 7066, NBER Working Paper Series, Cambridge, MA

Scherer FM, Perlman M (eds) (1992) Entrepreneurship, technological innovation, and economic growth. Studies in the Schumpeterian Ann tradition. The University of Michigan Press, Arbor

Schleifer A (1994) Establishing property eights. Discussion Paper Number 1683, Harvard Institute of Economic Research, Harvard University, Cambridge, MA

Schumpeter JA (1911) Theorie der wirtschaftlichen entwicklung. Dunker und Humblot, Jena, English ed., 1934, The theory of economic development: An inquiry into profits, capital, credit, interest and the business cycle, Vol. XLVI. Harvard University Press, Cambridge, MA

Shavell S, Van Ypersele T (1999) Rewards versus intellectual property rights. Working Paper 6956, NBER Working Paper Series, Cambridge, MA

Solow RM (1963) Capital theory and the rate of return. NorthHolland, Amsterdam

Wihlborg C (1998a) The role of enabling and mandatory law for financial systems efficiency. In: Eliasson G, Karlson N (eds) The limits of government [see above]

Wihlborg C (1998b) Economic efficiency with enabling and mandatory law. In: Eliasson G, Green G (eds) The microfoundations of economic growth. University of Michigan Press, Ann Arbor, MI

Zucker LG, Darby MR, Brewer MB, Peng Y (1995) Collaboration structure and information on dilemmas in biotechnology: Organizational boundaries as trust production. Working Paper 5199, NBER Working Paper Series, Cambridge, MA

Capital in the new economy: A Schumpeterian perspective

Jim Stewart[*]

School of Business Studies, Trinity College, Dublin, 2, Ireland
(e-mail: jstewart@tcd.ie)

Abstract. The "dotcom boom" and subsequent collapse raises issues as to the nature of capital and the relationship between capital and investment. Capital in conventional finance, based on the Fisher-Hirshleifer analysis, is defined as postponed consumption and investment is defined as a trade-off between consumption now and in the future. This paper argues that a more satisfactory explanation of the relationship between investment and capital was developed by the Austrian economist Böhm-Bawerk, who identified capital goods as separate from consumption goods, and where the passage of time is fundamental to the accumulation of capital. Such a process assumes risk rather than uncertainty, and does not capture the essence of Schumpeterian investment.

Key words: Capital theory – Investment – Neoclassical – Austrian – Schumpeterian

JEL Classification: B41, D21, D81, G31

1 Introduction

Developed economies are currently experiencing a Schumpeterian gale of "creative destruction". The dot.com boom based on the new technology and the subsequent demise of many dot.com firms cannot disguise the profound impact of the new technology upon which e-business is based. Schumpeter (1970, p. 87) focused

---

[*] The author is grateful for comments from participants at the International Joseph A. Schumpeter Society, Ninth Conference, in Gainsville, Florida. The author would especially like to thank Bill Kingston of Trinity College, Dublin for helpful comments and discussions. He would also like to thank Paul Coughlan and Colm Kearney of Trinity College, Dublin and David Jacobson of Dublin City University. The author would also like to thank the editors and referee for helpful comments.

on the impact of new technology *"on the existing structure of an industry"* and, because of the failure of many firms dependent on e-technology, there is a danger of dismissing its impact. Schumpeter (1970, p. 90) noted that *"Situations emerge in the process of creative destruction in which many firms may have to perish that nevertheless would be able to go on vigorously and usefully if they could weather a particular storm"*. One of the striking differences between dot.com firms and firms in more traditional sectors of the economy (sometimes referred to as "old economy firms") is the large difference that exists between the value of assets in the Balance Sheet of many newly-quoted dot.com firms and their stock market value. For example, for seven Irish dot.com firms quoted on NASDAQ, fixed assets amounted to an average of 10% of total balance sheet values and accounted for less than 1% of the value of the firm at the flotation price (Beirne and Stewart, 2002, Table 1). The pendulum has swung so far that a number of dot.com firms now have cash balances net of all liabilities greater than their stock market value. This means that the stock market places a negative value on these firms' assets other than cash[1]. For example, Energis, a UK-based telecom networking firm, had a stock market value of £ 12 billion in 2001 and had a stock market value of less than 100 million (in 2002). The Balance Sheet for December 2001 shows assets in excess of £ 2 billion, but Energis has been reported as attempting to sell European assets on which it had spent £ 1 billion for £ 1[2]. Such large differences between the stock market value of a firm and the value of balance sheet assets draw our attention to the weakness of commonly-held ideas about the nature of capital and capital values.

This paper describes generally accepted concepts of capital in the finance literature and how these concepts are used to justify the use of investment decision rules such as net present value and internal rate of return. The paper argues that the investment decision process described is unrealistic and misleading in a number of respects. It is argued that the Austrian concept of capital is a more satisfactory alternative, which may also be used to justify investment decision rules. However, the paper also notes that optimal investment decision rules as described in the finance literature are infrequently used in practice. The most likely explanation is that the investment decision process is best thought of as a Schumpeterian process, characterized by uncertainty, and we find that it is difficult if not impossible to model.

The paper is organized as follows. In Sect. 2, we point to a paradigm change in the teaching of investment theory in finance; in Sect. 3, we describe investment and capital in conventional finance theory; in Sect. 4, we provide a critique of

---

[1] One UK based e-business had 36 million in cash and a market value of £ 29 m. The decision was then made to liquidate all the group's assets and return the remaining cash to shareholders, Source: Yahoo Finance UK and Ireland 21/8/01. The Financial Times (3/9/01) lists three dot.com firms with cash balances greater than market values, and a further three with market values slightly greater than their cash balances. The Financial Times (31/10/01) discusses a dot.com firm with a stock market value of Stg. £ 137 million and cash balances of nearly Stg. £ 200 million.

[2] Source: Guardian Newspaper, 22/2/02. There are other examples where current asset values are a small fraction of expenditures. For example, in the case of KPNQwest, a 25,000 Km fibre optic network linking 60 cities in Europe on which €5 billion was spent may be practically worthless, Source: Financial Times, 11/6/02.

this theory; in Sect. 5, we describe Austrian capital theory and we argue that this theory may provide a framework to justify the use of net present value techniques; in Sect. 6 we consider the role that real option analysis may play in investment decision-making. In Sect. 7, we consider whether there could be a Schumpeterian theory of capital, and finally, in Sect. 8, we give some conclusions.

2 An example of paradigm change?

There is an almost total neglect in current mainstream finance writing of the great controversies of the past on the nature of capital and the relationship between capital and investment. Some well known examples are: the extensive criticisms by Knight (1941) of the von Mises theory of capital, the criticisms by Hayek (1936, and 1941 esp. Chap. 1) of Knight, and more recently the Cambridge postwar debates on capital theory (see Harcourt, 1972, for a summary). The nature of capital has not received much critical attention in recent economic theory either. Many finance textbooks continue to give prominence to the particular view of capital associated with Fisher (1930) and extended by Hirshleifer (1958; 1970)[3]. Fisher's theory of investment, which is dependent on his theory of capital, is the basic foundation for resource allocation in financial economics (O'Sullivan, 2000, pp. 44–45). Given the prominence of the Fisher/Hirshleifer analysis in the financial economics literature, it is of interest that the latest edition of the best selling textbook on Corporate Finance by Brealey and Myers (2000) omits what in previous editions had been a seven page discussion entitled "Foundations of the Net Present Value Rule"[4]. There are now several broad-based textbooks on finance which omit any discussion of the nature of capital or the relationship between real investment and the capital stock[5]. The rejection of one paradigm (albeit by omission) without replacement would seem to be at variance with Kuhn's maxim. He states (Kuhn, 1970, p. 77) "The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgement leading to that decision involves the comparison of both paradigms with nature and with each other". (Kuhn (p. 137) also argues that changes in textbooks are indicators of major paradigm shifts.).

We should also note that the contemporary neglect of capital theory is not unique to finance and economics. A sister discipline of finance, accounting, defines capital on the basis of rules. A problem with such a 'rule based' definition of capital is that, while it may appear to be more readily implemented (for example via regulators) than a theory-based definition (such as the discounted sum of expected future revenues), as the rules change, so does the definition of capital. Further issues arise in deciding who sets the rules and who monitors their implementation - points that have been well illustrated by Enron and other cases. As noted above, the growth of 'new economy' firms, in which capital as traditionally viewed is a tiny fraction of overall stock market values, has emphasized the irrelevance of accounting concepts of capital to these types of modern business enterprise.

---

[3] Examples are: Lumby and Jones (2001), esp. pp. 58–74; Cuthbertson (1996, pp. 15–20)

[4] The omitted analysis is available at a web site associated with the text book http://www.mhhe.com/business/finance/bm/npv.mhtml.

[5] Examples are: Bodie and Merton (1998) and Van Horne (2001).

3 Capital in conventional finance theory

Hirshleifer (1970, pp. 40–41) notes that capital in finance (and economics) can have a number of different meanings. While the generally accepted definition of real capital in finance is postponed consumption, the value of capital is defined differently. In finance, capital value has long been defined as the discounted sum of expected future cash flows. For example, Fisher in his book "The Theory of Interest" summarizes his views on the value of capital as follows (1930 p. 12): Capital meaning the value of capital is *"simply the future income discounted or, in other words capitalized"*. Fisher defines income as a "series of events" (Fisher, 1930, p. 3) or the *"psychic experiences of the individual mind"* (Fisher, p. 4). This definition of capital value is, as noted by Schumpeter (1970, p. 96), synonymous with profits. Hirshleifer comments (1970, p. 12) that capital value *"is not a real object like a capital good, but is rather the result of a process of calculation or market valuation"*.

A modern discussion of real capital is most likely to be found in textbooks on finance in the introductory chapters and in the context of providing a theoretical foundation for the net present value rule, for example in Brealey and Myers (2000, pp. 17–24). The analysis is based on Fisher (1930) as extended by Hirshleifer in which capital is 'postponed consumption', investment is a trade-off between consumption now and in a future period, the capital market provides borrowing and lending opportunities and thus permits the separation of consumption from investment flows. This leads to the widely cited Fisher separation theorem, that is, the optimal consumption decision can be taken independently of the optimal investment decision (Copeland and Weston, 1988, p. 10). Some regard the views of Fisher on investment as constituting a theory of the firm. For example Miller (1988, p. 103) states:

> *"Irving Fisher's view of the firm [...] impounds the details of the technology, production, and sales in a black box and focuses on the underlying cash flow. The firm for Fisher was just an abstract engine transforming current consumable resources, obtained by issuing securities, into future consumable resources payable to the owners of the securities"*.

The analysis developed by Hirshleifer is used to give a theoretical justification to the net present value rule. In order to illustrate such theories, examples are often drawn from primitive agriculture or simple societies. Fisher (1930) has frequent references to farming and forestry (pp. 159–164) or to farming, forestry, and mining (pp. 130–141). Hirshleifer (1970, p. 2) in contrasting production with exchange, states that production "involves the physical transformation of some commodities into others" "as when Crusoe fells a tree to build a canoe"[6], so that "there is one less tree but one more canoe". He then defines a capital good (p. 154) as:

---

[6] Hirshleifer and others may prefer such 'unreal' examples because as he notes (1970, p. 11) economists are not clear on what a firm is and why it exists. We should also note that appeals to 'Robinson Crusoe' type economies are widespread in texts explaining optimum investment decisions and have a long history. Examples are Fisher (1930, p. 182 and p. 248); Bromwich (1976, pp. 54–63); Copeland and Weston (1988, p. 5).

> *"a physical object existing in the present, but constituting the source of incomes or consumption opportunities in the future as an apple tree is a source of future apples, corn seed a source of next year's corn, and a house is the source of future shelter. The individual's choice between this year's corn and next year's corn, for example, could be expressed as a choice between this year's corn and this year's seed".*

More recently, the textbook on financial theory by Copeland and Weston (1988, p. 3) states in the introduction that logical development is facilitated by assuming a *"one person, one good economy"* in which:

> *"The decision maker, Robinsoe Crusoe must choose between consumption now and consumption in the future. Of course, the decision not to consume now is the same as investment"*

The use of the Robinson Crusoe metaphor is not just restricted to illustrating investment decision rules in finance, but can also be found in other areas of economics[7]. In addition, illustrative metaphors of investment decision rules in the finance literature are not just restricted to Robinsoe Crusoe-type societies or agriculture. As an example, the text by Brealey and Myers (2000, pp. 17–24) uses the metaphor of the 'prodigal and the miser' to illustrate how capital markets can be used to increase current consumption by borrowing against future cash flow or to lend current cash flow[8].

4 Critique of neoclassical investment theory

This section argues, in contrast to neoclassical investment theory, that investment is *not* reversible into consumption goods. Capital goods are *not* the same as consumption goods. Investment in the sense of capital augmenting is difficult to define and measure. Because investment takes place in the context of uncertainty, the value of an investment may turn out to be zero.

There are many criticisms of neoclassical capital theory as described above. For example, in the two-period diagrams used to illustrate investment decision rules, it is assumed that there are diminishing returns to investment, that is, the relationship between consumption in period t_0 and t_1 is convex. In addition, utility curves may not have the assumed concave shape. Rather than criticize technical aspects of the analysis, O'Sullivan (2000) simply rejects the assumed investment decision model. She notes (p. 47) that Fisher ignores the relationship between the rate of interest,

---

[7] A search of a data base (JSTOR) of 15 of the main journals in economics (mostly U.S. based) for the period 1900- 1994 shows 203 different articles refer to Robinson Crusoe, 27 of which occur post 1980.

[8] McCloskey (1994, Chap. 4) argues that the use of rhetoric, including metaphors, to persuade is fundamental to economic reasoning. A possible explanation for omitting the Fisher/Hirshleifer analysis with references to 'Robinson Crusoe' type economies is that this metaphor no longer persuades, as a modern audience is less likely to be familiar with the story by Daniel Defoe. As McCloskey states (p. 51) "rhetoric is speech with an audience". Also see Henderson for a further discussion of the use of metaphors in economic texts.

the productive process, division of labour, and the role of management. Thus Fisher and financial economists assume that productive capacity and the alternative uses to which these resources may be allocated are given. This static concept of resource allocation ignores its 'developmental' aspect, that is, the *"irreversible commitment of resources today for uncertain returns in the future"* (O'Sullivan, 2000, p. 20). The difficulties are further compounded by adding another source of income, the risk premium, without explaining why and how a return to risk becomes generated in the real economy. She states (2000, p. 48), modern financial economists *"analyse why it is that portfolio investors would demand a return on the securities they hold without ever posing the question of why such a return might be forthcoming in the economy"*.

O'Sullivan (2000, p. 22) also argues that the standard neoclassical concept of resource allocation has three other aspects – it is reversible, individual and optimal.

1. The assumption of reversibility means that the allocation of resources today has no influence on subsequent allocation. Resources may have to be committed over long periods, and failure to invest still further resources may mean the entire investment is wasted. Furthermore, the value of an investment made in one period or for the life of a project may turn out to be zero.
2. Neoclassical resource allocation assumes that resources can be individually allocated without reference to the decisions and actions of other individuals. This ignores the reality that learning is likely to be collective and cumulative and is best able to take place within organizations. Not only does it involve cooperation within organizations, it may also involve cooperation *between* organizations. This latter aspect can be illustrated by the extensive literature emphasizing the positive role of inter-firm cooperation, in innovation and learning within networks of firms (see Cooke, 1996, pp. 6–9). More generally, Granovetter (1985) argues that *"social relations between firms are more important, and authority within firms less so, in bringing order to economic life than is supposed in the markets and hierarchies line of thought"*.
3. The third key assumption of neoclassical resource allocation is that resource allocation is optimal. Investments are selected, given markets and technology. However, investment, rather than being a choice among alternative economic outcomes, may seek to transform existing market and technological conditions.

Criticism of accepted investment theory (use of net present value) has also come from within the finance literature. Dixit and Pindyck (1994, p. 3) argue that investment decisions have three key characteristics: they are *"partially or completely irreversible"*, there is uncertainty, and there is scope in relation to the timing of investments.On this basis, they argue investments are best viewed and analyzed as an *"opportunity to invest"* (1994, p. 6), or as an option.

The most important criticism of standard investment theory is the failure to address the issue of uncertainty. Standard investment theory can incorporate risk, where Knight (1921, p. 233) argues that *"the distribution of outcomes is known* *while in the case of uncertainty this is not true"*, but as noted by Loasby (2001, p. 403), much of economics avoids *"Knightian uncertainty"*. LeRoy and Singell (1987, p. 395) argue that *"economists especially working in the neoclassical tradition invoke the distinction [between risk and uncertainty] only in order to rule out uncertainty"*. Nordhaus (1969, p. 55), in discussing uncertainty and risk in relation to the invention process, introduces the concept of 'mild uncertainty', thereby suggesting a continuum rather than a dichotomy. Others redefine uncertainty as risk. For example, Dixit and Pindyck (1995), while acknowledging the role of uncertainty in investment decision-making, equate uncertainty with risk. They state (p. 115) that *"uncertainty requires that managers become much more sophisticated in the ways they assess and account for risk"*. Elsewhere (1994) they state that underlying uncertainty can be described as a mathematical process (dynamic programming, p. 93 and an Ito process p. 114), so that with uncertainty *"the current state determines only the probability of future states, not the actual values"* (p. 12).

Knight (1921) also has extensive discussion about the theory of knowledge, psychology and decision-making in order to shed light on the distinction between risk and uncertainty (1921, pp. 197–222). In doing so, although not acknowledged in surveys of the area (for example Hirshleifer, 2001), Knight would appear to anticipate current extensive interest in behavioral finance.

Given the existence of uncertainty, the continued survival and profitability of insurance companies would appear to be an anomaly[9]. As noted by Schumpeter (1955, p. 894), Knight argued that risk is insurable but uncertainty is not. The answer may be that consumers of insurance products are willing to pay far more that the estimated actuarial value of insurance (Scherer, 2001, p. 17). Purchasing insurance is not an expected 'fair game' but involves an expected net reduction in wealth. Such behavior may in turn be explained by the relationship between utility and changes in wealth as developed by Friedman and Savage (1948).

5 Austrian theory

A theory of investment, or of capital accumulation in which capital is defined as postponed consumption with the implicit assumption that it is reversible at a future date into consumption goods, as noted above, has been subject to extensive criticism and may partly explain why the formerly generally accepted Fisher/Hirshleifer analysis has been omitted from recent textbooks. The issue is what can replace this theoretical framework. Is there a theory that can provide a justification for investment decision rules such as Net Present Value or Internal Rate of Return? This paper argues that the theory of capital accumulation in Austrian economics is one possibility for two main reasons: (1) it is more realistic in that it assumes capital is an identifiable object separate from consumption goods, and (2) the passage of time is an essential aspect of the productivity and value of capital.

As is well known, Schumpeter had great admiration for Austrian economists and in particular Böhm-Bawerk[10]. However, there are differences between what is

---

[9] The author is grateful to Gunnar Eliasson for this point.

[10] Schumpeter (1997, p. 156) describes the style of Böhm-Bawerk writing *"as direct, unadorned, reserved. The author lets the subject speak and does not distract us with his own fireworks"*. In contrast,

generally regarded as Austrian economics and Schumpeterian economics. For example, Jacobson (1992, p. 787) states that Schumpeter's notion of the market being, at times, in equilibrium separates him from the 'mainstream' Austrian viewpoint. Fisher, a contemporary of Schumpeter, and acknowledged as the original source of generally accepted ideas on capital and investment, dedicated his book the Theory of Interest (1930) to Böhm-Bawerk[11]. The writings of Böhm-Bawerk also had considerable influence on the theory of capital developed by Austrian economists such as Hayek and von Mises[12]. In the Austrian theory of capital, Schumpeter (1997, p. 156), quoting from Böhm-Bawerk, defines capital as 'nothing but the total of the intermediate products which are generated in the various stages of the roundabout method of production'. What is meant by the term 'roundabout methods of production' is the use of *"nonconsumables"* or *"tools"* (Schumpeter, 1997, p. 165) in the production process. In his History of Economic Analysis, Schumpeter (1955, p. 904) describes the Austrian view of capital as consisting of *"intermediate products such as tools and raw materials"*.

Capital is first of all identified as a *"problem"* and Schumpeter (1997, p. 164) identifies two radically different aspects of this *"problem"* in the writings of Böhm-Bawerk. That is, *"the problem of capital"* as a means of production and the problem of capital as a source of net return. Capital as a means of production involves two key concepts. First, Labor is applied in producing *"nonconsumables with whose help final products can be produced more efficiently"*. Secondly, production is necessarily 'roundabout'. Consumables are not produced directly, but rather capital is produced and this capital can then be used to produce consumables. Capital is then defined as 'nothing but the total of the intermediate products which are generated in the various stages of the roundabout method of production'. One writer, in summarizing the key elements of the Austrian theory, quotes Wicksell to the effect that all capital goods are ultimately *"saved up labour and saved up land"* Smithies (1941, p. 767).

Hicks (1973) further develops Austrian capital theory. He defines a capital good as one that *"can be used in any way to satisfy wants in subsequent periods"* (Hicks, 1973, p. 1). This means that some capital goods can satisfy both current and future wants, described as having the property of present-future jointness. The value of capital goods which do not have this property are a function only of satisfying expected future wants. Hicks develops a model which assumes that goods that may satisfy current and future wants are rented and hence all capital goods are owned by firms or producers. By assuming a continuous flow of production rather than as

---

Kuenne (1971, p. 6), while sympathetic to Böhm-Bawerk, comments that his work is *"...hurried, prolix, polemic, and deficient technically, offering many opportunities for sallies by opponents, it bears silent witness to the life of a harried civil servant unable to devote the full energies of his creative potential to the tedious but valuable academic dialectic"*.

[11] The book is also dedicated to John Rae, a 19^{th} century Canadian economist.

[12] In a review of *"The Pure Theory of Capital"* by Hayek the reviewer states *"since his [Hayek] belief is, and always has been, that the ultimate truth was revealed in varying degrees to Jevons, Böhm-Bawerk and Wicksell, his present book consists in the main of higher criticism and extension of the work of these authors, rather than of new developments in economic doctrine"* (Smithies, 1941, p. 767). In a critical review of a book by Von Mises, Knight (1941) states Von Mises *"is a defender of Böhm-Bawerk's theory of capital"*.

assumed in the Austrian theory, a point input-point output process, the concept of the 'period of production' is no longer a relevant concept.

Investment decision rules within Austrian theory. Kuenne (1971, p. 73), in his critique of Böhm-Bawerk, concludes that, despite all the flaws, the passage of time is a fundamental aspect of economic theory and of the real economy.

Dobrovolsky (1971) formalized the Austrian theory of Böhm-Bawerk to provide a coherent rationalization for the use of investment decision rules such as internal rate of return and net present value, albeit in the context of the neoclassical theory of resource allocation, so that in a competitive economy firms maximize profits by equating marginal revenue to marginal cost. A key difference is that time is an essential element of resource allocation. In other words, the decision rule for using a factor of production is to equate marginal revenue with marginal cost, but either costs are compounded forward, or revenues are discounted to take account of the time-value of money. Dobrovolosky extends this analysis to a number of different time periods. Capital is thus defined as resources used up in a production process.

We can thus appreciate that investment decision rules are in essence similar to the use of rules relating to the employment of an additional factor of production within the theory of the firm, that is, they equate marginal revenue to marginal cost, but with a major difference – the inclusion of time. The appendix gives a fuller explanation of this approach.

The Austrian theory thus identifies capital as a separate object from consumption goods. This theory also has the added advantage that investment decision rules widely used in finance and capital budgeting can be explained within the logic of equating marginal cost to marginal revenue, but with the additional requirement that costs or revenues must be adjusted for the time-value of money. Important issues still remain, for example, the extent to which such an investment decision process adequately describes investment decisions within the corporation and the long standing disparity between the theory of optimal investment decisions and the practice of capital budgeting[13]. The most likely explanation for this disparity is not that investment decision makers are irrational, but that the investment decision process is characterized by uncertainty in the Knightian sense, may require cooperation within as well as between firms, and may be path dependent. All of these factors may make optimal investment decision rules redundant, in particular for investments in new technology. Sharp (1991, p. 69) comments that reliance on the use of net present values may rule out *"very risky, but strategically vital"* investments, and suggests adding to discounted cash flows the value of the options

---

[13] Mao (1970) comments *"There exists a wide disparity between the theory and practice of capital budgeting"*. Rockley (1973), in a survey of investment appraisal techniques for U.K. based companies, found that, while a majority used DCF techniques, this was generally combined with other non-discounting procedures (pp. 136–137). Risk assessment was based on 'hunch' or 'judgement' (p. 148). More recently, in a survey of the use of capital investment appraisal techniques, Sangster (1993, p. 309) concludes that, for the UK, studies have consistently shown that simple non-discounting techniques are more widely used than discounting techniques. Dixit and Pindyck (1995, p. 108–109) report research that, in estimating net present values, managers use discount factors three or four times larger than that required. Conversely, implied discount rates in disinvestment decisions are considerably below those required.

created by the risky investment, such as organizational learning, from developing new products or production techniques. The next section examines the role of real options in investment decision-making.

6 The use of real options in investment decision making

The use of real options in business decision-making has been described as one of the most important developments in the last hundred years (Howell et al., 2001, p. 1). The key insight in the development of real options was to recognize that actions that create flexibility in relation to timing or greater understanding of risk have considerable economic value. The development of real options is an extension of options theory from financial products to non-financial products. Financial options are part of a set of financial assets known as derivatives. This is because their value derives entirely from the value of another asset or group of assets. Options have been introduced and extended to a broad range of financial products and can be used as a form of insurance in terms of risk reduction, as a financial asset within a portfolio, or as the main components of a portfolio. In contrast to financial options, which may trade on markets and in most cases have no effect on the firm (analogous to trading equity in the secondary market), real options are always internal to the firm and can affect firm value. An important decision in relation to real options is how much to pay (or how much to charge if selling). Expenditure on R and D may thus be viewed as purchasing an option. The more that is spent, the greater the cost of the option. The economic value of this option may arise from new information on existing products or new products that the firm can then choose to develop or abandon or new information that may be revealed which results in changes in investment programs (abandoning investment or increasing investment). Eastman Kodak for example, has claimed success in using option pricing in valuing R and D expenditures (Faulkner, 1996).

Another key decision is choosing when to exercise an option. Thus investment decisions which are flexible as to their timing have a valuable embedded option. While entrepreneurs may have always valued flexibility and options, recent developments in real options theory provides an intellectual framework in which ‘hunch’ or counter intuitive thinking can be formally analyzed. There are, however, also limitations in the use of real options. The first limitation (familiar to entrepreneurs) is that flexibility creates options but also has a cost. This may arise through two factors: increased cost of the investment (which can be incorporated in the value of an option) and increased discretionary behavior. Agency theory has long argued that there are conflicts of interest both between managers and shareholders and between different providers of finance. Increased flexibility coupled with information asymmetries increases agency costs because of increased possibilities for opportunistic behavior. Second, while real option theory claims to value options, there are uncertainties in relation to option values. Increased volatility in the value of the underlying asset increases the value of an option. There is likely to be considerable uncertainty in relation to future cash flows and hence in option values estimated as the net present value of the potential investment. Real option theory assumes that the firm is a price taker (Howell et al., 2001, p. 193). This may not be the case where market structure is dominated by a key firm (such as Microsoft in PC applications). The length of time to expiry of an option influences option value, so that underestimating this time reduces option value. Estimating this time period in the case of a real option may be difficult, as for example where the value of an option created through R and D is a function of the ability of competitor firms to create alternative technologies. The major variables in valuing real options are thus characterized by Knightian uncertainty rather than risk.

Baldwin and Clark (2000) have argued that changes in industrial organization, in particular the growth of modular organization, has boosted innovation because of the real options created. They argue that the computer industry, as an example, reflects a modular structure (Baldwin and Clark, 2000, Chap. 1), but manufacturing and service industries are increasingly characterized by modular organizations. A product can be broken up into subsystems or modules, resulting in much greater flexibility. While some firms may seek to own and to control each module, greater efficiency results from different firms owning different modules (Baldwin and Clark, 2000, p. 154). As long as each supplier meets the design rules, they are free to experiment with product design (Baldwin and Clark, 2000, pp. 150–154). There are exceptions. Japanese computer firms are much more integrated and diversified than US firms (as noted by Baldwin and Clark, 2000, p. 7). In addition, modularity in production and design does not appear to be the overriding feature of all successful innovating firms. For example, Siemens, while stressing the importance of innovation (citing the number of patents issued), also states that the four components of its strategy involve developing a portfolio of 14 diverse businesses which follow different cycles, a global presence, skilled management, and the pursuit of synergies across the different businesses[14].

Baldwin and Clark have further argued that, while modularity boosts the rate of innovation, it also creates ‘uncertainty’, which they define as a situation in which possible outcomes can be described by the normal distribution (Baldwin and Clark, 2000, p. 255). The key economic drivers of modularity are the valuable options created. Baldwin and Clark argue that a modular design process creates at least as many options as there are modules. The value of an option created via modularity is similar to the value of a real option. For example, volatility in the underlying asset is measured as the ‘technical potential’ of a module (2003, p. 162). If the dot.com boom is taken as a major example of modularity, where each dot.com firm represents one or more real options as argued by Baldwin and Clark, the question arises as to why economic value was lost in the aggregate, rather than created by the new dot.com firms. Baldwin and Clark suggest that a key problem was how investors valued the options represented by dot.com firms. As noted above, valuing real options is necessarily imprecise, but the value of the real options represented by dot.com firms was consistently grossly overestimated. Baldwin and Clark (2000, p. 153) state that one problem is that they assumed ‘rational expectations’, meaning that the probabilistic structure of outcomes was

---

[14] Press release 23/1/03 available at http://www.siemens.com/.

known to participants[15]. As argued in this paper, investors and entrepreneurs do not make decisions on the basis of a known probability distribution. Investment decision outcomes and hence rules (particularly amongst institutions such as pension funds) may emphasize benchmarking, that is, comparison with a peer group. If reward structures are based on such comparisons, a 'rational' portfolio strategy encourages convergence in investment choice. Further difficulties arise because of agency-type problems. As noted earlier, real options create flexibility and new investment opportunities and, if markets are not perfect, due to information asymmetries, there is considerable scope for opportunistic behavior. The mismatch between the market value of dot.com firms (reflecting the value of the options created) and their true economic value can be partly explained by agency-type problems coupled with the failure of regulators[16].

As Schumpeter (1970, p. 88), states *"Long-range investing under rapidly changing conditions, especially under conditions that change or may change at any moment under the impact of new commodities and technologies, is like shooting at a target that is not only indistinct but moving - and moving jerkily at that"*. The reconciliation of investment decision-making, particularly entrepreneurial decision making, with rational choice is difficult. The probability of successful innovation is very low (Kingston, 1994, p. 666). The profits from innovation are highly skewed. A small number of winners obtain the bulk of returns. (Scherer, 2001, p. 11), and the dispersion of successful innovation is such that risk may not be reduced by diversification (Nordhaus, 1969, p. 56). Loasby (1996, p. 29), following Shackle, argues that entrepreneurship cannot be reduced to logic, but can be best described *"as the imagined deemed possible"* and hence is not subject in many aspects to rational calculation, so that entrepreneurs and entrepreneurial firms *"appear to possess the capacity of seeing things in a way that afterward prove to be true"*.

7 Is there a Schumpeterian theory of capital?

Schumpeter (1951, p. 123) specifically says that capital is not *"a stock of consumption goods"*. However, Kendrick (1961) associates Schumpeter with a stock concept of capital, that is, the 'inventory' concept of capital. According to Kendrick (1961, p. 102), this concept was first developed by Fisher, who defined capital *"as the stock of wealth of all kinds that exists at any moment" "including human resources"* (p. 103). Kendrick states that Schumpeter (1955) credits Fisher (Nature of Capital, 1906) as *"... clearly defining capital as the stock of wealth of all kinds that exists at any moment"* (Kendrick, 1961, p. 102). However, a close reading of the source (Schumpeter, 1955, History of Economic Analysis, pp. 627–628) shows that Schumpeter does not discuss a definition of capital as the stock of wealth and does not attribute (in this reference) such a definition to Fisher.

---

[15] In developing a model to value options created by modularity, Baldwin and Clark (2000, pp. 255-56) assume the value of each module is normally distributed. The normal distribution is widely assumed in finance. Wilmott (2001, p. 116) states that *"the Normal distribution occurs naturally in many walks of life"*. Yet the only illustration given is from a sequence of coin tossing experiments.

[16] See, for example, the auditor regulation and independence debate, the collapse of Enron and Andersen, and issues relating to the independence of stock market analysts.

As noted earlier, it is well documented that Schumpeter thought highly of the work of Böhm-Bawerk. Haberler (1950), in a review of the life and work of Schumpeter, states (p. 343) that *"Schumpeter undoubtedly was influenced by Böhm-Bawerk's capital theory, and used the concept of roundabout methods of production in his theory of economic development. But it gave it a special dynamic slant and his interest theory is entirely different from the Böhm-Bawerkian"*. In addition, Schumpeter would appear to have two widely different views of capital. One is to define capital in money or accounting terms, as in The Theory of Economic Development (p. 122), where he states:

> *"We shall define capital then, as that sum of means of payment which is available at any moment for transference to entrepreneurs"*.

In the *History of Economic Analysis* (1954, p. 323) he states that, if economists had stuck to a monetary concept, *"meaning either actual money, or claims to money"*, a *"mass of confused, futile and downright silly controversies"* would have been avoided. This 'confusion' has also been commented upon by others. Leijonhufvud (1968) explains the 'confusion' as being due to the pre-paradigm state of capital theory. That is, there is no generally accepted agreement as to what issues capital theory should deal with. As a result, Leijonhufvud states (p. 218) "It is not surprising (though it is significant) that these past controversies were at the same time so heated and so imprecise". In discussing real capital, (Hayek, 1941, p. 9) comments that *"The consequent ambiguity of the term capital has been the source of unending confusion, and the suggestion has often been made that the term should be banned entirely from scientific usage"*. This alternative view of capital is ascribed to Schumpeter in a footnote[17]. Thus it would appear that modern textbooks in finance in omitting any discussion on the nature of real capital may be following a Schumpeterian prescript.

Although in the *History of Economic Analysis* (pp. 632–633) this viewpoint is qualified, where Schumpeter states that, while *"in a sense it is true"*, real capital may not have *"any generally received meaning"*, for most analytic purposes capital can be described as *"goods"*. Yet not all goods can be included as capital. If capital is defined as those goods required for production, then the appropriate definition is *"the concept of produced means of production"* or as called by Böhm-Bawerk *"intermediate products"*.

8 Conclusion

This paper has argued that a satisfactory discussion of capital is missing from much of the modern finance literature. The problem of what real capital is and how it relates to investment is acute, given a widespread process of Schumpeterian creative destruction. The returns from investment are uncertain, investment as a process is path dependent and is not reversible. The assumption of reversibility into consumption goods in generally accepted theories of investment excludes the possibility that

---

[17] The full reference given is Schumpeter, Handwörterbuch der Staatswissenschaften, 4. Aufl., vol. 5 p. 582.

investment may result in capital that is worthless (as in the dot.com boom). This paper has argued that a more satisfactory explanation of the relationship between investment and capital was developed by the Austrian economist Böhm-Bawerk, who identified capital goods as separate from consumption goods; here, the passage of time and hence discounting/compounding are fundamental to the accumulation of capital. Nevertheless, such a process does not capture the essence of the Schumpeterian process of investment. As argued by Schumpeter, the process of measuring capital is fraught with difficulties, and can serve to distract from more fundamental economic processes. Schumpeter (1970, p. 83) states *"The fundamental impulse that sets and keeps the capitalist engine in motion comes from new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates".*

Appendix

The following draws extensively on Dobrovolsky (1971, Chap. 3).

Output of a firm (Q) is defined to be a function of a number of different productive factors or inputs:

$$Q = f(a_1, a_2, a_3, ..., a_n) \tag{1}$$

It is assumed that, at the level of the firm, there is no need to distinguish between real assets and financial assets. It is also assumed that the production process can be described as a point-input, point-output process. Factors of production or resources are used at one point in time and revenues accrue at another point in time.

Marginal physical product is represented by the first partial derivative of output with respect to the input in question:

$$\frac{\delta Q}{\delta a_1}, \frac{\delta Q}{\delta a_2}, ... \tag{2}$$

If it is assumed that pure competition exists in the factor market, price is independent of quantity, so marginal cost of $a_1 = w_1$ (w = price per unit).

As output changes, price will remain constant under perfect competition, as firms are price takers. Hence, the marginal revenue product of using one additional unit of a factor of production equals the marginal physical product times the product price.

$$M.R.P.of a_1 = \frac{\delta Q}{\delta a_1} p \tag{3}$$

Total profit (P) is maximized when $\frac{\delta P}{\delta a_1} = p\frac{\delta Q}{\delta a_1} - w_1 = 0$. That is, the change in profit with respect to a unit change in an additional factor of production is equal to the change in the marginal revenue product minus the cost of the factor. The above expression can be rearranged so that:

$$w_1 = \frac{\delta Q}{\delta a_n} \tag{4}$$

The essence of the Austrian theory of capital accumulation is that there is a gap between the use of a factor of production and the resulting output. The theory also argues that the larger this gap, the greater the resulting increase in output. In a simple production function, output (Q) in the next period ($t + 1$) is a function of the use of a factor of production (L) in this period (t).

$$Q_{t+1} = f(L_t) \tag{5}$$

Revenue (R) is obtained from the sale of this output hence revenue is earned in the next period and, assuming all output is sold, revenue is equal to is equal to output by price.

$$R_{t+1} = (Q_{t+1})p \tag{6}$$

The marginal revenue product of this factor of production (using the product rule for differentiation), that is, the change in revenue with respect to a one unit change in the use of this factor, is:

$$\frac{\delta R_{t+1}}{\delta L_t} = \frac{\delta p}{\delta L_t} Q_{t+1} + \frac{\delta Q_{t+1}}{\delta L_t} p \tag{7}$$

If perfect competition is assumed, then the price of the product p cannot change as output changes, and so the first term on the right hand side equals zero, and the marginal revenue product of labor becomes equal to:

$$\frac{\delta R_{t+1}}{\delta L_t} = \frac{\delta Q_{t+1}}{\delta L_t} p \tag{8}$$

Total cost (C) incurred at the beginning of the period t is:

$$C_t = wL_t \tag{9}$$

In order to compare total cost incurred at the beginning of period t with revenue realized at the beginning of period $t + 1$, one must take account of interest on the capital funds invested. If the funds had not been used to pay for resources, they could have been invested outside the firm at a certain rate of interest (i). The maximum rate is the opportunity cost which must be added to the cost of resources used at the beginning of time period t in order to arrive at the firm's total cost of producing the output at Q_{t+1}. Total cost incurred over the entire production period is thus cost incurred in period t compounded forward to period $t + 1$.

$$C_{t+1} = wL_t(1 + i) \tag{10}$$

The marginal cost of resources used can be defined as the change in costs with respect to a one unit change in the use of resources:

$$\frac{\delta C_{t+1}}{\delta L_t} = \delta[wL_t(1 + i] = \frac{\delta w}{\delta L_t} L_t(1 + i) + w(1 + i) \tag{11}$$

Again, under the assumption of perfect competition in the market for resources, the first term on the right hand side is equal to zero because prices do not change as a result of one firm's change in demand, and hence the marginal cost of one unit of resources becomes $w(1+i)$. Within the neoclassical model of the firm, profits will be maximized by employing a unit of a resource such as labor up to the point at which the marginal cost of a factor of production (making allowances for the cost of timing differences) is equal to the marginal revenue product:

$$w(1+i) = \frac{\delta Q_{t+1}}{\delta L_t} p \tag{12}$$

In other words, the decision rule for using a factor of production is: equate marginal revenue to marginal cost but either costs or revenues must be adjusted to take account of the time value of money. Dobrovolosky extends this analysis to a number of different time periods. Capital is thus defined as resources used up in a production process.

References

Baldwin CY, Clark KB (2000) Design Rules Volume 1 The Power of Modularity. MIT Press, London

Beirne A, Stewart J (2002) The Financing of e-Business Firms. International Journal of Business Performance Management special issue on e-business Strategic Planning and Development, 4, 2

Bodie Z, Merton R (1998) Finance. Prentice-Hall, New Jersey

Brealey R, Myers S (2000) Principles of Corporate Finance, 6th ed. McGraw-Hill, London, and 5th ed. (1996)

Bromwich M (1976) The Economics of Capital Budgeting. Penguin Books, Harmondsworth,UK

Cooke P (1996) Enterprise Support Policies in Dynamic European Regions. NESC Networking for Competitive Advantage, 100

Copeland TE, Weston JF (1988) Financial Theory and Corporate Policy 3rd ed. Addison Wesley, London, and 1st ed. (1979)

Cuthbertson K (1996) Quantitative Financial Economics. John Wiley and Sons, Chichester,UK

Dixit AK, Pindyck RS (1994) Investment under Uncertainty. Princeton University Press, Princeton

Dixit AK, Pindyck RS (1995) The Options Approach to Capital Investment. Harvard Business Review, 105–115

Dobrovolsky SP (1971) The Economics of Corporation Finance. McGraw-Hill, London

Faulkner TW (1996) Applying Options Thinking to R and D Valuation. Research Technology Management, 39, 3: 50–56

Fisher I (1930) The Theory of Interest, Reprinted. Augustus Kelley Publishers, Fairfield, N.J.

Friedman M, Savage LJ (1948) The Utility Analysis of Choices Involving Risk. Journal of Political Economy, 56: 279–304

Granovetter M (1985) Economic Action and Social Structure: The problem of Embeddedness. American Journal of Sociology, 91: 481–510

Haberler G (1950) Joseph Alois Schumpeter 1883-1950. Quarterly Journal of Economics, 64, 3: 333–372

Harcourt GC (1972) Some Cambridge Controversies in the Theory of Capital. Cambridge University Press, Cambridge

Hayek FA (1941) The Pure Theory of Capital. Routledge and Kegan Paul, London

Hicks JR (1973) Capital and Time A Neo-Austrian Theory. Clarendon Press, Oxford

Hirshleifer D (2001) Investor Psychology and Asset pricing. Journal of Finance, 1, 4: 1533–1597

Hirshleifer J (1970) Investment, Interest and Capital. Prentice Hall, Englewood Cliffs, N.J.

Howell S, Stark A, Newton D, et al. (2001) Real Options. Financial Times and Prentice Hall, London

Jacobson R (1992) The Austrian School of Strategy. Academy of Management Review, 17, 4: 782–807

Kendrick JW (1961) Some Theoretical Aspects of Capital Measurement. American Economic Review Papers and Proceedings, 51, 2: 102–111

Kingston W (1994) Compulsory Licensing with Capital Payments as an Alternative to Monopoly Grants for Intellectual Property. Research Policy, 23, 5: 661–672

Knight FH (1921) Risk, Uncertainty, and Profit. Houghton Mifflin, Boston

Knight FH (1941) Professor Mises and The Theory of Capital. Economica, 32: 409–427

Kuenne RE (1971) Eugen Von Böhm-Bawerk. Columbia Essays on Great Economists. Columbia University Press, New York

Kuhn TS (1970) The Structure of Scientific Revolutions, 2^{nd} ed. International Encyclopedia of Unified Science. University of Chicago Press, Chicago

Leijonhufvud A (1968) On Keynesian Economics and the Economics of Keynes: a Study in Monetary Theory. Oxford University press, Oxford

LeRoy SF, Singell LD (1987) Knight on Risk and Uncertainty. Journal of Political Economy, 95, 2: 394–406

Loasby BJ (1996) The Imagined, Deemed Possible. In: Helmstädter E, M P (eds.) Behavioral Norms, Technological Progress and Economic Dynamics, University of Michigan Press, Ann Arbor

Loasby BJ (2001) Time, Knowledge and Evolutionary Dynamics: Why Connections Matter. Journal of Evolutionary Economics, 11, 4: 393–412

Lumby S, Jones C (2001) Fundamentals of Investment Appraisal. Thompson Learning, London

Mao JCT (1970) Survey of Capital Budgeting Theory and Practice. Journal of Finance, 25, 2: 349–360

McCloskey DN (1994) Knowledge and Persuasion in Economics. Cambridge University press, Cambridge

Miller MH (1988) The Modigliani-Miller Propositions After Thirty Years. Journal of Economic Perspectives, 2, 4: 99–120

Nordhaus WD (1969) Invention, Growth, and Welfare. M.I.T. Press, Cambridge, Mass.

O'Sullivan MA (2000) Contests for Corporate Control. Oxford University Press, Oxford

Rockley LE (1973) Investment for Profitablity. London Business Books, London

Sangster D (1993) Capital Investment Appraisal Techniques: A Survey of Current usage. Journal of Business Finance and Accounting, 20, 3: 307–332

Scherer FM (2001) The Innovation Lottery. In: Dreyfuss, Zimmerman, First (eds.) Expanding the Boundaries of Intellectual Property, Oxford University Press, Oxford

Schumpeter JA (1951) Theory of Economic Development. Harvard University Press, Harvard

Schumpeter JA (1955) History of Economic Analysis. University of Oxford Press, Oxford

Schumpeter JA (1970) Capitalism, Socialism and Democracy [first published 1942]. Unwin University Books, London

Schumpeter JA (1997) Ten Great Economists, with an introduction by Mark Perlman. Routledge, London

Sharp D (1991) Uncovering the Hidden Value in High-Risk Investments. Sloan Management Review Summer, 69–74

Smithies A (1941) Professor Hayek on the Pure Theory of Capital. American Economic Review, 31, 4: 767–779

Van Horne JC (2001) Financial Management and Policy, 12th ed. Prentice Hall, New Jersey

Wilmott P (2001) Paul Wilmott Introduces Quantitative Finance. John Wiley and Sons, Chichester

A comparative perspective on innovation and productivity in manufacturing and services

Hans Lööf[*]

Royal Institute of Technology, Centre of Excellence for Science and Innovation Studies
(e-mail: hansl@infra.kth.se)

Abstract. This paper serves as another complementary link in a chain of a rather limited number of investigations in the R&D-innovation-productivity relationship within service industries. Innovation has been found to be a major contributor to productivity growth in manufacturing. In this paper, the importance of innovation is explored by comparing manufacturing and service firms in a sample of knowledge intensive industries. In particular, we intend to find evidence on the following two issues. First, is there any evidence that the reported weak rate of productivity growth in knowledge intensive services can be explained by a low propensity to be innovative? Second, is it possible that knowledge-intensive service firms are less efficient in deriving benefits from innovation than knowledge-intensive manufacturing firms? Empirical results based on innovation survey data indicate a surprising similarity in innovation performance between the two categories of firms.

Key words: Productivity – Innovation – Manufacturing – Services – Applied Econometrics

JEL Classification: C31, D24, L60, L80 O31

1 Introduction

The purpose of this paper is to bring a comparative perspective into the relationship between innovation and productivity in manufacturing and services. A large number

---

[*] The author gratefully acknowledges financial support from Vinnova, Swedish Agency for Innovation Systems. The author would like to thank Almas Heshmati, Gunnar Eliasson, Roger Svensson and one anonymous referee for very helpful comments and suggestions on an earlier version of this paper.
Correspondence to: Hans Lööf, Royal Institute of Technology, Centre of Excellence for Science and Innovation Studies, SE-10044 Stockholm, E-mail: hansl@infra.kth.se. Phone +46-8-790 80 12.

of studies have been done on the innovation-productivity link in the manufacturing sector. This study reports new results on this link by incorporating services into the analysis.

The uniqueness of this paper is that it is one of the first attempts to estimate a quite large number of firm level observations from both manufacturing and service firms based on an identical questionnaire as well as an identical modeling framework.

During the last two decades, the rate of growth in the manufacturing sector in Sweden has been about twice as high as the growth rate in the service sector. Finding a similar pattern in the U.S. economy, Triplett and Bosworth (1999) refer to the classical Baumol (1967) article and argue that, if the data are right, one might infer that the shift in the economy towards a larger share of services suggests slower growth in the aggregate productivity of the total economy.[1]

This paper considers two different samples of knowledge intensive firms, which represent some of the most dynamic parts of the Swedish economy when value added is considered. First, we have a group of manufacturing industries: pharmaceuticals; machinery and equipment; office machinery and computers; electrical and communication equipment; instruments and watches; and transport equipment. They are defined as knowledge-intensive according to their high degree of R&D intensity. The second category is business service firms the knowledge intensity of which is reflected in the high ratio of employees with a university degree to the total number of employees, as well as in a high ratio of R&D investments to sales.

Since the beginning of the 1980s, the average annual growth rate in value added has been an impressive 5 percent for both knowledge-intensive manufacturing and business services. The reported growth rate in labor productivity, however, differs considerably. While the average value added per working hour in knowledge-intensive manufacturing increased by 130 percent between 1980 and 1998, the corresponding rate of growth in business services was only 20 percent or about one percent per year.

One possible explanation for this large difference in growth rates is that it is much harder to increase productivity in service producing industries than in manufacturing, in line with the Baumol hypothesis. But the figures might also be incorrect. It is well documented in the literature that there are notorious problems measuring service production.[2]

In this paper, two in several aspects homogenous samples of manufacturing and service firms are compared. Both have domestic manufacturing firms as the main customers, and knowledge is a crucial production factor for competitiveness

---

[1] There is also a large degree of heterogeneity in the annual productivity growth rates in the manufacturing sector. The extremes of the 2-digit ISIC code levels in Sweden are, on the one the end, electrical and communication equipment, with more than 10% annual growth rates, and, on the other, office machinery computers, with a negative rate of growth. The service sector shows smaller disparities between individual industry growth rates, but there are rather substantial differences. For example between post and telecommunications on the one hand, and personal services and hotels and restaurants on the other.

[2] For instance, see Fuchs (1969), Mohr (1992), Griliches (1992), Griliches and Mairesse (1993), Sherwood (1994), Armknecht et al. (1992), Dean and Kunze (1992), Armknecht et al. (1997), Baily and Zitzewitz (2001), Heshmati (2003).

and survival. An important difference, however, is that while manufacturing firms operate on the global market, service firms produce for the national market.

Two issues related to the differences in productivity growth are explored in some detail. First, since innovations have been found to be a major contributor to productivity growth in manufacturing, is there any evidence that the average business service firm has a lower propensity than the average knowledge-intensive manufacturing firm to be innovative? Second, is it possible that the business service firms are less efficient in deriving benefits from innovation?

The empirical part is based on matched firm-level data collected from Statistics Sweden and observations from an innovation survey, conducted in 1999, based mainly on the Community Innovation Survey (CIS) - methodology.[3] The survey was among the first to utilize the internationally harmonized CIS questionnaire in order to receive information about both innovation input and innovation output from manufacturing firms, as well as from service firms. Moreover, in addition to the ordinary CIS questions on cooperation, information, strategy and obstacles when innovations are considered, this particular survey is also enlarged with information on product life cycles, growth rate on the markets, factors important for competitiveness of the products and categories of customers.

A central issue in the analysis is the choice of the methodological approach. A simultaneous system of equations is here estimated, based on a production function approach, where the relationship between innovation input, innovation output and productivity is modeled and estimated jointly.

The remainder of this paper is organized as follows: Sect. 2 briefly discusses some problems concerning the measurement of productivity. Section 3 describes the specific features of the data set. The empirical model is provided in Sect. 4. The estimation results are given in Sect. 5. Section 6 summarizes the main findings.

2 Measuring of productivity

In the two-sector growth model of Baumol (1967), the production functions in manufacturing and services are represented by a technologically advanced manufacturing sector and by a stagnant service sector. The model predicts that the stagnant service sector will, over time, absorb more and more workers to satisfy growth in demand because there are fewer opportunities to substitute technological advancements in machinery and equipment for labor.

The slow growth reported in services such as hotels and restaurants, large parts of the health care industry, transportation, education, and various parts of the public sector, is probably measured correctly. It might reflect slow technological change resulting from the labor-intensive nature of such industries, and the high income elasticity for the demand for such services is in line with Baumol's predictions.

However, the concept of service covers a heterogeneous set of industries, and recent years have seen a sharp growth in investment in new technology in finance, post and telecommunications, business services and wholesale and retail sales. Given the fact that business services are an industry in which output is measured

---

[3] For more information about the Community Innovation Survey, see OECD and Eurostat (1997).

with difficulties, it is likely that the reported weak productivity growth partly is a result of the notorious problems in measuring quantitative as well as qualitative output in service production.[4]

Quantitative estimates of the growth in real output have a general and pervasive Achilles' heel not only in service industries. To measure the real output, one needs to know their total revenue and have access to adequate information in order to construct an appropriate price index. Similarly, in order to measure inputs in real terms, one needs information on inputs used in production (total costs and prices or units used). In both cases, we have to deal with the problems of quality changes and the continuous emergence of new products and services, and the disappearance of old ones.[5] This is a problem facing large parts of the modern knowledge intensive production, regardless of whether they are manufacturing or services. However, these kinds of difficulties are reduced to a minimum since we only look at cross-sectional data.

3 Data and variables

The first two rounds of the Community Innovation Survey, launched by Eurostat and national agencies within the EU and some other countries during the 1990s, provided quantitative firm-level data on both innovation input and innovation output. However, the innovation output data is limited to manufacturing industries and there is a strong focus on technological innovations, which hampers a comparative analysis of the economic impact of innovations in both manufacturing and service industries.

In order to compare both categories of firms, Statistics Sweden launched a modified questionnaire of the CIS style in the 1999. The questionnaire was then sent out to both manufacturing and service firms. The questions are identical, which gives an opportunity to analyze differences and similarities in the link between innovation and productivity for the two industries. The survey data is further matched with register data adding economic variables for each of the observed firms.

The sample used in this study is drawn from a stratified random sample and consists of 607 knowledge-intensive manufacturing firms and 538 business service firms. To get as homogenous comparison samples as possible, censoring has eliminated the influence from extreme outliers.[6] Moreover, weighted factors are used for estimation, meaning that differences in the number of firms in a given strata and the

---

[4] The integration of computers in the economy has aggravated this problem. According to Triplett and Bosworth, the industries that are the largest purchasers of computers in the U.S. are all service industries – ranked by intensity of usage, financial services, wholesale trade, business services, insurance and communications. In these industries, computers have created new forms of service production that hardly is reflected in the statistics.

[5] The presence of a quality-based index, however, is by no means a guarantee of good quality in the measurement process. Nordhaus (1997) argues that the standard methodology of price indices has only captured a small amount of the revolutionary improvements in economic life from the massive changes in transportation, communications, lighting, heating and cooling and entertainment.

[6] Methods for censoring and treatment of missing variables are provided in Appendix A.

number of respondents in the survey is taken into account so that the observations represent the whole population of firms in a given size and class.

Tables 1-3 give some details of the sample used. It considers an overall sample of 1,145 observations of firms with 20 or more employees and a subsample of 526 innovative firms. In this paper, a firm is defined as innovative if it has both expenditures on innovation activities as well as positive revenues from innovations launched on the market during the last three years.[7]

Table 1 reports mean and standard deviations for some key variables of interest collected from firms' annual reports, educational statistics and the survey. Starting from Panel A, the overall sample shows that the average manufacturing firm is larger than the average service firm, 218 and 99 employees, respectively. The frequency of temporary hired employees is rather similar. The intensity of both human capital and R&D investment is highest among business services, while the average knowledge-intensive manufacturing firm proportionally invests more in physical capital (gross investment) and also has higher sales income per employee.

Innovation output measured as the ratio of innovation sales to total sales is about the same size for both sub-samples. Finally, innovation sales per employee is higher for the manufacturing sample, while the average business firm has higher level value added per employee (labor productivity).

The main difference between the overall samples in Panel A and the innovative samples in Panel B is that innovative firms have higher levels of sales, labor productivity and human capital.

The firm characteristics presented in Table 2 reveals some interesting similarities between the two samples. Surprisingly, about half or more of the firms report that the average life cycle (length of life) of the average product is 7 years or more. About two out of five firms in both groups consider quality to be of critical importance for competitiveness of their products. A majority of the firms in both knowledge-intensive manufacturing and business services carry out some kind of product innovation. Manufacturing firms are the main customers for both groups.

According to the definition used in this paper, 53 percent of the manufacturing and 40 percent of the service firms are innovative. In agreement with previous findings, e.g., Cohen and Levinthal (1990), Freeman (1991), Murray et al. (1996), Pyka (2000), a majority of both categories of innovative firms carry out innovation in

---

[7] It could be questioned whether or not the three-year period is too short to measure the economic results of innovations. Recent empirical research, however, supports the view of an ongoing process of shorter product life cycles, reduced time to market for new product development and an increasing share of sales deriving from new products. Choperena (1996) finds that many firms see cycle time acceleration as a way to increase their market competitiveness. Studying the time it takes from the basic idea to commercialization for 610 innovations in the Finnish economy, Palmberg et al. (2000) find that close to 50% of all innovations developed from the basic idea into commercialization in less than 2 years in both manufacturing and in services. The time from commercialization of the innovations to break even was also found to be surprisingly similar between services and manufacturing, within 2 years for about 70% of the innovations in both sectors. These findings are in sharp contrast to the results from similar studies carried out four decades ago. For example, Jewkes et al. (1958) identified fifty-one inventions and found that the mean lag between invention (the idea) and innovation (introduction of the idea to the market) was 12.5 years and the lag exceeded 20 years for 30 percent of these inventions. About two decades later, Ravenscraft and Scherer (1982) report a reduced lag structure of only 4-5 years from innovation input to innovation output.

Table 1. Descriptive Statistics: Major Variables

Means and Standard Deviations. Minimum and Maximum. Weighted Values

Panel A	Overall samples							
	Manufacturing N=607				Services N=538			
	Mean	SE	Min	Max	Mean	SE	Min	Max
Ordinary Employment	218	781	20	9,681	99	238	20	2,870
Temporary Employment (TE)	3	26	0	100	1	5	0	75
Proportion of firms with TE	0.19	0.39	0	1	0.16	0.37	0	1
Human capital: Engineers[1]	0.12	0.1	0	0.57	0.21	0.25	0	0.94
Human capital: Others[1]	0.02	0.03	0	0.33	0.12	0.15	0	0.69
Sales, 1000 Swedish Crowns[2]	1,312	764	100	73,000	1,079	964	113	10,000
Value added, 1000 Sw Crowns[2]	492	437	62	10,000	534	372	75	5355
Physical Capital Investment[3]	0.09	0.09	0	1.3	0.05	0.05	0	0.37
R&D-investment[3]	0.03	0.06	0	0.51	0.06	0.11	0	0.76
Innovation output[3]	0.16	0.24	0	1	0.14	0.25	0	1
Panel B	**Innovative samples[4]**							
	Manufacturing N=317				Services N=209			
	Mean	SE	Min	Max	Mean	SE	Min	Max
Ordinary Employment	230	756	20	9,681	102	269	20	2,870
Temporary Employment (TE)	5	35	0	600	2	6	0	59
Proportion of firms with TE	0.27	0.44	0	1	0.25	43	0	1
Human capital: Engineers[1]	0.14	0.11	0	1	0.22	0.25	0	1
Human capital: Others[1]	0.02	0.03	0	0.2	0.13	0.15	0	0.66
Sales, 1000 Swedish Crowns[2]	1,453	818	428	7,323	1,174	1,216	183	10,000
Value added, 1000 Sw Crowns[2]	513	229	69	2,242	569	508	124	5,355
Physical Capital Investment[3]	0.09	0.08	0.01	0.6	0.05	0.05	0	0.37
R&D-investment[3]	0.05	0.06	0.01	0.37	0.1	0.16	0.01	0.76
Innovation output[3]	0.29	0.26	0.01	1	0.32	0.29	0.01	1

Notes: (1) As a proportion of employment. (2) Per employee. (3) As a proportion of sales. (4) A firm is here defined as innovative if is has positive R&D and positive innovation output

cooperation with other partners such as universities, research institutes, customers, suppliers, consultancies, competitors and others. Moreover, about 50% of the innovative firms purchase professional services to support their internal innovative activities. A majority of firms innovating products also focus on organizational innovations and process innovations, indicating a close relationship between these forms of innovation.

The results presented in Table 3 indicate a large degree of similarity between the manufacturing firms and business firms when factors such as cooperation on innovation, strategy on innovation, sources of information for innovation and obstacles to innovation are considered. The only main difference is that the innovation partner of knowledge-intensive firms often is in the global arena, while business firms cooperate with regional or national partners of innovation.

Table 2. Descriptive Statistics: Firm Characteristics

Means and Standard Deviations. Overall Samples.

Firm characteristics	Manufacturing N=607		Services N=538	
	Mean	SE	Mean	SE
Product life cycle				
Less than 1 year	0.02	0.12	0.05	0.21
1-3 years	0.1	0.3	0.19	0.39
4-6 years	0.23	0.42	0.28	0.45
7-9 years	0.17	0.38	0.1	0.3
More than 9 years	0.48	0.50	0.38	0.46
Growth rate (t/t-2) on the firm's main market				
Strong	0.3	0.46	0.42	0.49
Weak	0.43	0.5	0.38	0.49
No growth rate	0.22	0.41	0.17	0.37
Negative growth rate	0.05	0.41	0.04	0.19
The firms' main customer				
Foreign customer	0.29	0.46	0.08	0.28
Domestic: Manufacturing firm	0.35	0.48	0.27	0.45
Domestic: Service firm	0.05	0.22	0.25	0.43
Domestic: Other firm	0.11	0.31	0.09	0.28
Domestic: Public sector	0.04	0.2	0.1	0.3
Domestic: Household	0.03	0.17	0.05	0.22
Domestic: Non profit organizations	0.01	0.09	0.04	0.2
Critical importance for competitiveness of the firm's products				
Quality	0.68	0.47	0.68	0.46
Content of knowledge	0.29	0.45	0.5	0.5
Price	0.31	0.44	0.2	0.4
Brand-name + Uniqueness + Design	0.3	0.46	0.27	0.44
Innovation activities				
Product (good or service) innovation	0.69	0.46	0.64	0.48
Process innovation	0.41	0.49	0.39	0.49
Organizational innovations	0.52	0.55	0.55	0.5
Cooperation on innovation (with external partner)	0.48	0.5	0.36	0.48
Purchase of professional service for innovation	0.44	0.5	0.33	0.47
"Innovative firms" defined as firms with positive innovation input and positive innovation output	0.53	0.49	0.4	0.49

Table 3. Descriptive Statistics: Firm Characteristics

Means and Standard Deviations. Innovation samples.

Firm characteristics	Manufacturing N=317		Services N=209	
	Mean	SE	Mean	SE
Strategy of innovation, critical importance				
Improving quality on existing products	0.44	0.47	0.67	0.47
Opening up new markets/increase market share	0.61	0.49	0.64	0.48
Extending product range	0.41	0.49	0.49	0.5
Reducing products being phased out	0.49	0.5	0.29	0.46
Very important sources of information for innovation				
Within the firm or within the group	0.8	0.4	0.74	0.43
Market (Customers, competitors, suppliers)	0.77	0.42	0.75	0.43
Other (Consultancies, universities, inventors, non-profit research institutes, patent disclosures, scientific literature or meetings, ICT and exhibitions)	0.29	0.46	0.36	0.48
Cooperation on Innovation				
Within the group	0.28	0.45	0.32	0.47
Market (Customers, competitors, suppliers).	0.61	0.49	0.63	0.48
Other (Consultancies, universities, investors and non-profit research institutes, patent).	0.24	0.43	0.27	0.44
Geographic localization of innovation partners (Here the group, or customers, competitors or suppliers.)				
Regional	0.28	0.45	0.43	0.5
National	0.46	0.5	0.52	0.5
Global	0.57	0.5	0.33	0.48
Obstacles to innovation				
Competence and knowledge problems	0.31	0.46	0.22	0.42
Economic and financial problems	0.26	0.44	0.22	0.42
Organizational problem	0.25	0.44	0.19	0.39
Market problem	0.11	0.31	0.13	0.33

Note: A firm is here defined as innovative if is has positive R&D and positive innovation output

4 The analytical framework

4.1 Formulation of the model

There are several difficulties in trying to establish a causal relationship between innovation input and innovation output, and also between innovation output and productivity in this study. One difficulty is simultaneity; most of the variables of interest in this study tend to change simultaneously, and it is not easy to untangle their separate effects. Another difficulty is sample selection. Sample selection arises when the observed sample is not randomly drawn from the population of interest.

Failure to take simultaneity and sample selection into account can potentially lead to inconsistent and biased estimates of the parameters of interest. In order to handle both these problems, this study relies on a simultaneity equation model as well as on a generalized Tobit model. The main idea behind this type of modeling is described by Crépon et al. (1998) and by Goux and Maurin (2000). Lööf and Heshmati (2002; 2004) describe the particular specification of the model used. We will briefly recapitulate some of the main features of the model here.

The theoretical framework for the model is a Cobb-Douglas production function with the main variables expressed as:

$$Q = AX^{\alpha}K^{\beta}e^{\lambda t+u} \tag{1}$$

where Q is the rate of productivity, A is a constant, X is a vector of conventional input variables such as labor, capital and so forth, K is a measure of technical knowledge, and u stands for all other unobserved determinants of productivity. α, β and λ are parameters that we are interested in estimating.

The empirical model is a structural model consisting of four equations and estimated in a multi-step procedure. In the first step, firms decide whether or not to engage in innovation activities (selection equation), and on the amount of money to invest in innovation. This is specified by a generalized Tobit model. Given the fact that the firm has decided to invest in innovative activities, the second step links the actual level of investment to its determinants. In the third and fourth steps, the generalized Cobb Douglas production function describes the effect of innovation investment on innovation output and innovation output on productivity in a simultaneous system. More specifically, the model is given by the following four equations:

$$y_{0i} = \begin{cases} 1 & \text{if} \quad y_{0i}^{*} = X_{0i}\beta_0 + \varepsilon_{0i} > 0 \\ 0 & \text{if} \quad y_{0i}^{*} = X_{0i}\beta_0 + \varepsilon_{0i} \leqslant 0 \end{cases} \tag{2}$$

$$y_{1i} = y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} \quad \text{if} \quad y_{0i} = 1 \tag{3}$$

$$y_{2i} = \alpha_{21}y_{1i} + \alpha_{23}y_{3i} + X_{2i}\beta_2 + \varepsilon_{2i} \quad \text{if} \quad y_{0i} = 1 \tag{4}$$

$$y_{3i} = \alpha_{32}y_{2i} + X_{3i}\beta_3 + \varepsilon_{3i} \quad \text{if} \quad y_{0i} = 1 \tag{5}$$

where y_{0i}^{*} is a latent innovation decision variable measuring the propensity to innovate; y_{0i} is the corresponding observed binary variable (being 1 for innovative firms); y_{1i}, y_{2i} and y_{3i} describe innovation input, innovation output and productivity; X_{0i}, X_{1i}, X_{2i} and X_{3i} are vectors of various variables explaining innovation decision, innovation input, innovation output and productivity. The β's and α's are the unknown parameter vectors. ε_{0i}, ε_{1i}, ε_{2i} and ε_{3i} are i.i.d. drawings from a multivariate normal distribution with zero mean, with the exception of two cases (equation (2) and (3), and equation (4) and (5)) not correlated.

For estimation purposes, we apply a two-step estimation procedure. In the first step, the generalized Tobit model, comprising the selection equation (2) and the innovation input equation (3), is consistently estimated by full maximum likelihood techniques, using observations on both innovative and non-innovative firms. The estimates of this first step are used to construct an estimate for the inverse Mills' ratio, which is incorporated as an explanatory variable in the estimation of both structural equations (4) and (5) to correct for potential selection bias. To take into account the possible problem that explanatory variables are determined jointly with the dependent variable, i.e. that they are not exogenously given, which highlights the simultaneity problem, the last two equations are estimated in a simultaneous equation system relying on the instrumental variable approach (2SLS and 3SLS respectively). For the estimation of this part of the model only, observations from the innovative firms are used.

4.2 Specification of the model

The specification of the model starts with the selection equation (2). As reported in the surveys by Cohen and Klepper (1996) and by Klette and Kortum (2002), size has been found to be a highly significant determinant of engagement in innovation activities. In addition to size, gross investment per capita is included. The selection equation also controls for product life cycle, other factors considered very important for the competitiveness of the firm's products, and whether the main customer is located abroad.

The size of innovation investment expenditure per employee (equation (3)) is explained by firm size, temporary employment, and by a number of indicator variables: obstacles to innovation; localization of innovation partners; critical sources of information for innovation; factors very important for the competitiveness of the firm's products.

In the innovation output (equation (4)), the most important explanatory variable is innovation input. The other continuous variables are capital intensity (proxied by gross investments per employee), labor productivity, and firm size. The discrete explanatory variables (indicator variables) are innovation strategy, information sources, cooperation partners, localization of innovation partners, product life cycle, and the growth rate of the firm's main market.

The final relationship is the productivity equation (5). Traditionally, the literature uses R&D as an independent variable. But thanks to an important novelty in the CIS style of data, we can use innovation output instead. In addition, we follow the literature and control for variations in firm size, physical capital, and human capital. Moreover, the productivity equation controls for process innovations and organizational innovations.

It is to be noted that, in all equations, the intensity variables are expressed in logarithmic terms and each of the four equations includes industry dummy variables.

5 The empirical results

The main hypothesis investigated in this study was whether there is any evidence for the notion that service industries have a lower propensity to be innovative or

Table 4. Correlation Analysis

Panel A
Innovative knowledge intensive manufacturing firms. Number of observations: 317. Logarithmic values.

	Labor productivity	Innovation output	Innovation input	Human capital Engineers	Human capital Other	Physical capital	Size Employees
Lab Product.	1						
Inn Output	.18***	1					
Inn Input	.17***	.26***	1				
Hum Cap E	.33***	.21***	.27***	1			
Hum Cap O	.29***	0.08	.31***	.27***	1		
Phys Cap	0.06	-0.08	0.11	-0.09	0.12	1	
Size	0.18	0.02	0.01	.25***	.16***	.34***	1

Note: Significant at the 1% level of significance(***)

Panel B
Innovative knowledge intensive service firms. Number of observations: 209 Logarithmic values.

	Labor productivity	Innovation output	Innovation input	Human capital Engineers	Human capital Other	Physical capital	Size Employees
Lab product.	1						
Inn Output	.42***	1					
Inn Input	.18***	.44***	1				
Hum Cap E	.21***	0.13	0.1	1			
Hum Cap O	0.11	.20***	0.15	-.38***	1		
Phys Cap	0.14	0.14	0.1	0.06	0.01	1	
Size	-0.11	-0.16	-0.13	-0.02	0.02	-0.04	1

Note: Significant at the 1% level of significance (***)

whether they are less efficient in deriving benefits from innovations. The correlation analysis presented in Table 4 indicates a close relation between R&D investments and innovation output, as well as between innovation output and productivity for both samples of firms.

5.1 Investigation of robustness of the model

In Table 5, regression results are presented which summarize, for a sample consisting both of knowledge intensive manufacturing firms and business service firms, the relationship between the probability of investment in R&D and labor productivity. Note that the full model contains all the variables described in the model specification. Only a few selected variables will be discussed here in order to conserve space.

Each of the four regressions includes industry dummy variables, allowing for separate industry intercepts. On the whole, the results are very encouraging. The estimated coefficients of firm size, R&D, innovation output, labor productivity, and human capital and physical capital (all in logs) are principally reasonable and highly significant. In equation (4), gross investment is used as a proxy for physical capital (stock) and the coefficient should be of an order of 0.1-0.2. However, the order of magnitude (0.03) is not quite what we expected and indicates that it is problematic to use a flow measure as a proxy for stock when physical capital is considered. The elasticity of innovation input with respect to R&D is quite large (0.4). In this case, using a flow of R&D investments as a measure of the R&D stock may overestimate the latter for large companies with a long R&D history (Griliches, 1998). Perhaps this is why the estimate is somewhat biased upwards.

The coefficients of innovation output (0.16) and human capital (0.5-0.7) in equation (4) are close to what has been reported in previous studies using a similar model (Crépon et al., 1998; Lööf and Heshmati, 2004).

The modern innovation literature (see for example Cohen and Klepper, 1996) finds that the probability of being engaged in innovation increases with firm size, but the R&D intensity is not higher for large firms than for small firms. The present model capture the latter fact and, in congruence with Janz et al. (2004), the sign is negative, meaning that smaller knowledge-intensive firms invest proportionally more in R&D per capita than larger knowledge-intensive firms. The non-significant size estimate in the probit equation can partly be explained by the fact that service firms are included in the sample.

Finally, the inverse Mills' ratio indicates the presence of a selectivity problem in step two of the model, and the method used to account for such effect by the selectivity equation is more efficient. The likelihood-ratio test (LR) indicates a necessary independence of the selection and input equation in the first of the two steps of the model.

5.2 Determinants of R&D

In Tables 6 and 7, the relationship between R&D[8] and its determinants is examined separately for both categories of firms. Starting with Table 6 and the selection equation, it is shown that the probability of doing R&D increases with size only for manufacturing firms (weakly significant). Another difference is that R&D engagement increases with the size of gross investment, but only for manufacturing firms. However, the inclusion of physical capital as an explanatory variable is a problem in this equation, since it might include R&D-embedded machinery and equipment

---

[8] In the CIS questionnaire, innovation input includes R&D and some other expenditures on innovation such as training related to R&D activities and market introduction of new goods. However, R&D is the vast major part of innovation expenditures.

Table 5. Summary Regressions: The Whole Model. Selected Variables.

Panel A:
Step 1 of the model; Generalized tobit. Overall sample of knowledge intensive manufacturing and business services. Number of observations: 1,145.

	Coeff	SE
Equation 1: Selectivity equation		
Dependent variable: Probability of R&D-investments		
Log firm size; employment	0.041	0.035
Equation 2: Research investment equation		
Dependent variable: Log innovation expenditures per employee		
Log firm size, employment	-0.247***	0.062
LR test of independent equations	25.56 (0.00)	

Panel B:
Step 2 of the model; Instrument variable method. 3SLS. Innovative sample of knowledge intensive manufacturing and business services. Number of observations: 526.

	Coeff	SE
Equation 3: Innovation output equation		
Dependent variable: Log innovation sales per employee	0.434***	0.122
Inverse Mills ratio	0.870***	0.2
Equation 4: Productivity equation		
Dependent variable: Log innovation expenditures per employee		
Log innovation sales per employee	0.162***	0.033
Human capital: engineers/total employment	0.473***	0.093
Human capital: others with a university degree/total employment	0.738***	0.157
Log gross investment per employee**	0.036**	0.015
R-square equation 3	0.27	
Root MSE equation 3	1.07	
R-square equation 4	0.17	
Root MSE equation 4	0.37	

Notes: Industry dummies and intercept are included in each regression. Significant at the 1% (***), 5% (**) and 10% (*) levels of significance.

and the exogeneity may therefore be questioned. Since a removal of this variable does not have any significant impact on equation (1) nor on the other regressions, it is used for comparability reasons. A third and not unexpected difference is that the likelihood of R&D-investment is significantly correlated with the main customer located abroad for the average manufacturing firm, but not for the average business service firm. Using product life cycle indicate a negative impact on R&D engagement for manufacturing and service firms. However, it is significant only for the former category of firms.

Table 6. Regression Results: Selection Equation.

Dependent variable: Probability of doing innovation.

	Manufacturing N=607		Services N=538	
	Coeff	SE	Coeff	SE
Firm Size: Ordinary employment[1]	0.074*	0.044	-0.002	0.058
Temporary employment[2]	0.520***	0.129	0.490***	0.139
Gross investment[1]	0.142***	0.046	0.041	0.041
The main customers are located abroad[2]	0.520***	0.104	0.177	0.168
Very important factors for the competitiveness of the products[2]				
Price	0.12	0.117	0.182	0.151
Quality	0.695***	0.127	0.797	0.155
Knowledge content	0.194	0.128	0.440***	0.136
Brand-name, design, uniqueness	0.575***	0.128	0.241*	0.135
Product life cycle[2]				
3 years or less	-0.272	0.666	0.05	0.171
4-7 years	-0.113	0.142	0.149	0.171
7-9 years: Reference	-	-	-	-
More than 9 years	-0.342**	0.137	-0.242	0.135
LR test[a]	36.67	0	9.06	0.002

Notes: Industry dummies and intercept are included in each regression. Significant at the 1% (***), 5% (**) and 10% (*) levels of significance. (1) Log variables. (2) Dummy variables. (a) LR test of independence of the selection equation (equation 2 reported in Table 6) and innovation input equation (reported in Table 7).

It should be noted that there are inconsistencies in the results obtained when the estimates for knowledge intensive manufacturing firms and business service firms are considered.

Table 7 provides more details on R&D determinants, showing the results of the elasticity of log R&D per employee. The coefficient for firm size is negative and highly significant for both industries. When looking at the log of temporary employment, a significant estimate is revealed (at the 5% level) in the order of a magnitude of 0.17 for manufacturing firms. Outsourcing is usually defined as the replacement of internally provided activities in the production process for externally produced goods and services. Concentration on core business, core technologies and specialized competence are commonly referred to as main causes of outsourcing. The hiring of workers in non-traditional jobs can also be defined as a kind of outsourcing.[9] Assuming that this variable can be interpreted as a proxy for total outsourcing, the regression results in both equation (1) and equation (2) indicate

---

[9] Falk and Koebel (2000) include temporary personnel services in their definition of outsourcing, and, using the same definition, Estevão and Lach (1999) identify several distinguished reasons for hiring ”temporary workers”. Primarily, this is one important aspect of the general trend toward flexible and specialized work arrangements by firms. They find that this kind of outsourcing is motivated by: (1) the potential for implementing a two-tier wage structure by contracting with firms that pay lower wages; (2) the possible realization of scale economies as a result of specialization in the provision of specific tasks; (3) the potentially higher productivity of specialized temporary workers, in relation to what could

Table 7. Regression Results: Innovation Input Equation.

Dependent variable: Logarithm of innovation expenditure per employee.

	Manufacturing N=317		Services N=209	
	Coeff	SE	Coeff	SE
Firm Size: Ordinary empoyment[1]	-0.148**	0.078	-0.408***	0.122
Temporary employment[1]	0.169**	0.072	-0.051	0.136
Obstacles to Innovation[2]				
Cost/Financing issues	0.309*	0.158	-0.025	0.265
Demand/Market issues	-0.138	0.22	0.161	0.307
Organizational issues	0.083	0.164	0.415	0.282
Lack of Skill/Knowledge	0.221	0.151	0.178	0.266
Localization of partners on innovation[2]				
Regional collaboration	-0.023	0.162	-0.146	0.223
National collaboration	-0.289*	0.158	-0.175	0.229
Global collaboration	0.422***	0.152	0.924***	0.226
Very important sources of information for innovation[2]				
Within the firm or the group	-0.084	0.165	0.277	0.298
Customers, suppliers, competitors	0.478***	0.15	0.332	0.221
Universities, consultancies etc.	-0.181	0.165	0.065	0.207
Very important factors for the competitiveness of the products[2]				
Price	-0.178	0.172	-0.616**	0.296
Quality	-0.553***	0.21	-1.148***	0.365
Content of knowledge	0.296	0.181	-0.309	0.272
Brand name, uniqueness, design	-0.343*	0.179	0.051	0.254

Notes: Industry dummies and intercept are included in each regression. Significant at the 1% (***), 5% (**) and 10% (*) levels of significance. (1)Log variables. (2) Dummy variables.

that outsourcing strategy has a positive impact on R&D in the average knowledge-intensive manufacturing firm. This confirms previous finds by for example Ulset (1996) and Deavers (1997). This variable is not significant for the service sample.

The dummy variable for global innovation partner is highly significant for both samples. Among manufacturing firms, the R&D intensity increases with "market" (customer, suppliers and competitors) as a crucial source of information for innovation.

Somewhat unexpectedly, the indicator variable for "quality" as a very important factor for the competitiveness of the product is negative and highly significant for manufacturing as well as for services. A tentative interpretation is that this variable captures marginal improvements rather than radical innovations.

---

have been gained if the firm had to learn and build up related knowledge internally; (4) and the ability to adjust the level of employment rapidly in response to temporary and/or uncertain changes in demand

Table 8. Regression Results: Innovation Output Equation.

Dependent variable: Logarithm of innovation sales per employee.

	Manufacturing N=317				Services N=209			
	3SLS		2SLS		3SLS		2SLS	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Innov. Input[1]	0.451***	0.141	0.506***	0.147	0.432**	0.198	0.571*	0.29
Labor Prod[1]	1.101**	0.48	1.145**	0.495	1.252*	0.693	1.318	0.841
Mills Ratio[1]	-0.793***	0.237	-0.649***	0.275	-1.226***	0.397	-1.116***	0.537
Gross invest[1]	-0.190***	0.074	-0.168*	0.087	-0.043	0.067	-0.011	0.076
Firm size[1]	-0.03	0.128	-0.014	0.059	-0.005	0.128	0.058	0.157
Innovation strategy[2]								
Replace prod	0.253**	0.105	0.275**	0.137	0.381**	0.167	0.23	0.204
Improve qual	-0.009	0.103	0.047	0.14	0.201	0.222	0.551*	0.281
Extend range	-0.294***	0.108	-0.219	0.137	-0.092	0.146	-0.145	0.189
Open up mark	0.069	0.106	0.181	0.145	0.011	0.149	0.012	0.213
Very important sources of information. for innovation[2]								
Firm/Group	0.196	0.108	0.308*	0.162	-0.429**	0.172	-0.591***	0.22
Market	-0.479***	0.143	-0.495***	0.182	0.073	0.167	0.1473	0.233
Other	-0.019	0.112	0.102	0.153	0.145	0.185	-0.03	0.229
Very important cooperation partners on innovation[2]								
The group	-0.289***	0.129	-0.277**	0.138	0.223	0.185	0.245	0.238
Market[3]	-0.047	0.121	-0.083	0.131	0.076	0.167	0.086	0.205
Other	-0.402***	0.137	-0.414**	0.154	-0.126	0.181	-0.059	0.206
Localization of partners on innovation[2]								
Regional	-0.186	0.121	-0.217	0.153	0.167	0.161	0.173	0.228
National	0.332***	0.12	0.398***	0.159	-0.123	0.185	-0.049	0.239
Global	-0.306**	0.12	-0.386***	0.172	-0.416	0.268	-0.657	0.376
Product life cycle[2]								
<= 3 years	0.658***	0.209	0.660***	0.211	0.196	0.289	0.048	0.368
4-7 years	0.078	0.142	0.144	0.169	0.078	0.243	-0.25	0.287
7-9 years[a]	-	-	-	-	-	-	-	-
> 9 years	0.023	0.138	-0.002	0.171	0.158	0.423	-0.326	0.327
Growth rate on the firms' main market[2]								
High	0.069	0.123	-0.084	0.158	0.275*	0.16	0.36	0.228
Low	-	-	-	-	-	-	-	-
No growth	-0.263**	0.131	-0.157	0.186	0.059	0.23	0.265	0.278
R-Square	0.25		0.26		0.39		0.41	
Root MSE	0.99		1.03		1.05		1.11	

Notes: Industry dummies and intercept are included in each regression. Significant at the 1% (***), 5% (**) and 10% (*) levels of significance. (1) Log variable. Intensity terms; per employee. (2) Dummy variable (3) Market: Customers, suppliers and competitors. Consultancies, universities, inventors, non-profit research institutes, patent disclosures, scientific literature or meetings, ICT and exhibitions (4) Other: Universities, consultancies, non-profit research institutes, inventors. (a) Reference.

5.3 Innovation and productivity

Table 7 and 8 present 2SLS and 3SLS estimates of the simultaneous equations model for both samples. The coefficient estimates for the two alternative estimation methods are rather close. Of more interest, the results for the key variables are roughly similar for manufacturing and services. The elasticity of labor productivity with respect to innovation output, and the elasticity of innovation output with respect to innovation input, are somewhat lower for manufacturing firms. But the differences are not statistically significant.

The estimates for the innovation output regression are presented in Table 7. When looking at the main variables, it shows nearly identical R&D elasticities in the 3SLS regressions, 0.45 versus 0.43. However, the slightly higher coefficient for manufacturing is somewhat more efficient and significant at the 1% level, while the service estimate is significant at the 5% level of significance. The coefficients indicate that a 10 percent increase in innovation expenditures per employee at the margin raises innovation sales per employee by 4 percent. When using the 2SLS, it results in coefficient estimates of 0.50 for the average manufacturing firm and 0.57 for the average service firms. The latter is only weakly significant.

The model allows for feedback effects of productivity on innovation output. The order of magnitude of the resulting estimate is about 1.0, but significant only in the manufacturing sample. Gross investment, the proxy for physical capital, correlates negatively with innovation output for manufacturing firms and is neutral when service firms are considered. There is no indication of a significantly increased or decreased research productivity with respect to firm size for knowledge intensive firms. The Mills' ratio, correcting for possible selection bias, is highly significant and the size of the coefficient is larger in the service sample.

The indicator variables are mostly insignificant in the service sample with the exception of innovation output, which increases with the replacement of products phased out as an important cause of R&D investments, and decreases with the firm or the group as a very important source of information for innovation. The regression results for the manufacturing sample show that the R&D productivity increases with short product life cycles and national innovation partners. On the contrary, the sign is negative for (i) bad growth rate on the main markets, (ii) global innovation partners, (iii) the own group, (iv) a bunch consisting of consultancies, universities, inventors, and non-profit research institutes as a very important cooperation partner for innovation, (v) the market as a very important source of information, and (vi) an extension of the product range as innovation strategy. Due to limited space, a detailed analysis of these estimates is not possible. The conclusion is that, despite differences among the coefficients for various indicator variables, there is a striking similarity between knowledge intensive manufacturing and service firms when the link between R&D and innovation output is considered.

The relationship between productivity and its determinants is investigated for the two samples in Table 9. The elasticity of log labor productivity with respect to log innovation output per employee is surprisingly similar for both categories of firms. The order of magnitude is about 0.11 among manufacturing firms and about 0.14 among service firms in the 3SLS regressions. Both estimates are highly

Table 9. Regression Results: Productivity Equation.

Dependent variable: Logarithm of value added per employee.

	Manufacturing N=317				Services N=209			
	3SLS		2SLS		3SLS		2SLS	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Innov. Output[1]	0.110***	0.038	0.116***	0.037	0.141***	0.048	0.150***	0.053
Hum Cap Eng[1]	0.793***	0.19	0.704***	0.239	0.297*	0.157	0.327**	0.138
Hum Cap Oth[1]	3.115***	0.692	3.321***	1.034	0.584***	0.217	0.440**	0.222
Gross invest	0.028	0.022	0.032	0.029	0.049**	0.023	0.05	0.036
Firm Size[1]	0.015	0.017	0.02	0.015	-0.011	0.037	-0.003	0.027
Process Innov[2]	0.026	0.034	-0.029	0.04	-0.079	0.059	-0.170**	0.072
Org. Innov[2]	-0.001	0.033	-0.024	0.041	-0.120**	0.064	-0.101	0.099
Very important cooperation partners on innovation[2]								
The group	0.056	0.04	0.051	0.044	0.073	0.064	0.092	0.104
Market[3]	0.015	0.036	0.015	0.038	-0.037	0.061	-0.049	0.073
Universities etc[4]	0.077*	0.046	0.084*	0.05	0.019	0.065	0.019	0.072
R-Square	0.21		0.21		0.27		0.28	
Root MSE	0.31		0.32		0.39		0.42	

Notes: Industry dummies and intercept are included in each regression. Significant at the 1% (***), 5% (**) and 10% (*) levels of significance. (1) Log variable. Intensity terms; per employee. (2) Dummy variable (3) Market: Customers, suppliers and competitors. (4) Other: Universities, consultancies, non profit research institutes, inventors.

significant. Using 2SLS as a check on the robustness of the results only resulted in minor changes in estimates, 0.12 versus 0.15. In both samples, labor productivity increases quite extensively with human capital, but not with firm size. The sign for gross investment as a proxy for physical capital is expected, but the coefficient estimate is low and not significant for the average manufacturing industry.

To conclude, we find only minor differences between knowledge-intensive manufacturing and knowledge-intensive services when the importance of innovation is considered. However, two major issues are comparability of innovation output between manufacturing and services, and the lag structure within both industries. In the model used, an underlying assumption is than an innovation is an innovation, irrespective of whether it is created in a service firm or in a manufacturing firm. Concerning lags, the main drawback with the R&D variable used is that it is a flow variable and observed only in the same year as we observe innovation output. This means that the lag between investment in research and the actual product innovation is ignored, and so is the lag between product innovation and introduction to the market. The underlying assumption here is that firms' R&D behaviors are constant over time.

6 Conclusions

This paper can be viewed as another complementary component in the chain of a rather limited number of investigations on the R&D-innovation-productivity relationship within service industries. Two main research issues were addressed here. First, since innovations have been found to be a major contributor to productivity growth in manufacturing, we asked whether there was any evidence that the average service firm has a lower propensity than the average manufacturing firm to be innovative. Second, we asked whether it was possible that the service firms were less efficient than manufacturing firms in deriving benefits from innovation?

The data set used for the analysis is taken from the largest innovation survey ever conducted in Sweden. The questionnaire launched by Statistics Sweden was answered by 50 percent of all manufacturing and service firms that existed in Sweden in 1998 and 20 or more employees. From this overall sample, this study is based on subsample of 1,145 knowledge intensive manufacturing and service firms. The motivation for this selection was an assumption of rather homogenous and comparable firms. This was also confirmed by the descriptive statistics.

In spite of the various limitations and difficulties faced, this paper finds a rather consistent positive relationship between R&D, innovation, and productivity for both knowledge-intensive manufacturing firms and business service firms. The R&D investments as a share of sales or per employee are higher in the average service firms, it is true, but using a Cobb-Douglas type production function, the estimates indicate an elasticity of innovation output with respect to innovation input of about 0.4 for both samples of firms. Considering the link between innovation and labor productivity, the regression results show highly significant estimates in the order of magnitude of about 0.10-0.15 for both industries. The manufacturing estimate is consistent with previous results in the literature, although there are few studies so far that have presented quantitative productivity analysis in the service sector based on rather large microeconomic datasets on innovation activities.

Given that the estimation results, the validity of the data, and the model used are not questioned, and assuming that the sample used are representative for the populations studied, will suggest that R&D and innovation contributes very little to the reported large differences in productivity growth between knowledge intensive manufacturing firms and knowledge intensive service firms.

Table A1. Treatment of Missing Values and Outliers

Panel A: Register data. Number of observations 1,145

Variable	Missing values	Treatment	Outliers	Treatment
Sales	0	-	3	Censoring
Value added	0		1	Censoring
Employment	0	-	0	-
Gross investment	2	MV=0	13	Censoring
Human capital	0	-	0	-

Panel B: Survey data. Number of observations 1,145

Variable	Missing values	Treatment	Outliers	Treatment
Temporary employment	0	-	0	-
Export	646	MV=0	0	-
R&D	174	MV=0	3	Censoring
Innovation output	172	MV=0	0	-
Other	Variation	Imputation	-	-

Notes: The observations on export contain an unsatisfactory amount of missing values. One alternative here might be to impute values. However the method used is to replace the missing values with a zero value. The missing values on R&D and innovation are assumed to be zero. The numbers of outliers are fairly low and a sensitivity analysis shows that the influence of outliers is neglectable.

Table A2. Productivity Performance, Knowledge Intensive Manufacturing and Services

Level and growth of productivity at industry level. Swedish Crowns and percent

Industry	Value added per working hour in the year 1991 prices.		Annual growth rate 1980-1998. Percent		
	1980	1998	Labor productivity	Value added	Hours
Knowledge-intensive manufacturing	**142**	**325**	**4.7**	**4.5**	**-0.2**
Electrical and communication equipment	84	577	11.3	10.8	-0.4
Pharmaceuticals, soap, perfumes	301	641	4.3	4.5	2.3
Instruments	190	331	3.1	5.7	2.5
Transport equipment	164	283	3.1	2.6	-0.4
Machinery & equipment	136	197	2	1.6	-0.5
Office machinery & computers	158	142	-0.6	-3.3	-2.6
Business services	**189**	**227**	**1**	**4.8**	**3.8**

Notes: The official statistics report that annual labor productivity has increased with about 5 percent for the aggregate of knowledge-intensive manufacturing studied in this paper during the two recent decades. This is in sharp contrast to the reported one percent change in labor productivity for business services. One possible explanation for this large difference in growth rates is that it is much harder to increase productivity in service producing industries than in manufacturing. But the figures might also be non correct. It is well documented in the literature that there are notorious problem measuring service production. Source: Statistics Sweden.

References

Armknecht PA, Lane WF, Ginsburg DH (1992) Improvements in Measuring Price Changes in Consumer Services: Past Present and Future. In: Griliches Z (ed.) Output Measurement in the service sector, 109–156, Studies in Income and Wealth, NBER and University of Chicago Press

Armknecht PA, Lane WF, Stewart KJ (1997) New products and the U.S. Consumer Price Index. In: Breshanan TF, Gordon RJ (eds.) The Economics of New Goods, vol. 58 of *Studies in Income and Wealth*, 375–391, NBER and The University of Chicago press

Baily MN, Zitzewitz E (2001) Service Sector Productivity Comparisons: Lessons for Measurements. In: Hulten CR, Dean ER, Harper MJ (eds.) New Development in Productivity Analysis, vol. 58 of *Studies in Income and Wealth*, 419–455, NBER and The University of Chicago press

Baumol WJ (1967) Macroeconomics of Unbalanced growth: The Anatomy of Urban Crises. American Economic Review, 57: 415–26

Choperena AM (1996) Fast cycle time-driver of innovation and quality. Research-Technology Management, 36–40

Cohen WM, Klepper S (1996) A reprise of size and R&D. The Economic Journal, 106: 925–951

Cohen WM, Levinthal DA (1990) Absorptive Capacity. A New Perspective on Learning and Innovation. Administrative Science Quarterly, 35: 128–152

Crépon B, Duguet E, Mairesse J (1998) Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level. NBER Working Paper No 6696

Dean ER, Kunze K (1992) Productivity Measurement in Service industries. In: Griliches Z (ed.) Output Measurement in the service sector, vol. 56 of *Studies in Income and Wealth*, 73–106, NBER and University of Chicago Press

Deavers KL (1997) Outsourcing: a corporate competitiveness strategy, not a search for low wages. Journal of Labor Research, 18, 4: 503–519

Estevão M, Lach S (1999) Measuring temporary labor outsourcing in U.S. manufacturing. NBER Working paper No 7421

Falk M, Koebel B (2000) Outsourcing of Services, Imported Materials and the Demand for Heterogeneous Labor: An Application of a Generalized Box-Cox Function. Discussion 00-51, ZEW

Freeman C (1991) Networks of innovators. A synthesis of research issues. Research Policy, 20: 499–514

Fuchs VR (1969) Introduction and Summary. In: Fuchs VR (ed.) Productivity in the Service Industries, 1–13, Conference on Research in Income and Wealth, NBER and Columbia University Press

Goux D, Maurin E (2000) Returns to firm-provided training: evidence from French worker-firm matched data. Labor Economics, 7

Griliches Z (1992) Introduction. In: Griliches Z (ed.) Output Measurement in the service sector, vol. 56 of *Studies in Income and Wealth*, 1–22, NBER and University of Chicago Press

Griliches Z (1998) Returns to R&D Expenditures in the Private Sector, Chapter 3. In: R&D and Productivity: The Econometric Evidence, Chicago

Griliches Z, Mairesse J (1993) Introduction. The Journal of Productivity Analysis, 4: 5–8

Heshmati A (2003) Productivity Growth, Efficiency and Outsourcing in Manufacturing and Services. Journal of Economic Surveys, 17, 1: 79–112

Janz N, Lööf H, Peters B (2004) Firm Level Innovation and Productivity - Is thedre a Common Story Across Countries? Problems & Perspectives in Management, Forthcoming 2004

Jewkes J, Sawers D, Stillerman R (1958) The Sources of Invention. St. Martin's Press, New York

Klette TJ, Kortum S (2002) Innovating Firms and Aggregate Innovation. NBER Working Paper No 8819

Lööf H, Heshmati A (2002) Knowledge Capital and Performance Heterogeneity: An Innovation Study at Firm Level. International Journal of Production Economics, 76, 1: 61–85

Lööf H, Heshmati A (2004) On the Relationship between Innovation and Performance: A Sensitivity Analysis. Economics of Innovation and New Technology, forthcoming

Mohr MF (1992) Recent and Planned Improvements in the Measurement of Deflation of Services Outputs and Inputs in BEA's Gross Product Originating Estimates. In: Griliches Z (ed.) Output Measurement in the service sector, 25–72, NBER and University of Chicago Press

Murray R, Millson S, Ray P, Wilemon D (1996) Strategic partnering for developing new products. Research-Technology Management, 41–49

Nordhaus DN (1997) Do Real-Output and real-Wage Measures Capture Reality? The History of Lightning Suggest Not? In: Bresnahan TF, Gordon RJ (eds.) The Economics of New Goods, vol. 58 of *NBER Studies in Income and Wealth*, The University of Chicago Press, Chicago

OECD and Eurostat (1997) Oslo-Manual, Proposed Guidelines for Collecting and Interpreting Technological Innovation Data. Organisation for Economic Co-operation and Development, Paris

Palmberg C, Niininen P, Toivanen H, Wahlberg T (2000) Industrial innovation in Finland, First results of the Sfinno-project. VTT , Group for technology studies, Working Papers No.47/00

Pyka A (2000) Informal networking and industrial life cycles. Technovation, 20: 25–35

Ravenscraft DJ, Scherer FM (1982) The lag structure of returns to research and development. Applied Economics, 14: 603–620

Sherwood MK (1994) Difficulties in the measurement of service outputs. Monthly Labor Review

Triplett J, Bosworth B (1999) Productivity in the Service Sector. Unpublished manuscript

Ulset S (1996) R&D outsourcing and central governance: an empirical study of commercial R&D projects. Journal of Economic Behavior & Organization, 30, 1: 63–82

Tracing empirical trails of Schumpeterian development

Michael Peneder

Austrian Institute of Economic Research (WIFO)
(e-mail: michael.peneder@wifo.ac.at)

Abstract. Schumpeterian development is characterized by the simultaneous interplay of growth and qualitative transformations of the economic system. At the sectoral level, such qualitative transformations become manifest as variations in the sectoral composition of production. Following the implementation of Harberger's method of visualizing the impact of differential productivity growth, dynamic panel estimations are applied to a standard growth model modified to include specific structural variables for both the manufacturing and the services sectors. Covering 28 countries over the period between 1990 and 2000, the results give empirical substance to the evolutionary emphasis on Schumpeterian development as opposed to mere aggregate growth.

Key words: Structural change – economic growth – Schumpeterian development – evolutionary economics

JEL Classification: O11, O30, O41

> "As a matter of fact, capitalist economy is not and cannot be stationary. Nor is it merely expanding in a steady manner. It is incessantly being revolutionized from within by new enterprise, i.e. by the intrusion of new commodities or new methods of production or new commercial opportunities into the industrial structure as it exists at any moment" (Schumpeter, 1942/50, p. 31).

1 Introduction

A major source of distinction between neoclassical and evolutionary economics is that the latter replaces conventional emphasis on mere growth with the broader

---

Correspondence to: Austrian Institute of Economic Research (WIFO); P.O. Box 91; A-1030 Vienna, Austria, Tel.: + 43-1-798 26 01 - 480; Fax.: + 43-1-798 93 86

concept of *Schumpeterian development*, which is characterized by the simultaneous processes of growth and qualitative transformations in the economy (Schumpeter, 1911/34).[1] Qualitative transformations arise as a result of the continual tension between the disequilibrating force of entrepreneurial 'creative response' and the equilibrating tendencies in terms of 'adaptive response' to such disturbances (Schumpeter, 1947).[2] Although the central idea of endogenous technological change through creative destruction also resides prominently in the latest generation of endogenous *Schumpeterian growth models*,[3] it is fair to say that these remain subservient to the restricted framework of steady state equilibrium analysis and thus bypass the more complex evolutionary dynamics of Schumpeterian development.[4]

Conversely, evolutionary analysis in the tradition of Nelson and Winter (1982) characterizes the market economy as a process of continuous change and qualitative transformation. Its dynamics are also driven by innovation; but "[i]nnovation is a matter of differential behaviour and differential behaviour is the basis for structural change" (Metcalfe, 1998, p. 37). As the fundamental diversity of micro-behavior involves dynamics that are much richer than those in steady state growth models, simple aggregation cannot do away with the fact that the potential paths of development are various and depend on the idiosyncratic characteristics of an economy, including among other elements its sectoral composition of production. From the evolutionary perspective, structural change is therefore an inevitable companion of growth, and the notion of Schumpeterian development the more relevant description of aggregate dynamics.

The importance of structural change to economic growth might prove to be the ultimate empirical test of the general relevance of the evolutionary agenda. Due to omnipresent diversity in firm behavior and performance, most sceptics would accept that an evolutionary view offers a more realistic description and explanation of micro-economic processes. But, based on the claim that such variations only represent unsystematic noise, tied down sufficiently by the forces of gravitation towards general equilibrium, they can still insist on its overall irrelevance to the analysis of aggregate phenomena. If we are not able to demonstrate the importance of evolutionary change to aggregate development, the neo-classical reduction by

---

[1] Building upon the classic contribution of Nelson and Winter (1982), recent examples of evolutionary approaches to growth and development are presented e.g. in Silverberg (1988), Silverberg and Verspagen (1997), Metcalfe (2001), and Montobbio (2002). Fagerberg (1994) and Verspagen (2001) survey the empirical evidence of technological change and growth from an evolutionary perspective.

[2] Our interpretation of Schumpeterian development thus encompasses both radical and incremental changes.

[3] The Schumpeterian kind of endogenous growth models are typically characterised by innovation races, a replacement mechanism and temporary monopoly power. Examples are the models in Grossman and Helpman (1991), Aghion and Howitt (1998) or Dinopoulos and Thompson (1998). Cheng and Dinopoulos (1996) present a rare example of a multisectoral endogenous growth model, where, depending on the choice of assumptions, the steady-state equilibrium can be replaced by deterministic cycles.

[4] For example, Aghion and Howitt (1998) are very precise about the limitations of steady state analysis: "The economy is always a scaled up version of what it was years ago, and no matter how far it has developed already the prospects for future developments are always a scaled-up version of what they were years ago" (Aghion and Howitt, 1998, p. 65).

means of assuming away diversity at the micro- and meso-levels can claim to be sufficiently accurate.[5]

This paper will focus on the presence of variation in industrial structure as one such example of qualitative transformations and examine its impact on aggregate development.[6] Sequentially permeating different layers of observation, Sect. 2 opens with a simple illustration of the correlation in time between structural change and aggregate income, followed by the more demanding visualization of Harberger's decomposition of aggregate productivity growth. Section 3 turns to econometric estimations of a macro-panel, where a standard growth model is augmented by structural variables. Section 4 provides a brief summary and concludes.

2 Visual inspections

2.1 Readily observable trails

To begin with a rather simple visualization, the lines in Fig. 1 depict developments in real GDP per capita on the vertical axis, in combination with the systematic shifts in the sectoral composition of production on the horizontal axis. Structural change is illustrated by the relative shares of particularly technology-driven manufacturing industries, classified in Peneder (2002), and the group of business related services taken from the OECD services statistics. The three marks on each line in Fig. 1 refer to the years 1985, 1992 and 1999, the two marks in Fig. 2 to 1992 and 1999. Since GDP p.c. grew in all European countries, the upper marks always indicate the later years.

In short, the figures reveal that the development of aggregate levels of income and structural change cannot be considered as independent processes. At the very least, we are able to observe co-movement over time, whereby GDP per capita and the shares of both industry types tend to increase. What makes the two figures particularly appealing in our context is that they encompass the typical elements of heterogeneity and selection of the evolutionary process mentioned above. First of all, even among EU countries, there is wide variety in the paths of development, with much variation in both industrial structure and GDP per capita. Second, since income levels, as well as industrial structure, are highly persistent, movements along the vertical and horizontal axes can easily be recognized as path dependent.

---

[5] See, for instance, Kongsamut et al. (2001) or Meckl (2002). Both papers propose a "generalized balanced growth path", where – despite the presence of differential growth at the level of disaggregated sectors – some very specific knife-edge conditions produce steady state constant growth rates in the aggregate. As a general intuition, structural change does not affect an economy's resource constraints in such models. Despite the very particular assumptions presented in the paper, Meckl (2002, p. 244) is very straightforward about his conclusion: "Our analysis indicates that as long as we are only interested in the behavior of aggregate variables, there is simply no need to disaggregate". Opposite conclusions were drawn by Echevarria (1997), who assumed differential productivity growth between sectors and presented according simulations with up to two percentage points of variation in growth rates explained by the sectoral composition of production.

[6] Recent and complementary empirical evidence for a significant link between trade specialisation in the manufacturing sector and economic growth are provided by Plumper and Graff (2001), Lewer and Van den Berg (2003), or Peneder (2003).

Third, despite the various directions taken by individual paths of development, their long run course exhibits a common (non-random) orientation. Countries are heading towards the north-eastern segments of the map, which are characterized by higher shares of technology-driven manufacturing and business-related service industries.

2.2 Harberger's decomposition: "mushrooms" vs. "yeasts"

At a more refined level, Arnold C. Harberger 1998 offers an attractive visualization of what he calls the "yeast versus mushroom" issue: "The analogy with yeast and mushrooms comes from the fact that yeast causes bread to expand very evenly, like a balloon being filled with air, while mushrooms have the habit of popping up, almost overnight, in a fashion that is not easy to predict" (Harberger, 1998, p. 4). Each analogy illustrates a different mechanism in the meso-macro link of productivity growth. The yeast analogy corresponds to a vision of the growth process driven by economies of scale and broad externalities applicable to the entire economy. Conversely, the mushroom analogy refers to advances in productivity "stemming from 1001 different causes" (p. 5) and appearing in irregular, often clustered patterns, which are more prevalent in some industries than in others. This implies that, within specific periods of time, productivity growth would be highly concentrated in relatively few industries, while over time, the clustered appearance of productivity growth might also shift between different branches of production. Based on his empirical results, Harberger argues in favor of the 'mushroom' analogy and ultimately relates his vision of the growth process directly to Schumpeter and his idea of 'creative destruction'.

Similar illustrations can be based upon data for manufacturing industries in the EU member countries, plus the USA. Due to the disaggregated breakdown of NACE 3-digit industries and the according data limitations, we refer only to labor productivity. This is in contrast to Harberger, who based his work on total factor productivity (TFP). Value added is measured at constant 1995 prices, using the industry-specific deflators presented in Egger and Pfaffermayr (2001).

The cumulated shares of each of i industries in the total value added of manufacturing in the base year (by) are indicated on the horizontal axis:

$$CS(VA_{i,by}) = \frac{\sum_{c=1}^{i} VA_{c,by}}{VA_{tm,by}} \tag{1}$$

CS = cumulated shares; VA = value added; tm = total manufacturing

The cumulated contributions of each industry i to the changes in aggregate labor productivity between the final year (fy) and the base year (by) of the observation period are indicated on the vertical axis:

$$C\Delta(LP_i) = \frac{\sum_{c=1}^{i} (LP_{c,fy} * \frac{L_{i,fy}}{L_{tm,fy}} - LP_{c,by} \frac{L_{i,by}}{L_{tm,by}})}{LP_{tm,fy} - LP_{tm,by}} \tag{2}$$

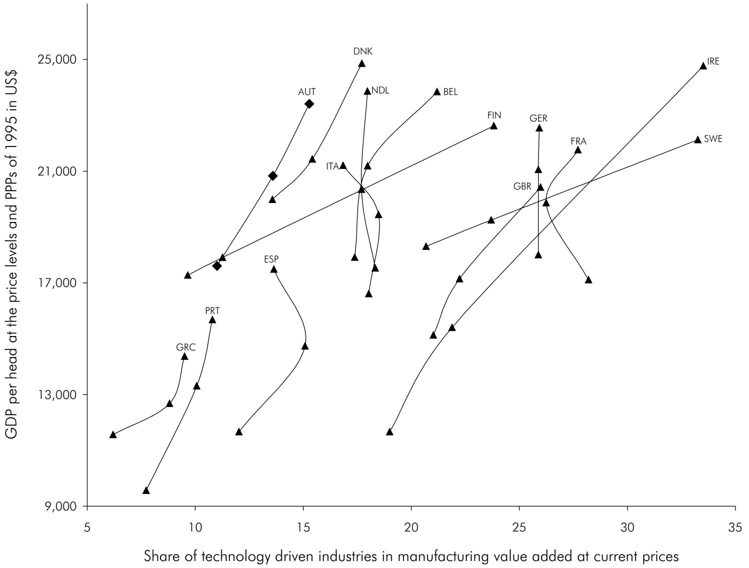


Fig. 1. GDP and the value added share of technology-driven industries: 1985/92/99

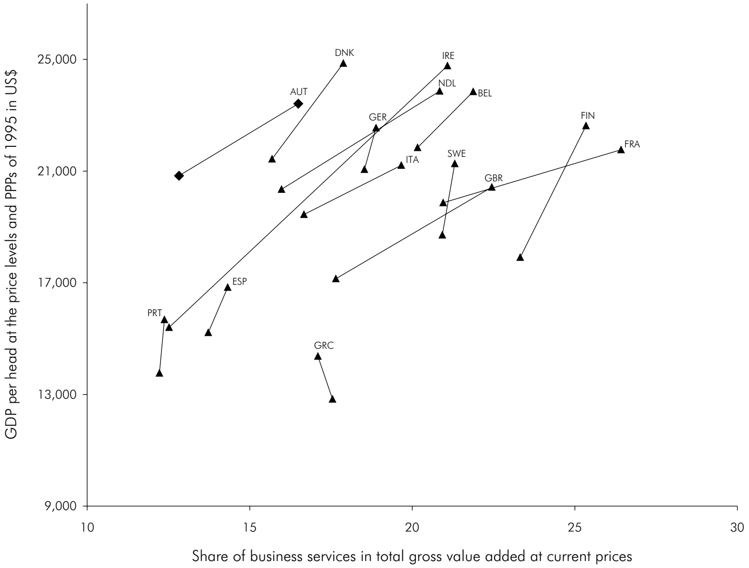


Fig. 2. GDP and the value added share of business services: 1992/99

Note: First year 1993 (Sweden); 1995 (Belgium, Greece, Portugal, Spain); final year 1998 (Spain, Sweden).

$C\Delta$ = cumulated changes; LP = labor productivity; L = employment

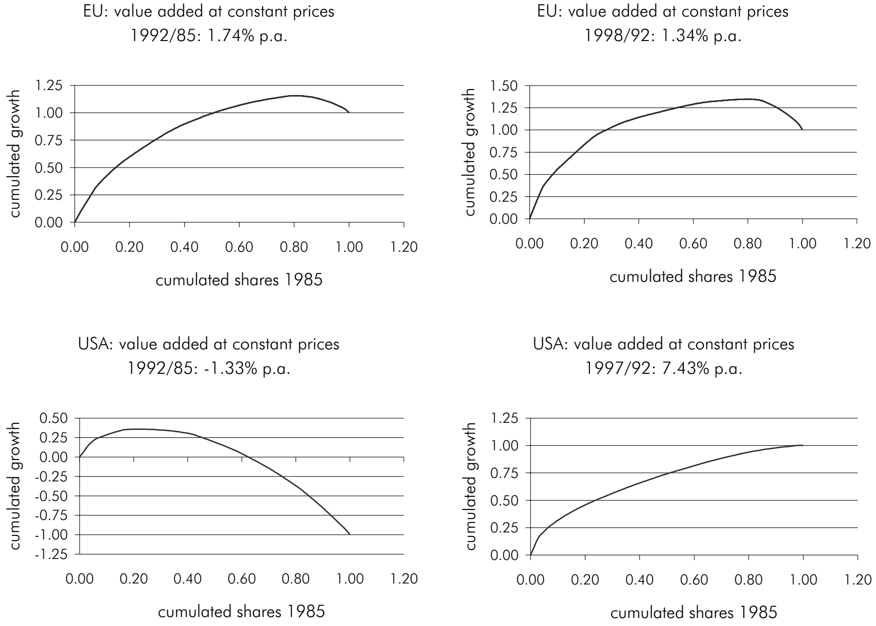
Through straightforward substitution, this expression can be reduced to

$$C\Delta(LP_i) = \frac{\sum\limits_{c=1}^{i} VA_{i,fy} - VA_{i,by}}{VA_{tm,fy} - VA_{tm,by}} \tag{3}$$

$C\Delta$ = cumulated changes; LP = labor productivity; VA = value added

Before the cumulated shares can be calculated, industries must be sorted according to the ratio of their share in productivity growth and their share in the total value added of the base year. The resulting Lorenz-type curve is a visual representation of the degree of concentration with regard to the contribution of individual industries to the changes in aggregate labor productivity. Finally, we re-scale the vertical axis, so that it corresponds to the average annual growth of value added per employee for total manufacturing.

The graphs in Fig. 3 and Fig. A1 in the Appendix are easy to interpret. A straight line from the origin to the end, where the cumulated shares of value added in the base year amount to unity, implies that the contribution of all industries to the aggregate growth of labor productivity was in exact proportion to their initial size. Conversely, a strong curvature of the line indicates that the contributions to aggregate productivity growth are unevenly distributed across industries, even after accounting for variations in initial size.



EU: value added at constant prices 1992/85: 1.74% p.a.

EU: value added at constant prices 1998/92: 1.34% p.a.

USA: value added at constant prices 1992/85: -1.33% p.a.

USA: value added at constant prices 1997/92: 7.43% p.a.

Fig. 3. Contributions to the aggregate growth of labor productivity in manufacturing

As the importance of structural change is easily recognizable due to the strong curvature of the lines in most of the graphs, we generally reject the yeast analogy (as did Harberger). Industries do not contribute proportionally to overall growth in labor productivity. The great variation in productivity growth among industries implies that diversity and heterogeneity at the levels of markets and industries are undeniable facts of economic development.

Beyond this initial observation, two additional features suggest a more complex Schumpeterian interpretation of industrial development:

- First, structural change itself is not a uniform process, but rather appears in clusters. The productivity growth of industries exhibits a more uniform pattern in some periods and a more varied performance in others.
- Second, the graphs suggest a certain tendency of structural change to be more pronounced during periods of low aggregate growth, whereas smoother developments are often evident in conjunction with larger productivity increases.

The case of the USA best illustrates the latter pattern of development. The graph demonstrates that the impressive US surge in productivity throughout the New Economy boom was preceded by a phase of severe restructuring within the manufacturing sector during the period prior to 1992. Contrary to conventional wisdom, the period following 1992 was characterized by relatively smooth overall development, lacking any pronounced structural changes within the manufacturing sector.[7] Although this pattern has been much less pronounced in Europe, the last point suggests a complex non-linear relationship between structural change and economic development. We should keep this in mind as a warning that even the straightforward econometric setting applied in the next section cannot fully uncover the meso-macro link in Schumpeterian development.

3 The econometric evidence

So far we have established that differential growth between industries is an undeniable empirical fact, which is apparently related (be it through correlation over time or simple mechanical decomposition) to aggregate patterns of development. However, the visual inspections only suggest plausible interpretations, without making any allowances for causal inferences. In this final section, we will ask more specifically, whether the patterns of industrial specialization and structural change have any significant impact on income levels and growth. In other words, we want to know whether qualitative transformations of the productive system, which in our case are exemplified by variations in industrial structure, have an impact on aggregate development.

---

[7] The overall pattern invokes a Schumpeterian interpretation of creative destruction, clustered in certain periods of major technological breakthroughs and enabling productivity growth to be more evenly distributed during the phase of widespread adaptive response thereafter. It is consistent with neo-Schumpeterian views on the processes of innovation and diffusion, where, for instance, Freeman et al. (1982), Perez (1983) or Freeman and Louca (2001) argue that technological and organizational innovations are more likely in periods of slow growth, especially in a few leading sectors. Periods of rapid growth are often characterised by a more uniform process of diffusion across other sectors as well. I am grateful to one of the anonymous referees for making that point.

The panel structure has special merits, because in traditional cross-country regressions we can only account for those determinants that are tangible enough to enable the proper measurement of internationally comparable indicators. However, intangible factors such as knowledge, organization, and institutions, which comprise many dimensions of social interaction that cannot be readily observed or measured, also have a decisive impact on an economy's path of development and growth (Nelson, 1998). Taking individual country effects into consideration, the panel econometric framework enables us to control for heterogeneity in those unobserved country-specific factors, which we can reasonably assume to remain constant over the period under investigation.[8]

We will test for both income levels and growth. The variables used in the two models are explained in Table 1. The first will be a fixed effects panel regression (LSDV) with GDP per capita at purchasing power parities for 1995 as the dependent variable. Among the regressors, we include data on demography, the business cycle, labor markets, and capital accumulation, augmented by a vector X of various structural indicators of relative specialization patterns. Since these indicators refer to relative shares of typically fast growing industrial branches (technology driven manufacturing, total services, and business services), our specific hypothesis is that these structural indicators have a significant and positive impact, which can be due either to differential opportunities for entrepreneurial discovery and productivity growth, or positive spillovers.[9] We use time lags to keep out the opposite effect of growth on structural change via differential income elasticities of demand.

The basic model for the estimation of income levels is the following:

$$\begin{aligned} lY_{i,t} &= \alpha + \beta_1 lPOP_{i,t} + \beta_2 lPOPWA_{i,t} + \beta_3 EMR_{i,t} + \beta_4 EMR_{i,t-1} \\ &\quad + \beta_5 lINVT_{i,t-1} + \beta_6 \Delta lINVT_{i,t} + \beta_j \mathbf{X}_{i,t-1} + \eta_t + \mu_i + \varepsilon_{i,t} \end{aligned} \tag{4}$$

In the second model, we take the growth of GDP p.c. as the dependent variable. The dynamic specification requires the presence of a lagged dependent variable among the set of regressors. Correlations between the lagged dependent variable and the error term are then resolved by first differencing, which also removes the country specific fixed effects (μ_i). The following expression provides the corresponding model for our estimation of growth rates:

$$\begin{aligned} \Delta lY_{i,t} &= \alpha + \beta_1 \Delta lY_{i,t-1} + \beta_2 \Delta lPOP_{i,t} + \beta_3 \Delta lPOPWA_{i,t} + \beta_4 \Delta EMR_{i,t} \\ &\quad + \beta_5 \Delta EMR_{i,t-1} + \beta_6 \Delta lINVT_{i,t-1} + \beta_7 \Delta 2lINVT_{i,t} + \beta_{j>7} \Delta X_{i,t} \\ &\quad + \Delta \eta_t + \Delta \varepsilon_{i,t} \end{aligned} \tag{5}$$

---

[8] For a detailed discussion of the benefits and limitations of the panel framework see e.g. Baltagi (1995). The application of panel econometrics to issues of economic growth was pioneered by Islam (1995), who allowed for heterogeneity only in intercepts. The same approach is also taken in this analysis. The controversy as to whether or not allowances should also be made for heterogeneity in slope parameters is presented in Pesaran and Smith (1995), Lee et al. (1998) and Islam (1998). An intermediate approach, the so called "pooled mean group estimation", was recently applied to growth regressions by Bassanini and Scarpetta (2001) as well as Bassanini et al. (2001). Because of the limited sample size mean group estimations were not feasible for the purpose of this exercise.

[9] The latter is also consistent with endogenous growth theory. For a detailed discussion of causal linkages between industrial structure and aggregate growth, see Peneder (2001; 2003).

Table 1. Denomination of the variables

Label	Description
	Dependent variable
Aggregate level of income	
$lY_{i,t}$	GDP per capita at PPPs of 1995 (logarithm)
Aggregate growth	
$\Delta lY_{i,t}$	Growth of $Y_{i,t}$
	General explanatory variables (Macro-level)
Catching up	
$\Delta lY_{i,t-1}$	Lagged dependent variable in the growth estimation
Demography	
$lPOP$	Total population (logarithm)
$\Delta lPOPWA$	Growth of POP
$lPOPWA$	Population at working age (logarithm)
$\Delta lPOPWA$	Growth of $POPWA$
Business cycle / labour markets	
EMR	Employment rate (‘national business cycle’ effects)
ΔEMR	First differences of EMR
$EMR_{(t-1)}$	Lagged employment rate (‘tightness of labour market’)
$\Delta EMR_{(t-1)}$	First differences of $EMR_{(t-1)}$
η_t	Time trend (‘global business cycle’ effects)
Capital accumulation	
$lINVT_{(t-1)}$	Lagged gross fixed capital investment (logarithm)
$\Delta lINVT$	Growth of $INVT$
$\Delta lINVT_{(t-1)}$	Lagged growth of $INVT$
$\Delta 2lINVT$	Second differences of $lINVT$
	Structural explanatory variables (Meso-level)
Manufacturing	
$XSR\_tdi_{(t-1)}$	Lagged share of technology driven industries in total exports relative to OECD
ΔXSR_tdi	First differences of XSR_tdi
$\Delta XSR\_tdi_{(t-1)}$	Lagged first differences of XSR_tdi
$\Delta 2XSR\_tdi$	Second differences of XSR_tdi
$MSR\_tdi_{(t-1)}$	Lagged share of technology driven industries in total imports relative to OECD
Services	
$SOTS_{(t-1)}$	Lagged share of total services in gross value added (at current prices)
$\Delta SOTS_{(t-1)}$	Lagged first differences of $SOTS$
$\Delta 2SOTS$	Second differences of $SOTS$
$SOBS_{(t-1)}$	Lagged share of business services in gross value added (at current prices)
$\Delta SOBS_{(t-1)}$	Lagged first differences of $SOBS$
$\Delta 2SOBS$	Second differences of $SOBS$

Data sources: OECD (ECO) for the dependent and the general explanatory variables; UNO (COMTRADE) for structural variables on manufacturing; OECD (Services statistics) for structural variables on services.

For the dynamic specification, we apply the generalized method of moments (GMM) estimator developed by Arellano and Bond (1991), which resolved the correlation between the differenced dependent variable and the transformed error term by means of an extensive instrument matrix. We calculate both the one-step and two-step GMM estimators, which should be asymptotically equivalent, if the error terms are independent and homoscedastic across countries and over time (Arellano and Bond, 1991, p. 279). In our calculations, the coefficients are indeed very similar, but the one-step estimations are considerably less favorable in terms of the statistical significance of the individual variables. The major source of this problem seems to be the heteroscedasticity of the error term. Since we only present the two-stage results in Table 3, we should keep in mind this caveat regarding the robustness of our findings. A critical assumption for the validity of Arellano-Bond GMM estimator is the lack of second order serial correlation. The test results (A-B(2)) are reported at the bottom of Table 3. The estimates are inconsistent when the null hypothesis of no second-order autocorrelation in the first-differenced residuals is rejected at a significant level.[10]

The regressors are divided into two groups of general and structural explanatory variables. The first refer only to the macro-level and together comprise the baseline specification, in which no structural factors are included. Among them, we find demography (total population and population at working age), capital accumulation (lagged levels and growth in gross fixed capital investment), and those variables that were primarily intended to control for business cycle effects. Whereas the overall time trend (dummies for individual years) captures the influence of the global business cycle, we use the employment rate as the best available proxy to control for the influence of national business cycles. Additionally, the lagged rate of employment is applied as an indication of the relative tightness of the labor market.[11]

The second set of structural variables is again divided into two groups of measures, namely, one for relative specialization in a specific type of manufacturing and one for service industries. The first refers to particularly technology driven industries, a group that was classified in Peneder (2002). Calculating export and import shares relative to the OECD, an overall time trend was eliminated right from the beginning. The second consists of the share of total services in gross value added as a broad measure of tertiarization, as well as the respective share of business services. Due to limited data availability, business services are comprised of ISIC codes 70 to 74 (including real estate). Because of the many missing values, no reliable global benchmark could be constructed, and the overall time trend has to be captured by the year dummies.

The analysis is based upon a data panel comprising $i = 28$ OECD countries[12], covering the years 1990 to 2000. The OECD ECO database was the source of data on GDP, population, employment, and capital investment. Value added shares in

---

[10] The Wald test of the null hypothesis that all the coefficients except the constant are zero is rejected in every specification and is not reported separately. However, the Sargan test of the validity of the restrictions concerning over-identifying restrictions reveals extreme discrepancies in the one-step and two-step estimators under the assumption of homoscedastic error terms, always rejecting when the former, but never when the latter is applied. This also indicates the aforementioned problems of heteroscedasticity (see also Arellano-Bond, 1991, p. 287).

[11] Research and development expenditures are not included for lack of equally comparable time series in many of the OECD countries included in the regressions.

[12] The selection of countries reflects the availability of data for the variables listed in Table 1 in the OECD national accounts data base. The sample exceeds the one used for Harberger's visualisation, because the latter relied on a more demanding disaggregation of production statistics at the 3-digit level.

Table 2. Fixed effects panel regression of Log GDP p.c.: 1990 to 2000

Dependent variable:	I	II	III	IV	V	VI
$lY_{i,t}$	$\beta(t)$	$\beta(t)$	$\beta(t)$	$\beta(t)$	$\beta(t)$	$\beta(t)$
$lPOP$	-2.2779^{***}	-1.9243^{***}	-2.1282^{***}	-2.4481^{***}	-2.4447^{***}	-1.7657^{***}
	(-8.87)	(-6.74)	(-7.22)	(-7.61)	(-7.58)	(-5.34)
$lPOPWA$	1.6155^{***}	1.4541^{***}	1.5270^{***}	1.8087^{***}	1.8058^{***}	1.3988^{***}
	(8.15)	(6.83)	(6.79)	(7.49)	(7.46)	(5.80)
EMR	0.5963^{***}	0.7161^{***}	0.8496^{***}	1.0225^{***}	1.0157^{***}	0.6487^{**}
	(3.97)	(3.26)	(3.58)	(4.01)	(3.97)	(2.59)
$EMR_{(t-1)}$	-0.1088	-0.4156^{*}	-0.5819^{**}	-0.7290^{**}	-0.7084^{**}	-0.4657^{*}
	(-1.00)	(-1.71)	(-2.16)	(-2.54)	(-2.43)	(-1.68)
$lINVT_{(t-1)}$	0.2126^{***}	0.2146^{***}	0.2361^{***}	0.2248^{***}	0.2230^{***}	0.2216^{***}
	(8.57)	(7.88)	(7.73)	(7.18)	(7.03)	(7.44)
$\Delta lINVT$	0.2112^{***}	0.1588^{***}	0.1658^{***}	0.1457^{***}	0.1452^{***}	0.1707^{***}
	(8.43)	(4.80)	(4.58)	(3.95)	(3.93)	(4.89)
$XSR\_tdi_{(t-1)}$		0.0636^{***}				0.0720^{***}
		(3.47)				(3.72)
ΔXSR_tdi		0.0005				
		(0.02)				
$MSR\_tdi_{(t-1)}$		0.1187^{***}				0.1193^{***}
		(3.95)				(3.64)
$SOTS_{(t-1)}$			0.0009		-0.0006	-0.0024^{*}
			(0.68)		(-0.42)	(-1.84)
$SOBS_{(t-1)}$				0.8646^{***}	0.8938^{***}	0.8788^{***}
				(4.18)	(4.09)	(4.23)
Year dummies (η_t)	Yes	Yes	Yes	Yes	Yes	Yes
No. observations	330	272	255	231	231	231
No. countries	29	29	29	29	29	29
R-sq within:	0.9006	0.9212	0.9158	0.9197	0.9198	0.9306

Note: GDP at PPP of 1995; XSR = shares in total exports relative to OECD; MSR = shares in total imports relative to OECD; tdi = technology driven industries; hs = high skill industries.

the services sector were extracted from OECD (2001). All other structural variables stem from the UN COMTRADE database.

The estimations for aggregate income levels are reported in Table 2, while those on aggregate growth are presented in Table 3. The basic specification excludes all structural variables and is consistent with prior expectations based on general considerations found in the growth literature. *Ceteris paribus*, GDP per capita must fall with the size of the population, whereas the size of the population at working age has a positive impact. The employment rate is pro-cyclical and the respective coefficient therefore positive. Finally, both the lagged levels and growth of capital investment foster per capita income. The only variable which fails to be significant is the lagged employment rate in the fixed effects panel regression, although it is nevertheless retained in the model, thanks to its significant (negative) impact in the other specifications augmented by the structural indicators. The lagged dependent variable in the dynamic estimations is significant and positive, thereby controlling for ‘catching up’ effects.

Table 3. Dynamic panel regression of growth in GDP per capita: 1990 to 2000

Dependent variable:	I	II	III	IV	V
$\Delta lY_{i,t}$	$\beta(t)$	$\beta(t)$	$\beta(t)$	$\beta(t)$	$\beta(t)$
$\Delta lY_{(t-1)}$	0.7166^{***}	0.6632^{***}	0.5751^{***}	0.4549^{***}	0.7100^{***}
	(9.95)	(7.87)	(4.93)	(6.49)	(6.07)
$\Delta lPOP$	-0.8071^{***}	-0.8208^{***}	0.0188	-0.4762^{**}	0.1257
	(-5.18)	(-2.92)	(0.05)	(-2.50)	(0.37)
$\Delta lPOPWA$	0.4477^{***}	0.6635^{***}	-0.0909	0.4725^{***}	-0.2889
	(3.57)	(2.91)	(-0.24)	(3.08)	(-0.94)
ΔEMR	0.4071^{***}	0.3632^{***}	0.5992^{***}	0.4054^{***}	0.2038
	(3.16)	(3.43)	(3.04)	(3.63)	(1.41)
$\Delta EMR_{(t-1)}$	-0.3453^{*}	-0.4212	-0.9123^{***}	-0.8143^{***}	-0.2768
	(-1.96)	(-1.53)	(-2.78)	(-4.75)	(-0.98)
$\Delta lINVT_{(t-1)}$	0.0495^{***}	0.0579	0.1386^{***}	0.1666^{***}	0.0849^{***}
	(3.65)	(1.60)	(2.80)	(6.09)	(4.92)
$\Delta 2IINVT$	0.1965^{***}	0.1816^{***}	0.2125^{***}	0.2127^{***}	0.2547^{***}
	(19.79)	(14.14)	(16.40)	(18.01)	(21.47)
$\Delta XSR\_tdi_{(t-1)}$		0.0258^{***}			
		(3.12)			
$\Delta 2XSR\_tdi$		0.0405^{***}			
		(4.88)			
$MSR\_tdi_{(t-1)}$		0.0049^{***}			
		(2.60)			
$SOTS_{(t-1)}$			-0.0003^{*}		-0.0004^{**}
			(-1.91)		(-2.46)
$\Delta 2SOTS$			-0.0023^{***}		-0.0037^{***}
			(-4.56)		(-13.36)
$\Delta SOBS_{(t-1)}$				0.8827^{***}	0.5674^{**}
				(3.21)	(2.07)
$\Delta 2SOBS$				0.5269^{*}	0.3432^{**}
				(1.76)	(2.25)
Year dummies (η_t)	Yes	Yes	Yes	Yes	Yes
No. observations	275	242	197	173	173
No. Countries	29	29	28	28	28
A-B test (2)	0.1720	0.9365	0.9033	0.7162	0.6621

Note (1): GDP at PPP of 1995; Δvar = variable in first differences; Δvar_{t-1} =.. lagged differences; $\Delta 2var$ =.. second differences; $XSR(MSR)$ = shares in total exports (imports) relative to OECD; tdi = technology driven industries; hs = high skill industries.

Note (2): Time dummies were only used for 1992, 1994, 1995, 1998, and 1999, which were selected because of being significant in the base model of specification I and in order to avoid second order serial correlation in A-B test (2).

Note (3): $MSR\_tdi_{(t-1)}$ and $SOTS_{(t-1)}$ were introduced as strictly exogenous variables without first differencing.

In a world of Schumpeterian development, we expect that, in addition to all these factors, aggregate income and growth cannot remain totally unaffected by the specific production structure of the economy. This proposition is in sharp contrast to the conventional neo-classical growth models, in which (due to their purely macroeconomic focus) industrial structure has no role to play. Augmenting the basic model with selected structural variables, we find ample evidence of its significant impact in Tables 2 and 3. With regard to the manufacturing sector, both the relative export and import shares of technology driven industries matter. While the first can be explained by differential growth, as well as producer related externalities, the second finding clearly indicates the presence of user related spillovers in this type of industry. With respect to services, an interesting differentiation also surfaces. The share of total services in overall value added is insignificant in the regression on income levels, but is significantly negative in the growth equations. Taken on its own, this result is largely consistent with Baumol's cost disease argument.[13] Services, however, comprise an extremely heterogeneous sector, which requires further differentiation (Peneder et al., 2003). For instance, if we focus on the value added share of business services, a reversed image emerges with significant positive coefficients. This pattern of a negative impact exerted by the share of total services, contrasted with a positive coefficient for business services, also persists when both kinds of structural variables are included.

In conclusion, the econometric evidence convincingly demonstrates that variations in industrial structure have a significant impact on aggregate development and thus substantiates our concern for the meso-macro link in economic development. But in closing this section, we should also emphasize the weaknesses of the analysis. Besides the aforementioned econometric problems with the dynamic model, we must acknowledge that the interpretation of the structural effects is not as straightforward and precise as we might wish. Remember that we have not been able to capture the more complex time patterns of creative destruction suggested in some of the Harberger visualizations. But also with respect to the structural effects, which were actually identified, we face the problem of multiple alternative explanations.

In a straightforward evolutionary interpretation, it is tempting to treat structural effects as a result of the direct impact of differential growth. Some industries tend to expand faster and achieve higher growth in labor productivity than others. Assuming an under-utilization of productive resources, the greater specialization of a country in industries of this type enhances its prospects for aggregate growth. Second, industries might differ in their propensities to generate positive externalities to the rest of the economy. For instance, producer related spillovers may stem from the enhanced knowledge diffusion of technologically sophisticated production within a relatively small area. Similarly, user related spillovers may stem from new technology that is embodied in capital goods. Although not at odds with an evolutionary interpretation, this explanation is at the same time consistent with the steady state endogenous growth models. Third, variations in sectoral specialization might capture correlations with certain intangible and location bound factors of competitive capabilities (e.g. the aggregate R&D ratio, the national innovation system, or other institutional factors for which sufficiently comparable data are not available), which are not constant over time and therefore have not been eliminated by the country dummies or first differences.

---

[13] It predicts a decrease in overall growth due to rising shares of relatively stagnant service industries (Baumol, 1967; Baumol et al., 1985).

4 Summary and conclusions

This paper opened with the presumption that Schumpeterian *development* is characterized by the simultaneous interplay of growth and the qualitative transformation of the economic system. At the sectoral level, such qualitative transformations become manifest as variations in the sectoral composition of production, i.e. structural change. In contrast to Schumpeter's broader notion of development, theories of economic growth tend to focus exclusively on macroeconomic phenomena. For the sake of analytic tractability and clear identification of the steady state equilibrium solutions, the meso-level of industrial structure is bypassed by the assumption of balanced steady-state growth, uniformly spread across all industries.

We presented an empirical validation of this evolutionary emphasis on Schumpeterian development, focusing on variations in industrial structure and its impact on aggregate income and growth. We traversed three different layers of visibility. Within the first layer of easily recognizable trails, an apparent co-movement in time involving aggregate income and certain selected types of industry motivated our further investigations. In the second layer, the application of Harberger's visualization not only demonstrated that differential productivity growth is an undeniable fact, but also revealed some interesting time patterns in its relationship to aggregate development. Specifically, the boom in the U.S. New Economy in the late 1990s, preceded by a phase of painful but creative destruction in the years prior to 1992, invites a very Schumpeterian interpretation. In the final layer, (dynamic) panel estimations of a standard empirical growth model augmented by various structural variables for 28 OECD countries during the period 1990 to 2000, revealed that variations in industrial structure do have a significant impact on both aggregate income levels and growth. While (consistent with Baumol's cost disease argument) the share of the services sector in total value added exerted a negative influence, the coefficients for the value added share of business services and the export shares of particularly technology driven manufacturing industries were positive and significant. Potential explanations range from differential growth between industries to their different propensities to generate producer-related spillovers. For technology-driven industries, we additionally found a positive impact of relative import shares, indicating the presence of user-related spillovers from embodied technology flows.

The essential message of this study is that variations in industrial structure are significant determinants of aggregate income levels and growth. The empirical evidence thus substantiates the evolutionary emphasis on Schumpeterian *development*, which in addition to the endogeneity of innovation in Schumpeterian *growth* models, comprises growth and structural change as two inseparable elements.

Acknowledgements. My especial thanks are due to Jan Fagerberg, Paul Geroski, Serguei Kaniovski, Michael Landesmann, Markus Marterbauer, Stan Metcalfe, Michael Pfaffermayr and two anonymous referees for helpful comments and suggestions, as well as to Eva Sokoll and Traude Novak for their data-related assistance.

Appendix

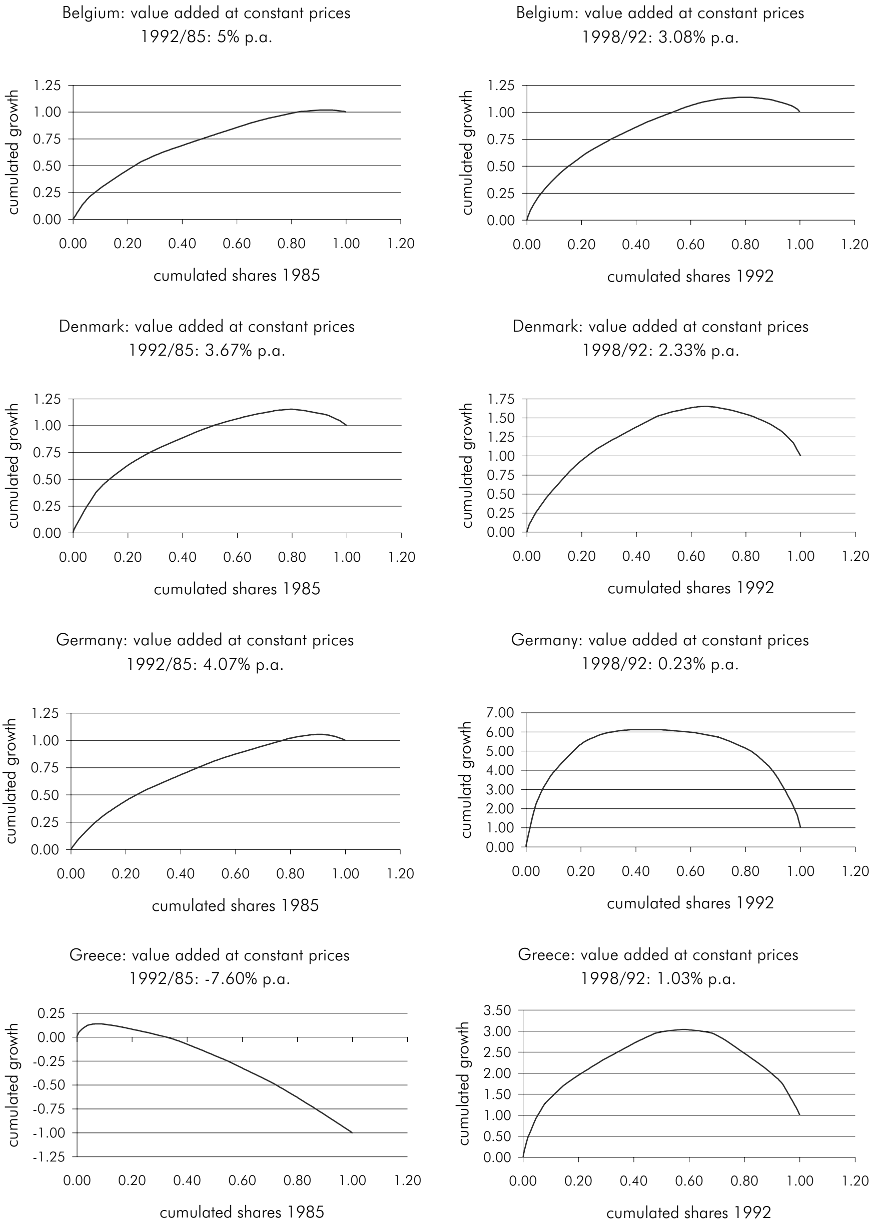
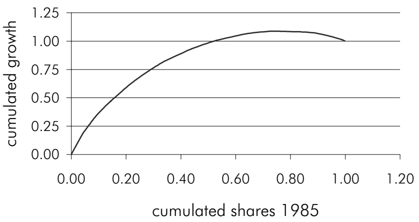
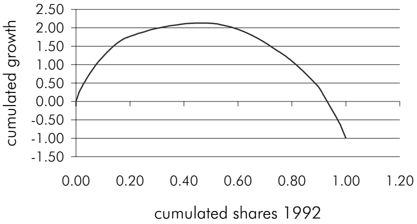


Fig. A1. Harberger's visualisation for individual EU countries

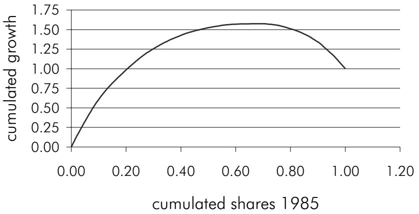
Spain: value added at constant prices
1992/85: 3.97% p.a.



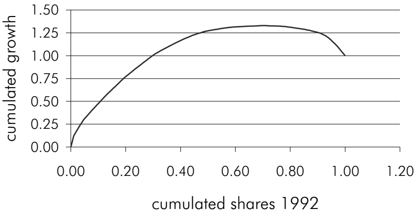
Spain: value added at constant prices
1998/92: -0.53% p.a.



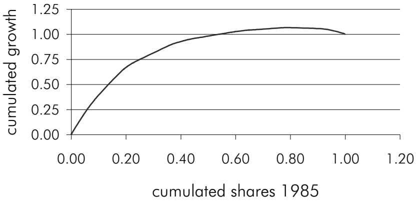
France: value added at constant prices
1992/85: 1.40% p.a.



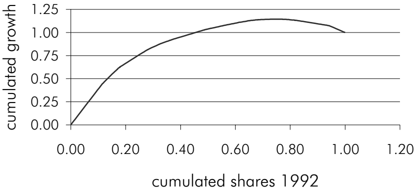
France: value added at constant prices
1998/92: 1.88% p.a.



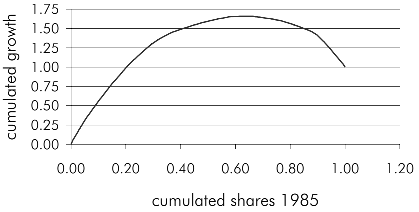
Ireland: value added at constant prices
1992/85: 4.88% p.a.



Ireland: value added at constant prices
1998/92: 8.23% p.a.



Italy: value added at constant prices
1992/85: 1.24% p.a.



Italy: value added at constant prices
1998/92: -0.83% p.a.

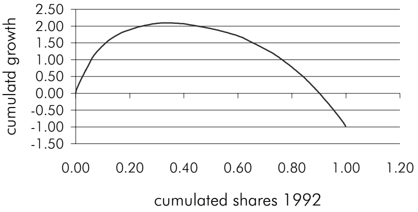


Fig. A1. (continued)

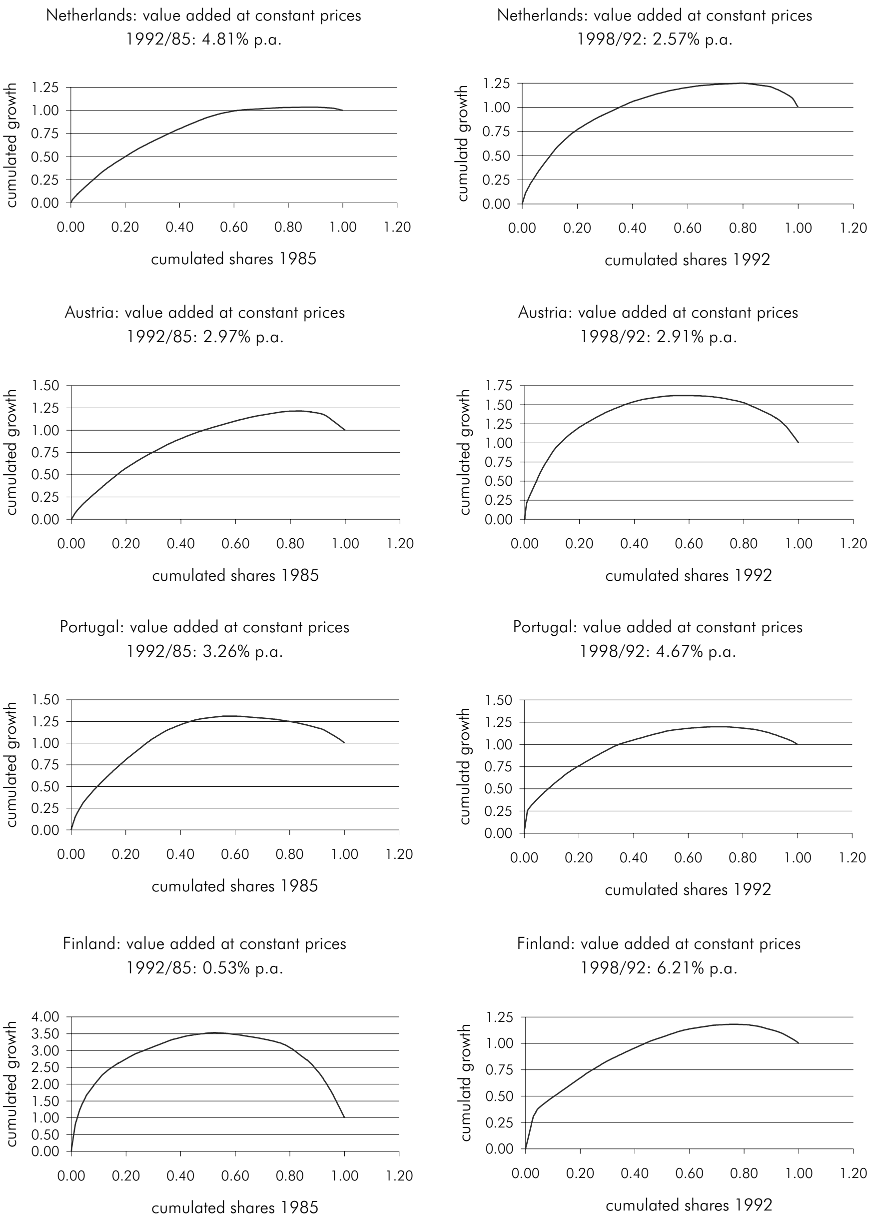


Fig. A1. (continued)

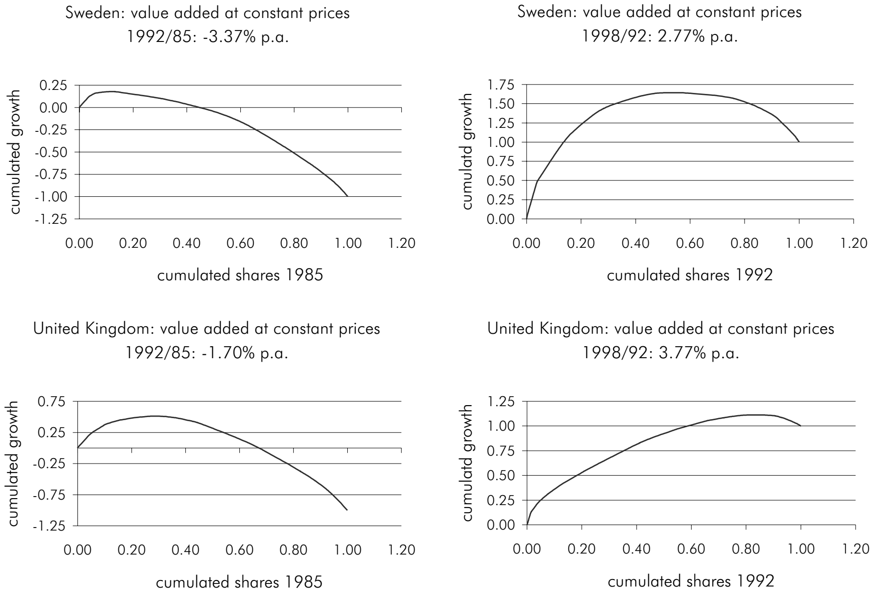


Fig. A1. (continued)

References

Aghion P, Howitt P (1998) Endogenous Growth Theory. Cambridge, MA: M.I.T. Press

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Review of Economic Studies 58:277–297

Baltagi BH (1995) Econometric Analysis of Panel Data. Chichester: Wiley

Bassanini A, Scarpetta S (2001) Does human capital matter for growth in oecd countries? evidence from pooled mean group estimates. OECD, Economics Department Working Papers, No. 282

Bassanini A, Scarpetta S, Hemmings P (2001) Economic growth: The role of policies and institutions. panel data evidence from oecd countries. OECD, Economics Department Working Papers, No. 283

Baumol WJ (1967) Macroeconomics of unbalanced growth: The anatomy of urban crisis. The American Economic Review 57:415–426

Baumol WJ, Blackman SAB, Wolff EN (1985) Unbalanced growth revisited: Asymptotic stagnancy and new evidence. The American Economic Review 75:806–817

Cheng LK, Dinopoulos E (1996) A multisectoral general equilibrium model of schumpeterian growth and fluctuations. Journal of Economic Dynamics and Control 20:905–923

Dinopoulos E, Thompson P (1998) Schumpeterian growth without scale effects. Journal of Economic Growth 3:313–335

Echevarria C (1997) Changes in sectorla composition associated with economic growth. International Economic Review 38:431–452

Egger P, Pfaffermayr M (2001) The impact of (changes in) corporate structure on price dispersion, productivity and profits in eu manufacturing. In De Voldere I, Reynaerts J, Rommens K, Sleuwaegen L, Veugelaers, R. (eds.) Determinants of Industrial Concentration, Market Integration and Efficiency in the European Union, Study for the European Commission

Fagerberg J (1994) Technology and international differences in growth rates. Journal of Economic Literature 32(9):1147–1175

Freeman C, Clark J, Soete L (1982) Unemployment and Technical Innovation. London: Pinter

Freeman C, Louca F (2001) As Time Goes by: From the Industrial Revolutions to the Information Revolution. Oxford: Oxford University Press

Grossman GM, Helpman E (1991) Innovation and Growth in the Global Economy. Cambridge, MA: M.I.T. Press

Harberger AC (1998) A vision of the growth process. The American Economic Review 88:1–32

Islam N (1995) Growth empirics: A panel data approach. The Quarterly Journal of Economics 110:1127–1170

Islam N (1998) Growth empirics: A panel data approach - a reply. The Quarterly Journal of Economics 113:325–329

Kongsamut P, Rebelo S, Yie D (2001) Beyond balanced growth. Review of Economic Studies 68:869–882

Lee K, Pesaran MH, Smith R (1998) Growth empirics: A panel data approach - a comment. The Quarterly Journal of Economics 113:319–323

Lewer JJ, Van den Berg H (2003) Does trade composition influence economic growth? time series evidence for 28 oecd and developing countries. Journal of International Trade & Economic Development 12:39–96

Meckl J (2002) Structural change and generalized balanced growth. Journal of Economics (Zeitschrift für Nationalökonomie) 77:241–266

Metcalfe JS (1998) Evolutionary Economics and Creative Destruction. London: Routledge

Metcalfe JS (2001) Consumption, preferences, and the evolutionary agenda. Journal of Evolutionary Economics 11:37–58

Montobbio F (2002) An evolutionary model of industrial growth and structural change. Structural Change and Economic Dynamics 13:387–414

Nelson RR (1998) The agenda for growth theory: A different point of view. Cambridge Journal of Economics 22:497–520

Nelson RR, Winter SG (1982) An Evolutionary Theory of Economic Change. Cambridge, MA: Belknap

OECD: (2001) Services statistics on value added and employment, 2001 edition. Paris: OECD

Peneder M (2001) Entrepreneurial Competition and Industrial Location. Cheltenham: Edward Elgar

Peneder M (2002) Intangible investment and human resources. Journal of Evolutionary Economics 12:107–134

Peneder M (2003) Industrial structure and aggregte growth. Structural Change and Economic Dynamics 14:427–448

Peneder M, Kaniovski S, Dachs B (2003) What follows tertiarisation? structural change and the role of knowledge based services. The Service Industries Journal 23:47–66

Perez C (1983) Structural change and assimilation of new technologies in the economic and social systems. Futures 357–375

Pesaran MH, Smith R (1995) Estimating long-run relationships from dynamic heterogenous panels. Journal of Econometrics 68:79–113

Plumper T, Graff M (2001) Export specialization and economic growth. Review of International Political Economy 8:661–88

Schumpeter JA (1911/34) Theorie der wirtschaftlichen Entwicklung, 4th edn. Berlin: Duncker & Humblot

Schumpeter JA (1942/50) Capitalism, Socialism and Democracy, 3rd edn. New York: Harper & Row

Schumpeter JA (1947) The creative response in economic history. Journal of Economic History 7:149–159

Silverberg G (1988) Modelling economic dynamics and technical change: Mathematical approaches to self-organization and evolution. In Dosi G, Freeman C, Nelson R, Silverberg G, Soete, L. (eds.) Technical Change and Economic Theory 531–559

Silverberg G, Verspagen B (1997) Economic growth: An evolutionary perspective. In Reijnders, J (ed.) Economics and Evolution 137–170. Cheltenham: Edward Elgar

Verspagen B (2001) Economic growth and technological change: An evolutionary interpretation. STI Working Papers

Towards an evolutionary interpretation of aggregate labor market regularities*

Giorgio Fagiolo[1], **Giovanni Dosi**[2], **Roberto Gabriele**[3]

[1] Sant'Anna School of Advanced Studies, Pisa, Italy
(e-mail: fagiolo@sssup.it)
[2] Sant'Anna School of Advanced Studies, Pisa, Italy
(e-mail: gdosi@sssup.it)
[3] D.I.S.A., University of Trento, Italy and Sant'Anna School of Advanced Studies, Pisa, Italy
(e-mail: rgabriel@cs.unitn.it)

Abstract. In this paper, we present an agent-based, evolutionary, model which formalizes from the bottom up individual behaviors and interactions in both product and labor markets. We describe vacancy and wage setting, matching and bargaining, demand and price formation as endogenous processes. Firms enjoy labor productivity improvements and are selected in the product market. Simulations show that: (i) the model is able robustly to reproduce Beveridge, Wage and Okun curves; (ii) Okun coefficients greater than one emerge even if individual firms employ linear technologies; (iii) changes in institutional, behavioral, and technological parameters induce statistically detectable shifts in Okun and Beveridge curves.

Key words: Labor Markets – Dynamics – Aggregate Regularities – Beveridge Curve – Okun Curve – Wage Curve – Matching Models

JEL Classification: J63, J64, O12, J41

1 Introduction

Over the last couple of decades, a quite large literature has been trying to investigate the process through which firms and workers meet in the labor market, how

---

* Thanks to Uwe Cantner, Herbert Dawid, Peter Flaschel, Alan Kirman, Willi Semmler, Mauro Sylos Labini, Leigh Tesfatsion, two anonymous referees, and the participants to the conference "Wild@Ace: Workshop on Industry and Labor Dynamics. An Agent-based Computational Economics approach", Laboratorio Revelli, Turin, October 3-4, 2003, for valuable comments and suggestions.

Correspondence to: Giorgio Fagiolo, Sant'Anna School of Advanced Studies, Laboratory of Economics and Management, Piazza Martiri della Libertà, 33. I-56127 PISA (Italy). Tel: +39-050-883341. Fax: +39-050-883344.

this matching process affects wage setting and (un)employment dynamics, and the extent to which unemployment and output interact over the business cycle[1].

Three well-known empirical aggregate regularities seem to provide a quite complete picture of the interplay between the forces at work. First, the Beveridge curve predicts a negative relationship between rates of vacancies and rates of unemployment. Second, the Phillips curve suggests that changes in wage rates are negatively related to unemployment rates. Alternatively, the Wage curve predicts a negative correlation between levels of real wages and unemployment. Third, the Okun curve posits a more than proportional increase in real GDP for every one percentage point reduction in the unemployment rate.

Most of the literature has tried so far to explain these phenomena on the grounds of a standard "toolbox" based on micro-foundations which postulate hyper-rational firms and workers. The "representative individual hypothesis" is often employed to overcome difficulties entailed by aggregation of heterogeneous agents. Moreover, static equilibrium conditions are largely used to interpret macroeconomic dynamics.

Despite their formal sophistication, the degrees of success of this class of models is, at best, mixed. In particular, existing literature seems to lack a joint explanation of the foregoing three aggregate regularities.

In this paper, we propose a radically different interpretative strategy. The model that we present in the following might be taken as an exploratory attempt to provide a micro-foundation of the interactions between labor-market and output dynamics from an evolutionary perspective[2].

The underlying philosophy builds on the acknowledgement that both firms and workers live in complex systems which evolve through time and might be characterized by endogenous, persistent, novelty. Agents are heterogeneous in their endowments, wealth, and, possibly, in their behavioral rules and rationality skills. Given the complexity of the environment they have to cope with - which changes endogenously as the outcome of individual behaviors - agents can only be boundedly-rational and hold an imperfect understanding of the system (Dosi et al., 2004).

Expectations employed to revise control variables (e.g. demanded and offered wages, output produced, etc.) are typically assumed to be adaptive. Workers and firms interact directly and their choices are affected by those undertaken in the past by other agents. Interaction networks (e.g. matching rules in labor market) are themselves endogenous and may change across time. Firms interact both in the labor market and in the product market, wherein their revealed "competitiveness" is affected also by their hiring and wage-setting behaviors.

Macroeconomic dynamics is generated in the model via aggregation of individual behaviors. Typically, non-linearities induced by heterogeneity and far-from-equilibrium interactions induce a co-evolution between aggregate variables

---

[1] For a quite exhaustive overview of the state-of-the-art of both theoretical and empirical labor market literature, cf. Ashenfelter and Layard (1986), Ashenfelter and Card (1999) and Petrongolo and Pissarides (2001).

[2] More on the general *Weltanschauung* of the evolutionary approach is in Dosi and Nelson (1994) and Dosi and Winter (2002). The model we present has large overlappings with the "Agent-Based Computational Economics" (ACE) approach (Tesfatsion, 1997; Epstein and Axtell, 1996; Aoki, 2003), as well as with self-organization models of labor markets pioneered by Lesourne (1992).

(employment, output, etc.). Statistical properties exhibited by aggregate variables might then be interpreted as emergent properties grounded on persistent micro disequilibria.

Consequently, even when some equilibrium relationship exists between aggregate variables (e.g. inflows and outflows from unemployment), the economy might persistently depart from it and follow some disequilibrium path. The observed stable relations amongst those same aggregate variables might emerge out of turbulent, disequilibrium, microeconomic interactions.

Here, making use of a model built on such premises, we shall address two types of questions. First, we shall ask whether the model is able to reproduce robustly over a large set of behavioral and institutional settings the main aggregate regularities that we observe in real-world labor-market data. For instance: Does our model generate jointly Beveridge, Wage, and Okun curves for a sufficiently large region of system parameters? Notice that this would lend support to a disequilibrium foundation of aggregate regularities: despite the fact that the economy always departs from equilibrium (if any), aggregate regularities emerge as the outcome of decentralized interactions, adaptive behavioral adjustments, and imperfect coordination.

Second, we shall try to map different behavioral and institutional settings into statistically distinct patterns of labor market dynamics. For example: Are there institutional and technological settings wherein the economy is *unable* to display robustly a downward-sloping relation between vacancy and unemployment rate? Under which conditions can one observe shifts of the Beveridge curve? And, similarly: Under which technological regimes the Okun curve displays a greater than one absolute elasticity?

The paper is organized as follows. In Section 2 we start by briefly surveying the main empirical findings about the foregoing three aggregate regularities. Next, we discuss how mainstream economic theory has been trying to provide explanations of such stylized facts. In Section 3, we present the model and we discuss the extent to which it departs from existing theoretical frameworks. Section 4 presents the results of simulation exercises. Finally, Section 5 draws some concluding remarks.

2 Individual behaviors, interactions, and aggregate regularities in labor market dynamics: an assessment of the state of the art

2.1 A brief overview of empirical regularities

When dealing with labor market dynamics, a familiar angle of inquiry regards the extent to which "rigidities" and "frictions" are able to account for the observed unemployment levels (Phelps, 1972; Blanchard and Wolfers, 2000). In this respect, the Beveridge curve (BC) is a good starting point. The BC postulates a negative relationship (over time) between the rate of unemployment u and the rate of vacancies v, where rates are defined in terms of total employment[3].

---

[3] Observation of reliable proxies for actual vacancies entails many empirical problems, especially in Europe, see Solow (1998). For instance, one is typically bounded to observe only *ex-ante* vacancies (i.e. job openings). *Ex-post* vacancies (i.e. unfilled job openings) are much more affected by frictions than *ex-ante* ones and thus should be in principle preferred as objects of analysis.

The intuition is simple: if an economy exhibits higher level of vacancies - in turn plausibly corresponding to a higher level of aggregate demand - it is easier for workers to find a job. Thus, one should also observe a lower level of unemployment. Movements along the curve should be typically induced by the business cycle. For instance, contractions should imply - *ceteris paribus* - a reduction in aggregate demand. This should in turn induce a decrease of vacancies and an increase in unemployment.

Moreover, the position of the BC in the (u, v) space is typically related to the degree of “frictions” existing in the labor market and, more generally, to its institutional setting: the closer the curve to the axes, the lower - *ceteris paribus* - market “frictions”. Shifts of the curve are attributed to all factors influencing: (a) *directly* the process of matching between vacancies and unemployed workers (e.g. unemployment benefit system, employment protection laws, active labor market policies, etc.); and (b) *indirectly* the (u, v) relationship via the impact they have on the wage rate (union power, union coverage, degree of coordination of wage bargaining, etc.)[4].

Empirical findings seem to be quite controversial (Blanchard and Diamond, 1989; Nickell et al., 2001; Belot and Van Ours, 2000; Fitoussi et al., 2000). In fact, casual inspection of scatter plots of rough (u, v) data across countries does not show any clear-cut negative relationships. Even if some weakly downward-sloping curves do emerge across sub-samples, it seems that shifts and twists prevail. Heterogeneous cross-country patterns also emerge[5]. However, once controls for institutional factors, time, and country dummies are introduced in panel-data regressions, then quite robust, statistically significant, negative elasticities between u and v typically emerge. That is, BCs emerge within data cells containing homogeneous groups of observations. Together, one observes that the impact of variables that *indirectly* affect unemployment through the wage rate is not significant. As far as shifts are concerned, it seems that all OECD countries present a shift to the right of the BC over time (implying higher “rigidities”). Nevertheless, after the 80’s, some countries, including Germany, Sweden, and Japan, seem to exhibit reverse patterns.

Econometric analyses have helped in highlighting the role of institutional variables in shaping the dynamics of jobs and vacancies. However, current analyses still display some major drawbacks. First, on the methodological side, econometric testing is typically not parsimonious. This could lead to the emergence of negative relationships only in over-homogeneous cells, thus weakening the “robustness” of the regularity. Second, unemployment and vacancy rates are computed as ratios between non-stationary variables, possibly entailing too much variability over time[6].

---

[4] On these points, cf. Nickell et al. (2001).

[5] All this would demand a careful discussion on what does we mean by “aggregate regularities” (i.e. “Is it any BC only in the eye of the observer?”) and their relationship with theory. This is, however, beyond the scope of this paper.

[6] The denominator of both vacancy and unemployment rates is total employment (instead of labor force or population), which does not appear to be I(0); cf. however Layard et al. (1991) for an alternative point of view. The choice of total employment is required if one wants to keep a tight relation with

Finally, the role of technical progress is typically not investigated in econometric analyses and almost always treated as a business cycle effect.

A complementary empirical regularity is the famous Phillips curve, or the alternative Wage curve (Blanchflower and Oswald, 1994). As is known, they both posit the existence of co-movements between unemployment and wages. Two almost alternative worlds can be envisaged. If an economy experiments a negative relationship between *changes* of the wage rate and the unemployment rate, one is in a Phillips curve (PC) regime. Conversely, a Wage curve (WC) world is characterized by a negative relationship between *levels* of the wage rate and unemployment rates (Blanchard and Katz, 1997; Card and Hyslop, 1996)[7].

Some remarks are in order. While the WC is typically taken as a proposition about homogeneous areas (e.g. regions or location-specific labor markets), the PC is assumed to bear a more general validity. Hence, the two may not be mutually exclusive: it is possible to think of homogeneous areas characterized by contemporaneous co-movements of both wage growth and levels in response of unemployment shifts. However, empirical studies (Blanchflower and Oswald, 1994; Card, 1995) show that, in homogeneous areas, WC is in general valid, while PC is not. This seems to be a quite robust finding, holding true across regions, countries, etc.. At the same time, the elasticity of wage levels to unemployment rates varies - although not dramatically - across regions and countries. Notice that since different wage-unemployment elasticities imply different degrees of responsiveness of wages to labor market conditions (as reflected by unemployment rate), workers can earn different wages - holding other conditions fixed - when they choose to work in regions with high or low unemployment rates.

As the WC pertains to homogeneous data cells, one cannot "see it" in rough data. Panel data estimation must be performed in order to control for variables such as personal characteristics of workers, labor market institutions, "fixed" effects allowing discrimination among sectors or regions, etc. A strong result here is that a statistically significant, negative, relationship between the wage and unemployment rate still holds across different institutional setups (Börsch-Supan, 1991; Bleakley and Fuhrer, 1997).

The interpretation of a WC is quite controversial. In fact, Card (1995) prefers to argue about what a WC *is not*. In particular, a WC *is not* a Phillips curve, because it does not emerge as a mispecification of a PC regression. Moreover, a WC *is not* a supply function, as it cannot be obtained as a short-run inverted labor supply function (i.e. a relationship linking wage and unemployment through a given supply of labor in the short-run).

Nevertheless, once one has acknowledged the fact that the WC robustly emerges as an aggregate empirical regularity, some important implications follow. On the one hand, the market-clearing (equilibrium) interpretation underlying a PC cannot be invoked anymore. On the other hand, the competitive equilibrium framework does not easily account for WC emergence. In fact, a competitive labor market with all

---

the BC theoretical counterpart modeled through a homogenous of degree one matching function (see below). In the model which follows, we define all rates in terms of total population (or labor force).

[7] For additional evidence on the wage vs. Phillips curves debate - especially concerning "wage-price spirals" - see Flaschel et al. (2003) and references therein.

its canonical features would lead to a *positive* correlation between unemployment and the wage rate. Climbing up a downward demand for labor schedule - i.e. raising wage - would indeed induce higher levels of unemployment, as the unmet supply of labor would grow.

The third aggregate regularity we address here - i.e. the Okun curve (OC) - characterizes the interplay between labor markets and economic activity (Okun, 1962, 1970). In fact, one typically observes a negative, linear, relationship between changes in unemployment rate and GDP growth rates, with an absolute value of the slope larger than one. The standard interpretation[8] runs as follows. Suppose that in the economy there is a under-utilization of labor resources with respect to the full employment level (i.e. unemployment rates are higher than the "natural" level). Then, the effect on economic activity of the cost associated to such under-utilization is more than proportional.

Therefore, whatever the causes, empirical evidence suggests an amplifying feedback between unemployment dynamics and output dynamics. A decrease of one percentage point in the unemployment rate - *ceteris paribus* - is associated with a growth rate of GDP of about two to three percentage points (according to original Okun estimations). Notice that a coefficient greater (less) than one entails some form of increasing (decreasing) returns. The "*ceteris paribus*" assumption is, however, far from innocent: it means that, over different periods of expansions and recessions, all other variables affecting GDP growth should remain nearly stable.

The debate about the existence (and the slope) of the Okun curve is not yet settled. First, the empirical value of the Okun coefficient (i.e. absolute value of the slope of the regression between changes in unemployment rate and GDP growth) is still a subject of controversy. So, for example, Prachowny (1993) challenges the "*ceteris paribus*" assumption and shows that taking into account all variables that increase GDP (e.g. changes in weekly hours, movements in capacity utilization, labor productivity) leads us to a decreasing returns regime: in Prachowny's exercises, a 1-point decrease in unemployment rate only induces an increase in GDP of 0.66%. Conversely, Attfield and Silverstone (1997), by taking into account cointegration relationships between I(1) variables, recover an Okun coefficient in line with an increasing returns economy. Moreover, they show that additional control variables introduced by Prachowny are no longer significant when ECM (Error Correction Models) are employed and estimates are computed using dynamic OLS.

A second issue concerns whether the Okun coefficient is stable over time and across countries (Moosa, 1997; Sögner and Stiassny, 2000). Evidence shows that Okun coefficients are weakly stable over time but quite heterogeneous across countries. Moreover, the Okun relationship seems to be stronger in North-America than in Europe.

From a methodological point of view, the interpretation of the Okun curve must be carefully spelled out. The traditional interpretation is a static one. The joint bivariate process simply implies an invariant relationship with an implicit causality arrow going from economic activity to unemployment. Blanchard and Quah (1989)

---

[8] Notice that an alternative interpretation can be given in terms of labor-productivity / unemployment changes (i.e. as a cyclical "Verdoorn-Kaldor" type of law), displaying rising productivity as unemployment falls.

and Evans (1989) have instead challenged this interpretation and introduced some dynamics in a stationary bivariate VAR framework. Their aim was to consider the reversed effects from unemployment back to economic activity. Despite the fact that their estimations seem to support Okun's conclusions, the bivariate system does not exhibit a clear-cut structural value for the elasticity between economic activity and unemployment. An implication is that the OC does not seem to be very robust in a dynamic perspective.

Finally, as it happens to both BC and WC, one typically faces a few important, data-related, problems. For example, while many econometric studies employ as measures of unemployment and GDP changes their deviations from some equilibrium values (i.e. "natural" unemployment rate and potential GDP, respectively), Okun's original analysis was in terms of growth rates (Okun, 1962). In turn, the contemporary re-formulation might entail many estimation biases, e.g. those related to the estimation of "natural" levels. Furthermore, one has to assume that the unemployment rate and GDP are stationary around a deterministic trend (which instead might be stochastic). For all those reasons, in the following we shall use one-period growth rates instead of deviations.

2.2 Theoretical explanations of aggregate labor-market regularities

Mainstream economic theory has been trying to explain the foregoing aggregate regularities in the familiar *equilibrium-cum-rationality* framework, building the explanation on the shoulders of hyper-rational, maximizing, representative worker and firm. Hence, any aggregate regularity is interpreted as the equilibrium outcome of some maximizing exercises carried out by such agents. Thus, even when the sign in the equilibrium correlation between any two aggregate variables (e.g. vacancy and unemployment rates) is derived from an intertemporal optimization problem, the hyper-rationality assumption allows one to compress the entire (infinite) stream of choices in a unique, simultaneous, decision implying non reversible, consistent, choices.

A paradigmatic example of such modeling strategy can be found within the theoretical literature aimed at micro-founding and explaining the BC. Suppose we start from a standard "matching model" (Pissarides, 2000; Blanchard and Diamond, 1989). Then, the total number of hires from unemployment (i.e. the number of matches) M in the economy can be given by $\varepsilon \cdot m(cU, V)$, where U is unemployment, V is the number of vacancies, c is search effectiveness of unemployed workers and ε is matching efficiency.

All search and matching, which in reality is an inherently dynamic process, is thus described in a static setting by means of a deterministic matching function m, which is assumed to be well-behaved, homogeneous of degree one, and increasing in both arguments. In equilibrium, given employment level N and the exogenous inflow rate into unemployment s, it is assumed that $sN = M = \varepsilon \cdot m(cU, V)$. Exploiting constant returns to scale, one thus gets a BC:

$$s = \varepsilon \cdot m(c \cdot u, v), \tag{1}$$

where $u = U/N$ and $v = V/N$ are unemployment and vacancy rates.

It is worthwhile noticing that the BC relationship is directly implied by the functional form and the parametric assumptions of the matching function m. In particular, the BC is treated here as a static (long-run) equilibrium locus in the $u - v$ space, requiring that all flows in and out of unemployment must always compensate[9]. Needless to say, this is at odds with any empirical observation.

Moreover, in order to get the desired results, many over-simplifying assumptions are required. *First*, the environment must be strictly stationary, ruling out any form of technological and organizational change, as well as any type of endogenous selection amongst firms and workers. *Second*, the presence of a hyper-rational, representative individual rules out the possibility of accounting for any form of heterogeneity across firms and workers. More than that: it excludes the very possibility of analyzing any *interaction* process among agents[10]. *Third*, as a consequence, one is prevented from studying the dynamic outcomes of multiple (reversible) decisions of hiring, firing, quitting, and searching which unfold over time.

Similar critiques also apply to the purported micro-foundations of Wage and Okun curves[11]. Consider the Wage curve first. Since a competitive equilibrium market framework cannot account for a downward sloping equilibrium relationship between wage and unemployment rates (Blanchflower and Oswald, 1994; Card, 1995), other frameworks departing from perfect competition have to be devised in order to provide a rationale for this robust piece of aggregate evidence. Models generating a WC belong to two strands. First, *bargaining models* build on the idea that higher levels of joblessness produce lower bargaining power for workers and thus a reduced ability to elicit some kind of surplus. This effect can be amplified by the existence of a union in the labor market. This interpretation employs *implicit contract theory* and assumes that a contract does not only consist of a wage level, but also of some implicit temporary insurance against unemployment. Second, *efficiency wage models* (Shapiro and Stiglitz, 1984) assume that unemployment functions as a "discipline device" for workers. Other things being equal, higher unemployment levels induce a higher probability of job loss. Therefore, rational employees should exert a higher effort in the high-unemployment equilibrium, even if they receive a lower wage.

Note that, in these alternative WC models, what varies are the assumptions on what causes the departures from the perfect competition set-up, but they all continue to share a *rationality-cum-equilibrium*, static framework. Similar considerations apply to the *state-of-the-art* of contemporary interpretations of the Okun curve. Also in this case, the evidence is hard to reconcile with the "pure" neoclassical view in which one assumes that markets always clear: in such a setting, there is no easy way to generate downward-sloping relationships between unemployment changes and economic activity.

---

[9] On the contrary, the model we present below allows the economy to evolve on a permanent disequilibrium path.

[10] In this respect, the far-reaching observations by Kirman (1992) on the pitfalls of any "representative agent" reduction of market interactions fully apply also to most contemporary models of the labor market.

[11] Cf. Hahn and Solow (1997) for a thorough discussion on this and related points.

Since only structural and frictional unemployment is allowed to exist, a negative relation between unemployment and GDP growth is hard to sustain, insofar as it is difficult to assume that structural or frictional unemployment declines in upswings and increases in downswings. In general, theoretical explanations must rely on a careful and often *ad hoc* modeling of expectation formation. For instance, one could assume that in an upswing people searching for a new job still hold low wage aspirations and are therefore more willing to take a particular job. This should result in shorter search times in upswings and lower unemployment[12].

Conversely, both a old-fashioned and a new Keynesian perspective allow us to explain Okun law in more straightforward ways. A possibility is to assume fixed prices and wages. Then, changes in aggregate demand induce firms to alter their output plans; labor demand changes and hence the unemployment rate is affected. Another possibility is to consider models of monopolistic competition (Blanchard and Kiyotaki, 1987) with menu costs (nominal rigidity) on the market for goods and real rigidities on labor market (e.g. efficiency wages): there, changes in aggregate demand can be easily shown to affect output and therefore unemployment[13].

Notwithstanding the existence of some competing, although not entirely persuasive, interpretations of each of the three aggregate regularities *taken in isolation*, the economic literature witnesses a dramatic lack of theories attempting *jointly to explain* Beveridge, Okun and Wage curves. The over-simplifying assumptions needed in order to derive analytically-solvable models (to repeat: hyper-rational, optimizing representative agents, static frameworks, commitment to equilibrium, etc.) strongly constrain the possibility of providing a unified theory of the interplay between the microeconomics of labor market dynamics and the macroeconomics of unemployment and economic activity.

In the following, we begin indeed to explore a radically different path and study the properties of a model in which the most stringent assumptions of standard formalizations are abandoned, and we explicitly account for the processes of out-of-equilibrium interaction among heterogeneous agents.

3 An evolutionary approach to labor market dynamics

3.1 The model

Consider an economy composed of F firms and N workers[14]. Time is discrete: $t = 0, 1, 2, ...$ and there is a homogeneous, perishable good g whose price is $p_t > 0$. In each period, a firm $i \in \{1, ..., F\}$ produces q_{it} units of good g using labor as the sole input under a constant returns to scale (CRTS) regime:

$$q_{it} = \alpha_{it} n_{it}, \tag{2}$$

---

[12] See also Aghion and Howitt (1994) and Schaik and Groot (1998) for attempts to explain the OC within the framework of endogenous growth models.

[13] An interesting by-product of this type of models is that productivity shocks can lead to OC as well. Indeed, GDP and employment move in the same direction as long as the effects of productivity shocks on efficiency-wages are not too strong.

[14] The ratio between the number of workers and the number of firms (N/F) can be interpreted as a measure of the concentration of economic activity.

where α_{it} is the current labor productivity of firm i and n_{it} is the number of workers hired at t by firm i. Workers are homogeneous as far as their skills are concerned. If the firm offers a contractual wage w_{it} to each worker, current profits are computed as:

$$\pi_{it} = p_t q_{it} - w_{it} n_{it} = (p_t \alpha_{it} - w_{it}) n_{it}. \tag{3}$$

Contractual wages offered by firms to workers are the result of both a matching and a bargaining process. We assume that any firm i has at time t a "satisficing" wage w_{it}^s it wants to offer to any worker. Similarly, any worker $j \in \{1, ..., N\}$ has at time t a "satisficing" wage w_{jt}^s which he wants to get from firms. Moreover, any worker j can only accept contractual wages if they are greater or equal to his *reservation wage* w_j^R, which we assume to be constant over time for simplicity.

We start by studying an economy where jobs last only one period. Hence, workers must search for a new job in any period. Job openings are equal to labor demand and, at the same time, to "ex-ante" vacancies. However, workers can be unemployed and firms might not satisfy their labor demand.

Let us turn now to a brief description of the flow of events in a generic time-period. We then move to a detailed account of each event separately.

Dynamics

Given the state of the system at the end of any time period $t-1$, the timing of events occurring in any time period t runs as follows.

1. Firms decide how many jobs they want to open in period t.
2. Workers search for a firm posting at least one job opening and queue up.
3. Job matching and bargaining occur: firms look in their queues and start bargaining with workers who have queued up (if any) to decide whether to hire them or not.
4. After hiring, production takes place according to eq. (2). Aggregate supply and demand are then formed simply by aggregating individual supplies and demands. Subsequently, a "pseudo-Walrasian" price setting occurs. We assume that the price of good g at t is given by:

$$p_t Q_t = W_t, \tag{4}$$

   where $Q_t = \sum_{i=1}^{F} q_{it}$ is aggregate (real) output and $W_t = \sum_{j=1}^{N} w_{jt}$ is total wage. Thus, total wage equals aggregate demand, as we assume that workers spend all their income to eat good g in any time period. Then, firms make profits:

$$\pi_{it} = (p_t \alpha_{it-1} - w_{it}) n_{it}.$$

5. Given profits, firms undergo a selection process: those making negative profits ($\pi_{it} < 0$) exit and are replaced by entrants, which, as a first approximation, are simply "average" firms (see below).
6. Firms and workers update their satisficing wages (w_{it-1}^s and w_{jt-1}^s).

7. Finally, technological progress (if any) takes place. We assume that in each period labor productivity may increase at rates which are exogenous but firm-specific (see below).

Job openings

At the beginning of period t, each firm creates a queue of job openings. Since in reality only *ex-ante* vacancies (i.e. new job positions) can be empirically observed, we will employ throughout the term job openings as a synonym of (ex-ante) vacancies. “Ex-post” vacancies will be computed as the number of unfilled job-openings.

Let us then call v_{it} the number of new positions opened by firm i at time t. As far as the firm’s decision about how many vacancies to open is concerned, we experiment with two alternative “behavioral” scenarios.

In the first one, a firm simply observes current (i.e. time $t-1$) price, quantity produced and contractual wage offered, and sets vacancies v_{it} as:

$$v_{it} = \overline{v}_{it-1} = \left\lceil \frac{p_{t-1} q_{it-1}}{w_{it-1}} \right\rceil , \tag{5}$$

that is, it creates a queue with a number of open slots equal to the “ceiling” of (i.e. the smallest integer larger than) the ratio between revenues and the contractual wage offered in the last period. We call this job opening scenario the “**Wild Market Archetype**”, in that no history-inherited institution or behavioral feature is built into the model.

In the second “behavioral” scenario (which we shall call the “**Weak Path-Dependence**” scenario), we introduce some rather mild path-dependence into the vacancy setting. We suppose that: (a) jobs opened by any firm at time t are a non-decreasing function of last-experienced profits growth rate; and (b) cannot exceed $\overline{v}_{it-1}$. More formally:

$$v_{it} = \min\{\overline{v}_{it-1}, v^*_{it}\}, \tag{6}$$

and:

$$v^*_{it} = \begin{cases} \lceil v_{it-1}(1+|X|) \rceil \text{, if } \frac{\Delta\pi_{it-1}}{\pi_{it-1}} \geq 0 \\ \lceil v_{it-1}(1-|X|) \rceil \text{, if } \frac{\Delta\pi_{it-1}}{\pi_{it-1}} < 0 \end{cases}, \tag{7}$$

where X is an i.i.d. random variable, normally distributed with mean zero and variance $\sigma_v^2 > 0$, and $\lceil x \rceil$ denotes the ceiling of x. Notice that the higher σ_v, the more firms react to any given profits growth rate by enlarging or shrinking their current queue size. Hence a *higher* σ_v implies *higher* sensitivity to market signals. Notice that, in both scenarios, firms always open at least one vacancy in each period.

Job search

In our model, workers can visit in any time period only one firm. Similarly to job opening, we consider two “behavioral” scenarios for the job search procedure employed by workers to find a firm that has just opened new job positions. In the first one, called "**No Search Inertia**", each worker j simply visits any firm i in the market with a probability proportional to the last contractual wage w_{it-1} offered. If the selected firm has places still available in the queue, the worker gets in and demands a wage equal to the "satisficing" one, i.e. w^s_{jt-1}.

In the second scenario, which we label "**Search Inertia**", we introduce some stickiness (loyalty) in firm visiting. If worker j was employed by firm i in period $t-1$, he visits first firm i. If i still has places available in the queue, the worker gets in and demands w^s_{jt-1}. Otherwise, the worker employs the random rule above ("No Search Inertia") to select among the remaining $F-1$ firms.

In both scenarios, a worker becomes unemployed if he chooses a firm that has already filled all available slots in its queue.

Job matching and bargaining

After workers have queued up, firms start exploring workers wage demands to match them with their *desiderata*. Suppose that, at time t, firm i observes $0 < m_{it} \le N$ workers in its queue. Then, it will compute the average wage demanded by those workers:

$$\overline{w}_{it} = \frac{1}{m_{it}} \sum_{h=1}^{m_{it}} w^s_{j_h t-1}, \tag{8}$$

where j_h are the labels of workers in i's queue. Next, it sets the contractual wage for period t as a linear combination of $\overline{w}_{it}$ and the satisficing wage w^s_{it-1}. Thus:

$$w_{it} = \beta w^s_{it-1} + (1-\beta)\overline{w}_{it}, \tag{9}$$

where $\beta \in [0,1]$ is an institutional parameter governing firms' strength in wage bargaining. A higher β implies a higher strength on the side of the firm in wage setting. If $\beta = 0$, firms just set contractual wage as the average of wages demanded by workers in the queue. If $\beta = 1$, firms do not take into account at all workers' *desiderata*.

Once the firm has set the contractual wage at which it is willing to hire workers in the queue, any worker j in the queue will accept the job only if w_{it} exceeds the reservation wage w^R_j.

As soon as a worker j accepts the job, he temporarily changes his satisficing wage to keep up with the new (actual) wage earned, i.e. $w^s_{jt-1} = w_{it}$. Similarly, a firm who has filled at least a job opening will replace w^s_{it-1} with w_{it}[15].

Given the number of workers n_{it} hired by each firm, production, as well as price setting and profits determination occur as explained above. *Ex-post* firm i's vacancies are defined as $\widetilde{v}_{it} = m_{it} - n_{it}$.

Selection, exit, and entry

---

[15] These new values of satisfying wages will then be employed in the updating process. Since satisfying wage can be interpreted as (myopic) expectations, satisfying wage updating plays in the model the role of expectation formation process.

Suppose that - given the new contractual wage, price p_t, and current productivity α_{it-1} - firm j faces negative profits, i.e. $p_t \alpha_{it-1} < w_{it}$. Then selection pressure makes firm j exit the market.

Each exiting firm is replaced by a new firm which starts out with the average "characteristics" of those firms still in the market at t (i.e. those making non-negative profits)[16]. Notice that this entry-exit process allows to keep an invariant number of F firms in the economy at each t.

Satisficing wages updating

Surviving firms, as well as the N workers, will then have the opportunity to revise their satisficing wage according to their perceptions about the outcome of market dynamics.

- **Firms**: We assume that each firm has an invariant desired ratio of filled to opened jobs $\rho_i \in (0, 1]$ which it compares to the current ratio:

  $$r_{it} = \frac{n_{it}}{v_{it}}.$$

  If firm i hired too few workers (as compared to the number of job positions it has decided to open), then it might want to increase the wage it is willing to offer to workers. Otherwise, it might want to decrease it. We capture this simple rule by positing that:

  $$w_{it}^s = \begin{cases} w_{it-1}^s(1 + |Y|) \text{ if } r_{it} < \rho_i \\ w_{it-1}^s(1 - |Y|) \text{ if } r_{it} \geq \rho_i \end{cases}, \tag{10}$$

  where Y is an i.i.d. random variable distributed as a standard normal. Notice that w_{it-1}^s is equal to w_{it} (i.e. contractual wage just offered) if the firm has hired at least one worker.
- **Workers**: If worker j remains unemployed after matching and bargaining, he might want to reduce his satisficing wage (without violating the reservation wage threshold). Otherwise, he might want to demand a higher wage during the next bargaining session. We then assume that:

  $$w_{jt}^s = \begin{cases} \max\{w_j^R, w_{jt-1}^s(1 - |Y|)\} & \text{if } j \text{ unemployed} \\ w_{jt-1}^s(1 + |Y|) & \text{if } j \text{ employed} \end{cases}, \tag{11}$$

  where Y is an i.i.d. random variable distributed as a standard normal. Again, $w_{jt-1}^s = w_{jt}$ if j has been just hired.

Technological progress

The last major ingredient of the model regards labor productivity dynamics. Here, we experiment with two "technological scenarios". In the first one ("**No Technological Progress**"), we study a system where labor productivity does not change

---

[16] All results we present in the next Section are robust to alternative assumptions concerning entry and exit.

Table 1. System Parameters

Parameter	Range	Meaning
N/F	R_{++}	Concentration of economic activity (Number of Workers / Number of Firms)
σ_v	R_{++}	Sensitivity to market signals in vacancy settings (only in a Weak Path-Dependence Scenario)
β	$[0,1]$	Labor-market institutional parameter governing the strength of firms in wage-setting
σ_Z	R_+	Technological parameter tuning the availability of opportunities in the system ($=0$ means no technological progress)

through time (i.e. $\alpha_{it} = \alpha_i, \forall i$)[17]. In the second scenario ("**Technological Progress**"), we allow for an exogenous, albeit firm-specific, dynamics of labor productivities. We start with initially homogeneous labor coefficients ($\alpha_{i0} = \alpha$) and we let them grow stochastically over time according to the following multiplicative process:

$$\alpha_{it} = \alpha_{it-1}(1+Z), \tag{12}$$

where Z, conditionally on $Z > 0$, is an i.i.d. normally distributed random variable with mean 0 and variance $\sigma_Z^2 \geq 0$[18]. The latter governs the opportunity setting in the economy. The larger σ_Z, the more likely firms draw large productivity improvements. Notice that if we let $\sigma_Z = 0$ we recover the "**No Technological Progress**" scenario.

3.2 Initial conditions, micro- and macro-dynamics

The foregoing model, as mentioned, genuinely belongs to an evolutionary/ACE approach. Given its behavioral, bottom-up, perspective, one must resort to computer simulations to explore the behavior of the system[19]. One of the main goals is to look for meta-stable properties (and rarely to equilibria in the traditional sense) which emerge as the result of the co-evolution among individual behaviors over time and persist for sufficiently long time spans.

In our model, the dynamics of the system depends on four sets of factors. *First*, we distinguish behavioral (e.g. concerning job opening and job search) and technological scenarios. We call such discrete institutional and technological regimes "system setups". *Second*, a choice of system parameters (F/N, σ_v, β, σ_Z) is required (see Table 1).

---

[17] Labor productivity may in turn be either homogeneous across firms ($\alpha_i = \alpha$) or not.

[18] Hence, there is a probability 0.5 to draw a neutral labor productivity shock ($Z = 0$), while positive shocks are distributed as the positive half of a $N(0,1)$.

[19] Simulation code is written in C++ and is available from the Authors upon request.

Third, one should explore the would-be importance of different initial conditions[20]. Since simulations show that the latter do not dramatically affect the long-run properties of aggregate variables, we typically define a “canonical” set of initial conditions. All results presented below refer to this benchmark choice. Finally, individual updating by firms and workers induces a stochastic dynamics on micro-variables (e.g. contractual wages, desired production, desired employment, etc.). By aggregating these individual variables over firms and workers, one can study the properties of macro-dynamics for the variables of interest. We will focus on unemployment:

$$U_t = N - \sum_{i=1}^{F} n_{it}, \tag{13}$$

vacancies:

$$V_t = \sum_{i=1}^{F} v_{it}, \tag{14}$$

output price p_t, total wages:

$$W_t = \sum_{j=1}^{N} w_{jt}, \tag{15}$$

and (real) GDP:

$$Q_t = \sum_{i=1}^{F} q_{it}, \tag{16}$$

as well as its growth rate:

$$h_t = \varDelta \log(Q_t). \tag{17}$$

Related literature on matching and labor-market dynamics: a necessary digression

One of the key features of the foregoing model is an explicit microfoundation - within an evolutionary framework - of labor market dynamics regarding the processes governing e.g. job opening, job search, matching, bargaining, and wage setting.

Standard theoretical literature on matching in labor markets, as mentioned above, has typically abstracted from any explicit account of decentralized interaction patterns. For example, matching models based upon a “search equilibrium” framework[21], while stressing the existence of frictions and imperfect information

---

[20] In the model this implies defining initial values $(n_{i0}, \alpha_{i0}, w^s_{i0}, w_{i0})_{i=1}^F$ for firms and $(w^s_{j0})_{j=1}^N$ for workers. Moreover, an initial price p_0, and some distributions for desired ratios $(\rho_i)_{i=1}^F$ and reservation wages $(w^R_j)_{j=1}^N$ have to be chosen.

[21] See *inter alia* Pissarides (2000), Petrongolo and Pissarides (2001), Mortensen (1986) and Mortensen and Pissarides (1994).

in labor markets, have implicitly assumed a sort of centralized, equilibrium, device matching the "representative firm" and the "representative worker" (eq. 1 stands precisely for that). Wage setting is then often assumed to be a Nash bargaining process. Given these strong assumptions, as well as the restrictions on the shape of the matching function itself, it is not surprising that the model delivers, e.g., Beveridge curves.

The bottom line of the exercises belonging to the "pure equilibrium" *genre* is that they turn out to be unable, almost by construction, to account for involuntary unemployment or even endogenous changes in the "equilibrium" rates of unemployment. Important advances, incrementally departing from the standard model, have tried to incorporate agents' informational limitations, in order to account for phenomena such as endogenous fluctuations in aggregate activity and persistent involuntary unemployment (see e.g. the seminal work by Phelps and Winter, 1970 and Phelps, 1994).

More recently, some efforts have been made to depart from exogenous and deterministic matching devices and assume some "endogenous matching" mechanism to describe the (Walrasian) decentralized process governing the meetings between firms and workers in the labor market[22]. For instance, Lagos (2000) studies an ex-ante frictionless and random decentralized matching process, while Peters (1991) describes wage offers as a sequential game with incomplete information where firm strategies can influence the search behavior of the workers. The main goal of these contributions is to study under what conditions a centralized, well-behaved, matching function can be ex-post generated, in equilibrium, by some decentralized, endogenous matching function. An important conclusion is that, if such centralized matching device exists, then its properties heavily depend on the fine details of market organization and institutional setups (and thus also on policy interventions).

This is certainly a point our model takes on board in its full importance, and it does so through an explicit account of the (disequilibrium) unfolding of the interaction process. In this respect, our model has three important antecedents in labor market literature. *First*, the out-of-equilibrium, interaction-based perspective that we pursue is a distinctive feature of "self-organization" labor market models[23]. They assume heterogeneous, boundedly rational workers and firms meeting at random over time in institutionally-shaped labor markets. For given institutional arrangements, the system self-organizes in long-run configurations where different unemployment and wage levels emerge as the result of individual choices and interactions. *Second*, the ACE model in Tesfatsion (2001) also assumes many heterogenous, interacting agents, characterized by "internal states" and behavioral rules, who exchange information in the market. Matching occurs in a decentralized way through a one-sided offer auction and individual work-site payoffs are modeled as in a Prisoner-Dilemma game. *Third*, Aoki (2003) extends the ACE model of fluctuations and growth proposed in Aoki and Yoshikawa (2003) to allow for unemployment dynamics. Similarly to our model, co-evolution between product

---

[22] See Lagos (2000), Peters (1991), Cao and Shi (2000), Burdett et al. (2001), Smith and Zenou (2003) and Julien et al. (2000).

[23] Cf. Lesourne (1992) and Laffond and Lesourne (2000). Self-organizing processes are discussed in Witt (1985).

and labor market dynamics is explicitly taken into account and simulations allow to reproduce (albeit in some benchmark parameterizations) Okun curves. However, matching and wage bargaining are not incorporated in the model as endogenous processes. Therefore, no implications about wage and Phillips curves can be derived from simulation exercises.

Notwithstanding many overlappings with "self-organization" and ACE formalizations, our model proposes advances, *vis-à-vis* the state of the art in this area, on at least four levels. *First*, it accounts for the co-evolutionary dynamics between the labor market and the product market. More specifically, we try to nest labor market interactions in what one could call a "general disequilibrium" framework with endogenous aggregate demand. This feature allows us to study market properties associated with an endogenous business cycle. *Second*, we explicitly model (as endogenous processes) job opening, matching, wage bargaining, and wage setting. *Third*, we allow for technical progress and the ensuing macroeconomic growth. *Fourth*, in the analysis of the results, we go beyond an "exercise in plausibility" and we explicitly compare the statistical properties of the simulated environments with empirically observed ones, specifically with respect to the emergence of Beveridge, Wage, and Okun curves.

4 Simulation results

The general strategy of our simulation experiments runs as follows. First, we attempt to identify some general conditions (i.e. setups and parameters choices) under which the model is able *jointly* to replicate the three aggregate regularities characterizing labor markets dynamics and economic activity discussed in Section 2.

Second, in order to wash out stochastic effects in micro- and macro-dynamics specific to single sample paths, we perform Monte Carlo exercises so as to understand how the statistical properties of labor-market dynamics and economic activity change across different parameterizations and setups.

4.1 Simulation setups

All simulation exercises we present in the paper refer to (and compare) the following behavioral and institutional scenarios, and combinations thereof:

1. **Walrasian Archetype** (**WA**): This economy is characterized by the "**Wild Market Archetype**" scenario as far as job opening is concerned and the "**No Search Inertia**" scenario for workers' job search. In this world, there is no path-dependence in job openings, nor in job search. Workers visit firms at random, while the latter open a number of new positions in each period without being influenced by past experienced profits.
2. **Institutionally-Shaped Environment** (**ISE**): In this economy, workers and firms face some path-dependence in job opening and job searching. We assume that firms open new job positions within a "**Weak Path-Dependence**" scenario (i.e. they adjust job openings according to last profits growth), while workers

Table 2. System Setups

Setups	Label	Job Opening	Job Search	Tech. Progress
1	Walrasian Archetype w/o Tech. Progress	Wild Market Archetype	No Search Inertia	NO
2	Walrasian Archetype w/ Tech. Progress	Wild Market Archetype	No Search Inertia	YES
3	Institutionally- Shaped Environment w/o Tech. Progress	Weak Path-Dependence	Search Inertia	NO
4	Institutionally- Shaped Environment w/ Tech. Progress	Weak Path-Dependence	Search Inertia	YES

search for a firm under the "**Search Inertia**" scenario (i.e. they try to stick to the last firm in which they were employed).

Each of the two foregoing behavioral choices can be associated with a different technological scenario (with or without technological change), in order to define a "system setup". Table 2 summarizes the four "worlds" which we extensively explore in our simulation exercises[24].

4.2 Some qualitative evidence

We start by investigating from a qualitative perspective the emergence of Beveridge, Wage, and Okun curves in an economy characterized by the "Walrasian Archetype".

In this world where agents decide myopically and do not carry over past information, the system does not allow the recovery of any aggregate, statistically significant, negative relationship between vacancy and unemployment rates. Simulations show that, irrespective of the technological scenario, the Beveridge curve does not emerge (cf. Figs. 1 and 2) in a large region of the system parameters (F/N, β, σ_Z) space.

Notwithstanding the fact that matching and search do not seem to affect the (u, v) relation, the unemployment rate turns out to be negatively related to wage levels. Moreover, higher unemployment growth entails smaller GDP growth. Therefore, both Wage and Okun curves robustly emerge no matter whether technological progress is shut down or not. Notice that if $\sigma_Z = 0$, the economy works as a dynamic allocation device trying to match in a decentralized and imperfect way individual labor demand and supply *for given resources*. It is then easy to see that both Okun and Wage relationships are a consequence (and not an emergent property) of the joint assumptions of quasi-Walrasian price-setting and constant returns to scale. Indeed, from (2) and (4), one gets: $W_t = -p_t U_t + p_t(N - N_t + \sum_i \alpha_i n_{it})$ and $Q_t = -U_t + (N - N_t + \sum_i \alpha_i n_{it})$. Thus, if $\alpha_i \neq \alpha$, both curves are implied by the

---

[24] In all exercises that follow, we set the econometric sample size $T = 1000$. This time span is sufficient to allow for convergence of the recursive moments for all variables under study.

Table 3. Emergence of Aggregate Regularities: Qualitative Results. (*) The associated aggregate regularity can be (partly) explained by the assumptions made in the model about micro-behaviors.

	Tech.	Aggregate Regularity			
Setup	Change	Beveridge	Wage	Phillips	Okun
WA	No	No	Yes*	No	Yes*
WA	Yes	No	Yes	No	Yes
ISE	No	Yes	Yes*	No	Yes*
ISE	Yes	Yes	Yes	No	Yes

assumptions. In particular, one should observe a unit coefficient for the wage curve. If labor productivities are heterogeneous, one should instead observe for both WC and OC some noise around negatively sloped lines.

If, on the contrary, technological progress occurs in a WA scenario, there is no apparent reason to expect both OC and WC to emerge robustly. Yet, as simulations show, they both characterize system dynamics for a large region of the parameter space, even if no path-dependent behavior drives the economy (cf. Figs. 3 and 4).

Consider now an economy in which firms are influenced by past profits when they adjust vacancies and workers try to stick to previous employers (i.e. what we call an "Institutionally-Shaped Environment"). Then, irrespective of the technological regime, the model is able robustly to generate Beveridge curves with statistically significant (negative) slopes: see Figs. 5 and 6. Furthermore, when technological progress is present, both Wage and Okun curves still characterize macro-dynamics as robust, emergent, properties of the system, cf. Figs. 7 and 8.

Table 3 summarizes our main qualitative results about the emergence of aggregate regularities. Notice that some path-dependence seems to be a necessary condition for a Beveridge relationship. Moreover, a standard Okun curve seems to be in place even when technological progress persistently boosts available production capacity.

Finally, despite persistent heterogeneity arising endogenously from labor productivity dynamics, Phillips-curve type of regularities are typically rejected by the simulated data in favor of a Wage curve relationship.

4.3 Monte Carlo experiments

In the last section, we singled out some broad behavioral and technological conditions under which aggregate regularities of interest emerge for a sufficiently large sub-region of the parameter space. We now turn to a more detailed and quantitative study addressing the robustness of emergence results. We present here two sets of exercises.

First, we study whether the implications summarized in Table 3 are robust, for any given parametrization, across independent realizations (i.e. time-series). To this end, in each of the four main "setups" under study, we identify a "benchmark" setting for system parameters, and we generate M independent (Monte Carlo) simulations. We then study the moments of the distributions of the statistics of

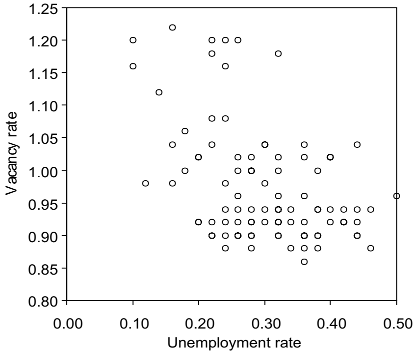


Fig. 1. Vacancy vs. Unemployment Rate in a "Walrasian Archetype" Economy **without** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5$.

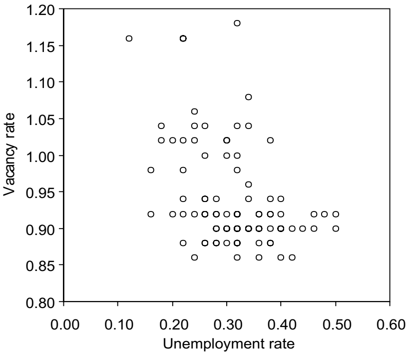


Fig. 2. Vacancy vs. Unemployment Rate in a "Walrasian Archetype" Economy **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$.

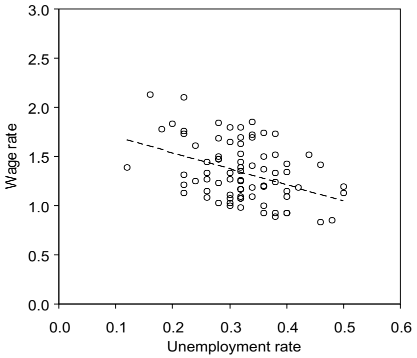


Fig. 3. Emergence of Wage curve in a "Walrasian Archetype" Economy **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$.

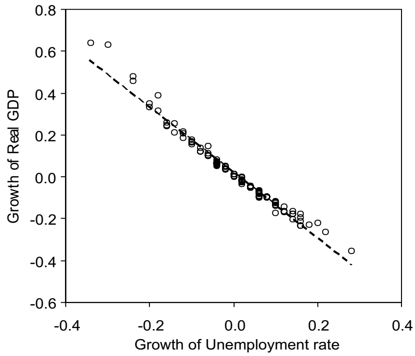


Fig. 4. Emergence of Okun curve in a "Walrasian Archetype" Economy **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$.

interest. We focus in particular on test statistics for the significance of coefficients in Beveridge and Okun regressions, the magnitude of the Okun coefficient, as well as test statistics discriminating between Wage and Phillips curves.

Second, we will perform some simple "comparative dynamics" exercises to investigate what happens to emergent regularities when one tunes system parameters within each "setup". We are in particular interested in detecting shifts (if any) in the Beveridge curve and changes in the Okun coefficients. Once again, we will discuss the outcome of Monte Carlo statistics coming from independent time-series simulation runs for any given parametrization[25].

Emergence of aggregate regularities: robustness tests

---

[25] All Monte Carlo experiments are undertaken using a Monte Carlo sample size $M = 100$. Initial conditions are always kept fixed (see above).

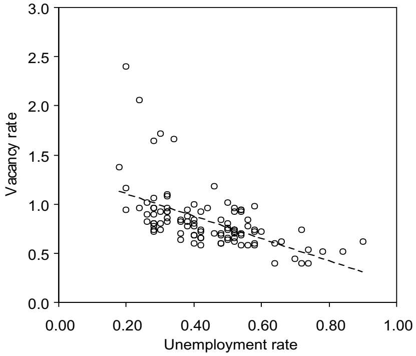


Fig. 5. Emergence of Beveridge curve in a "Institutionally-Shaped" Environment **without** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5$.

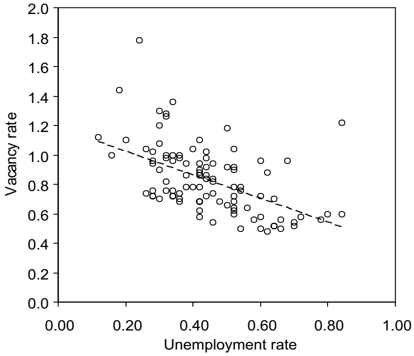


Fig. 6. Emergence of Beveridge curve in a "Institutionally-Shaped" Environment **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5, \sigma_Z = 0.1$.

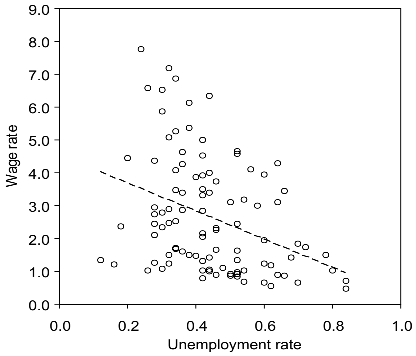


Fig. 7. Emergence of Wage curve in a "Institutionally-Shaped" Environment **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5, \sigma_Z = 0.1$.

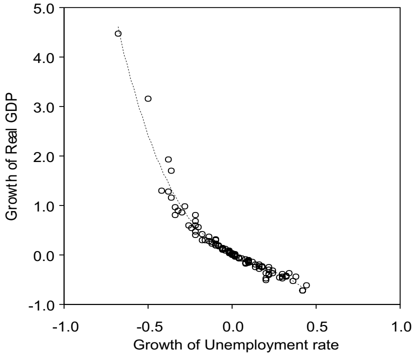


Fig. 8. Emergence of Okun curve in a "Institutionally-Shaped" Environment **with** Technological Progress. Parameters: $N/F = 5$, $\beta = 0.5, \sigma_Z = 0.1$.

To begin with, consider the emergence of Beveridge curves. Suppose that, for any setup under analysis, a benchmark parametrization under which the results in Table 3 hold is given. Following existing empirical literature, we computed, for each of M independent simulated time-series, estimates (and R^2) for the simple time-series regression:

$$u_t = b_0 + b_1 v_t + \epsilon_t, \tag{18}$$

where ϵ_t is white-noise, u_t is the unemployment rate and v_t is the vacancy rate (both defined as activity rates). We also performed two-tailed test statistics for the null hypothesis $b_1 = 0$ and computed the percentage of rejections (i.e. frequency of emergence of Beveridge curve, in case of a negative estimate). We then computed the Monte Carlo average and standard deviation of estimates $\widehat{b_1}$, of their standard errors $\sigma(\widehat{b_1})$ and goodness-of-fit R^2, together with the maximum value of the distribution of tail-probabilities for the test $b_1 = 0$.

Table 4. Emergence of the Beveridge curve in alternative setups. WA = "Walrasian Archetype". ISE= "Institutionally- Shaped Environment". Estimation of $u_t = b_0 + b_1 v_t + \epsilon_t$. Monte Carlo Standard Errors in parentheses. Monte Carlo sample size $M = 100$. Benchmark parametrization: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$ (when > 0), $\sigma_v = 0.1$ (under ISE).

	Setups			
	WA		ISE	
	$\sigma_Z = 0$	$\sigma_Z > 0$	$\sigma_Z = 0$	$\sigma_Z > 0$
MC Average of $\widehat{b_1}$	-0.176 (0.095)	-0.263 (0.178)	-0.422 (0.072)	-0.524 (0.078)
MC Average of $\sigma(\widehat{b_1})$	0.118 (0.015)	0.158 (0.013)	0.055 (0.007)	0.053 (0.007)
R^2	0.024 (0.011)	0.039 (0.016)	0.375 (0.066)	0.431 (0.082)
1st Quartile of Tail Prob. Distr. for H_0:b_1=0	0.081	0.050	*	*
3rd Quartile of Tail Prob. Distr. for H_0:b_1=0	*	*	0.000	0.000
Percentage of rejections (H_0:b_1=0) at 5%	10%	25%	100%	100%

As Table 4 shows, estimates for the Beveridge coefficient are, on average, negative. In more detail, the "institutionally-shaped environment" entails a 100% percentage of rejections for the test (i.e. a statistically significant Beveridge curve always emerges). However, when a WA is assumed, the frequency of rejections dramatically decreases. In this case, the distribution of tail probabilities is considerably shifted to the right as compared to a WA economy. This means that the emergence of a Beveridge curve in a WA economy may be considered as a quite rare event. This result is also confirmed by looking at goodness-of-fit: average R^2 are much lower in the WA case than in the ISE. Furthermore, the dispersion of the Monte Carlo distribution of estimates increases when one moves towards an "institutionally-shaped" system. Interestingly enough, the presence of technological progress seems to allow for an even more robust emergence of a BC: when $\sigma_Z > 0$, R^2 are higher and the average magnitude of the coefficient increases.

While the Beveridge curve tends to emerge robustly only in an "institutionally-shaped" economy, simulations show that a Wage curve always characterizes our system in all four setups. In particular, statistical tests aimed at discriminating between a Phillips and a Wage world, show that the latter is almost always preferred. Following Card (1995), we perform the lagged regression:

$$\Delta \log \widetilde{W_t} = g_t + a_1 \log u_t + a_2 \log u_{t-1} + \Delta e_t, \tag{19}$$

where $\widetilde{W_t}$ is the wage rate, u_t is the unemployment rate, g_t is a time trend, and first-differences are taken to avoid serial correlation in e_t. As Card (1995) shows, the Wage curve hypothesis implies $a_1 = -a_2$ (together with $a_1 < 0$), while the Phillips curve hypothesis requires $a_2 = 0$. Table 5 reports Monte Carlo testing exercises in our four setups. Notice that the percentage of rejections of a Phillips world is quite high, while we tend not to reject the hypothesis that wage levels are negatively correlated with unemployment rates in almost all simulations.

The R^2 is very high in all setups. This might be an expected result when $\sigma_Z = 0$, because, without technological progress, a Wage curve follows from price-setting and constant returns. However, when $\sigma_Z > 0$, the goodness-of-fit remains high (and standard errors very low). Our model seems to allow for well-behaved Wage curves also when technological progress induces persistent heterogeneity in labor productivity dynamics. Furthermore, a quite general and robust result (see also below) concerns the effect of technological progress upon the slope of the curve. As discussed above, the latter is expected to be around -1.0 when $\sigma_Z = 0$, but nothing can in principle be said about the expected slope when $\sigma_Z > 0$. Our results suggest that, even when technological progress is present, the Wage curve robustly emerges. Indeed, wage rates become even more responsive to unemployment than in the $\sigma_Z = 0$ case.

As with the Wage curve, the Okun curve, too, turns out to be a robust outcome of our labor market dynamics. Evidence of this effect simply appears by linearly regressing GDP growth rates against changes in the rates of unemployment:

$$\Delta \log(Q_t) = c_0 + c_1 \Delta \log(u_t) + \epsilon_t. \tag{20}$$

We computed Monte Carlo estimates of the Okun coefficient c_1 and we tested for $H_0 : c_1 = 0$ (i.e. emergence of an Okun curve - as long as $c_1 < 0$), see Table 6. Our economy allows for an Okun relationship in all settings, especially when technological progress is present. Again, this might be considered as a not-too-surprising result when $\sigma_Z = 0$, but it becomes a truly emergent property when $\sigma_Z > 0$.

Table 5. Emergence of the Wage curve in alternative setups. WA = "Walrasian Archetype". ISE= "Institutionally- Shaped Environment". Functional form tested: $\Delta \log \widetilde{W}_t = g_t + a_1 \log u_t + a_2 \log u_{t-1} + \Delta e_t$. Rejecting Phillips curve hypothesis means rejecting $H_o' : a_2 = 0$. Rejecting Wage curve hypothesis means rejecting $H_o' : a_1 = -a_2$. Monte Carlo Standard Errors in parentheses. Monte Carlo sample size $M = 100$. Benchmark parametrization: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$ (when > 0), $\sigma_v = 0.1$ (under ISE).

	Setups			
	WA		ISE	
	$\sigma_Z = 0$	$\sigma_Z > 0$	$\sigma_Z = 0$	$\sigma_Z > 0$
MC Average of $\widehat{a_1}$	-0.814 (0.025)	-1.643 (0.093)	-1.019 (0.072)	-2.329 (0.225)
MC Average of $\widehat{a_2}$	0.781 (0.019)	1.520 (0.083)	0.977 (0.020)	2.134 (0.169)
R^2	0.985 (0.003)	0.906 (0.023)	0.978 (0.017)	0.914 (0.026)
% of rejections (H_0 : a_2=0) at 5%	100%	99%	99%	100%
% of rejections (H_0 : a_1=$-a_2$) at 5%	10%	5%	5%	1%

Table 6. Emergence of the Okun curve in alternative setups. WA = "Walrasian Archetype". ISE= "Institutionally- Shaped Environment". Estimation of $\Delta \log(Q_t) = c_0 + c_1 \Delta \log(u_t) + \epsilon_t$. Monte Carlo Standard Errors in parentheses. Monte Carlo sample size $M = 100$. Benchmark parametrization: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$ (when > 0), $\sigma_v = 0.1$ (under ISE).

	Setups			
	WA		ISE	
	$\sigma_Z = 0$	$\sigma_Z > 0$	$\sigma_Z = 0$	$\sigma_Z > 0$
MC Average of $\widehat{c_1}$	-2.064 (0.042)	-2.196 (0.047)	-2.635 (0.068)	-3.072 (0.063)
R^2	0.939 (0.026)	0.925 (0.060)	0.928 (0.064)	0.936 (0.025)
Max of Tail Prob. Distrib. for $H_0 : c_1$=0	0.000	0.001	0.000	0.001
% of rejections ($H_0 : c_1$=0) at 5%	100%	99%	100%	99%

The absolute value of the Okun coefficient is larger than one (and indeed close to Attfield and Silverstone's empirical estimates), implying some emergent aggregate dynamic increasing returns to labor. The effect becomes stronger when an ISE is assumed: Monte Carlo averages of the Okun coefficient range from -2.196 to -3.072.

Notice that one did not assume any increasing returns regime at the individual firm level. In fact, firms produce using constant returns production functions; see (2). Moreover, no Phillips curve relationships is in place: our economy typically displays a negative relationship between unemployment rates and wage *levels*. This suggests that aggregation of imperfect and persistently heterogeneous behaviors leads to macro-economic dynamic properties that were not present at the individual level. Therefore, aggregate dynamic increasing returns emerge as the outcome of aggregation of dynamic, interdependent, microeconomic patterns (Forni and Lippi, 1997).

Some comparative dynamics Monte Carlo exercises

We turn now to a comparative dynamics Monte Carlo investigation of the effect of system parameters on emergent aggregate regularities. We focus on the "institutionally-shaped" setup, wherein the economy robustly exhibits well-behaved Beveridge, Wage, and Okun curves, and we study what happens under alternative parameter settings. In particular, we compare parameter setups characterized by:

1. low vs. high N/F ratio (i.e. degrees of concentration of economic activity);
2. low vs. high σ_v (i.e. sensitivity to market signals in the way firms set their vacancies);
3. low vs. high β (i.e. firms' bargaining strength in wage setting);
4. low vs. high σ_Z (technological opportunities).

We *first* ask whether a higher sensitivity to market signals in vacancy setting induce detectable shifts in aggregate regularities. As Table 7 shows, the smaller σ_v, the stronger the revealed increasing dynamic returns: GDP growth becomes more responsive to unemployment growth and the Okun curve becomes steeper. Notice that σ_v can also be interpreted as an inverse measure of path-dependence in firms' vacancy setting. The smaller σ_v, the more firms tend to stick to last-period job openings. Therefore, a *smaller* path-dependence implies a steeper Okun relation.

Analogously, we investigate the impact on the BC of simultaneously increasing N/F (i.e. increasing N for a given F) and σ_v (i.e. firms' "sensitivity to market signals"). Notice that a higher concentration allows firms, *ceteris paribus*, to more easily fill their vacancies. Similarly, the higher σ_v, the more firms are able to react to aggregate conditions and correspondingly adjust vacancies. Therefore, one might be tempted to interpret economies characterized by high values for both N/F and σ_v as "low friction" worlds, and expect the BC curve to lie closer to the axes. Notice, however, that, in our model, an "indirect" effect is also present. If labor demand is very low (e.g. because the economy is in a recession), then the unemployment rate might be high, irrespective of the value of N/F. Moreover, if σ_v is high, firms will fire more workers during downswings, thus inducing a sort of "accelerator" effect on the recession. Thus, the consequences on the BC of assuming a larger market concentration and a higher sensitivity to market signals are *ex-ante* ambiguous: if "indirect" effects dominate, we should observe various combinations between shifts to the right and "business-cycle" movements along the curve.

Notwithstanding all that, Monte Carlo simulations show that the model is able to reproduce the predicted shifts in the BC. We observe (cf. Table 8) that, as N/F and σ_v both increase in an ISE economy, Monte Carlo averages of estimated intercepts stay constant, while the BC becomes, on average, steeper (and thus closer to the origin). A steeper BC implies that firms adaptively learn to open less vacancies and to adjust their filled-to-open vacancy ratios in response to market signals.

Table 7. Shifts in the Okun coefficient in an "Institutionally- Shaped Environment" under alternative parameter settings. HSMS: High Sensitivity to Market Signals. LSMS: Low Sensitivity to Market Signals. Estimation of $\Delta \log(Q_t) = c_0 + c_1 \Delta \log(u_t) + \epsilon_t$. Monte Carlo Standard Errors in parentheses. Monte Carlo sample size $M = 100$. Benchmark parametrization: $N/F = 5$, $\beta = 0.5$, $\sigma_Z = 0.1$.

	ISE Setup			
	$\sigma_v = 1.0$ (HSMS)		$\sigma_v = 0.2$ (LSMS)	
	$\sigma_Z = 0$	$\sigma_Z > 0$	$\sigma_Z = 0$	$\sigma_Z > 0$
MC Average of $\widehat{c_1}$	-2.700 (0.082)	-2.960 (0.085)	-2.900 (0.064)	-3.270 (0.060)
R^2	0.928 (0.064)	0.936 (0.025)	0.939 (0.026)	0.925 (0.060)
Max of Tail Prob. Distrib. for $H_0 : c_1$=0	0.001	0.001	0.000	0.001
% of rejections ($H_0 : c_1$=0) at 5%	100%	99%	100%	99%

Table 8. Shifts in the Beveridge curve in an "Institutionally- Shaped Environment" under alternative parameter settings for: (i) concentration of economic activity N/F; (ii) sensitivity to market signals σ_v. Estimation of $u_t = b_0 + b_1 v_t + \epsilon_t$. Monte Carlo Standard Errors in parentheses. Monte Carlo sample size $M = 100$. Benchmark parametrization: $\beta = 0.5$. No technical progress is assumed to focus on BC shifts for given resources.

	Parameter Settings			
N/F	50	20	10	5
σ_v	1.0	0.6	0.2	0.1
MC Mean of $\widehat{b}_0$	0.684 (0.018)	0.689 (0.024)	0.691 (0.043)	0.692 (0.043)
MC Mean of $\sigma(\widehat{b}_0)$	0.020 (0.002)	0.027 (0.002)	0.040 (0.004)	0.033 (0.004)
Max of MC Tail Prob. Distr. for H_0: $b_0 = 0$	0.001	0.000	0.001	0.001
% of Rejections for H_0: $b_0 = 0$	99%	100%	98%	99%
MC Mean of $\widehat{b}_1$	−0.679 (0.030)	−0.631 (0.043)	−0.535 (0.071)	−0.413 (0.077)
MC Mean of $\sigma(\widehat{b}_1)$	0.031 (0.003)	0.044 (0.004)	0.065 (0.006)	0.056 (0.007)
Max of MC Tail Prob.Distr. for H_0: $b_1 = 0$	0.000	0.001	0.002	0.001
% of Rejections for H_0: $b_1 = 0$	100%	99%	98%	99%
MC Mean of R^2	0.816 (0.038)	0.677 (0.045)	0.408 (0.064)	0.410 (0.062)

Second, we explore what happens to (within-simulation) average and standard deviation of GDP growth time-series[26] when both σ_v and firms' bargaining strength β are allowed to vary. Recall that, the higher β, the less firms take into account workers satisficing wages when they decide their contractual wage. Figs. 9 and 10 show Monte Carlo means of average and standard deviation of GDP growth rates. We find that the *higher* firms' bargaining strength, the *smaller* both average growth rates and their variability. Thus, allowing for some bargaining power on the workers' side implies better aggregate performance, but also more fluctuations. Furthermore, if firms are *less* responsive to market signals (e.g. they employ a path-dependent vacancy setting rule), the economy enjoys persistently higher average growth rates and persistently smaller fluctuations.

Finally, we assess the consequences of "fueling" the economy with higher technological opportunities (i.e. higher σ_Z) for different levels of β (and setting σ_v to an intermediate level). While a higher σ_Z implies higher average growth rates in all parameter settings (Fig. 11), a stronger bargaining power for workers still implies better aggregate performances. Together, more technological opportunities also entail a higher volatility in the growth process (see Fig. 12). Volatility can be weakened if one increases firm strength in wage bargaining.

---

[26] That is, we compute average and standard deviation of GDP growth rates within a simulation $\{h_t, t = 1, ..., T\}$, $h_t = \Delta \log Q_t$.

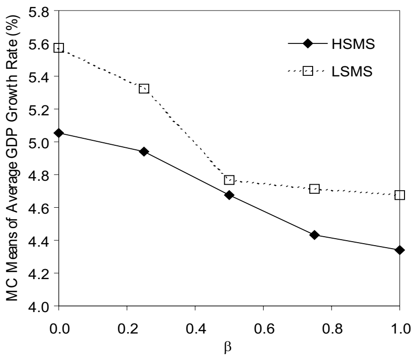


Fig. 9. Monte Carlo Means of (within-simulation) Average Real GDP Growth Rates as a function of firms strength in wage bargaining (β). LSMS vs. HSMS: Low ($\sigma_v = 0.1$) vs. High ($\sigma_v = 1.0$) sensitivity to market signals in vacancy setting. "Institutionally-Shaped" Environment. Parameters: $N/F = 5, \sigma_Z = 0.1$.

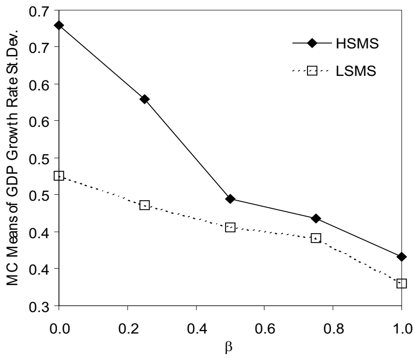


Fig. 10. Monte Carlo Means of (within-simulation) Standard Deviation of Real GDP Growth Rates as a function of firms strength in wage bargaining (β). LSMS vs. HSMS: Low ($\sigma_v = 0.1$) vs. High ($\sigma_v = 1.0$) sensitivity to market signals in vacancy setting. "Institutionally-Shaped" Environment. Parameters: $N/F = 5, \sigma_Z = 0.1$.

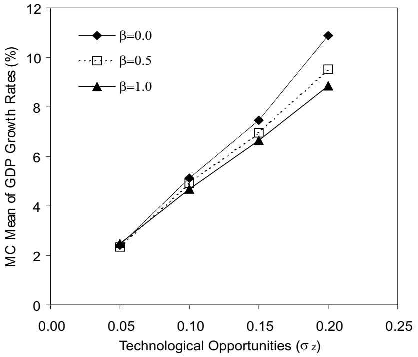


Fig. 11. Monte Carlo Means of (within-simulation) Average Real GDP Growth Rates as a function of technological opportunities (σ_Z) and firms strength in wage bargaining (β). "Institutionally-Shaped" Environment. Parameters: $N/F = 5, \sigma_v = 0.1$.

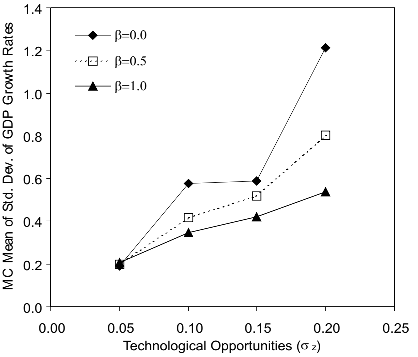


Fig. 12. Monte Carlo Means of (within-simulation) Standard Deviation of Real GDP Growth Rates as a function of technological opportunities (σ_Z) and firms strength in wage bargaining (β). "Institutionally-Shaped" Environment. Parameters: $N/F = 5, \sigma_v = 0.1$.

5 Conclusions

As far as the properties of labor market dynamics and the business cycle are concerned, three well-known aggregate regularities (i.e. Beveridge, Wage, and Okun curves) seem to provide a quite complete picture. Nevertheless, the existing theoretical literature still lacks micro-founded models which are able *jointly* to account for these three crucial stylized facts.

In this paper, we presented a preliminary agent-based, evolutionary, model trying to formalize from the bottom up individual behaviors and interactions in both product and labor markets.

In the model, vacancy and wage setting, as well as matching and bargaining, demand, and price formation, are all endogenous processes. Firms enjoy labor productivity improvements thanks to technological progress and undergo a selection pressure acting on their revealed competitiveness (which is also affected by their hiring and wage-setting behaviors).

Simulations show that the model is able robustly to reproduce Beveridge, Wage and Okun curves under quite broad behavioral and institutional settings. Moreover, the system generates endogenously an Okun coefficient greater than one (i.e. aggregate dynamic increasing returns) even if individual firms employ production functions exhibiting constant returns to labor.

Monte Carlo simulations also indicate that statistically detectable shifts in Okun and Beveridge curves emerge as the result of changes in institutional, behavioral, and technological parameters. For example, a higher concentration of market activity (i.e. a higher number of workers per firm) and a higher sensitivity to market signals in firms' vacancy setting rules imply Beveridge curves which lie closer to the axes. Finally, the model generates quite sharp predictions about how the average aggregate performance (and volatility) of the system change in alternative behavioral, institutional, and technological setups.

Many issues remain to be explored. First, additional Monte Carlo simulation exercises could be performed to more finely map (e.g. within a given "system setup") parameters and aggregate behaviors.

Second, the issue as to whether (and how) heterogeneity is able to affect the emergence of aggregate regularities might be addressed. For instance, one could explore the effects to endow workers (resp. firms) with increasingly heterogeneous distributions of reservation wages (resp. desired ratios of filled to open vacancies). Third, one might investigate the consequences of assuming alternative matching and bargaining processes to allow for a richer institutional setting. Finally, the structure of the model might be complicated in order to investigate economies where jobs last more than one period and firms are able to transfer profits across time.

References

Aghion P, Howitt P (1994) Growth and Unemployment. Review of Economic Studies, 61: 477–494

Aoki M (2003) A New Model of Output Fluctuation: Ultrametrics, Beveridge Curve and Okun's Law. Economics online papers, no. 234, UCLA

Aoki M, Yoshikawa H (2003) A Simple Quantity Adjustment Model of Economic Fluctuations and Growth. In: Cowan R, Jonard N (eds.) Heterogeneous Agents, Interaction and Economic Performance, Berlin, Springer

Ashenfelter O, Card D (eds.) (1999) Handbook of Labor Economics. Volume 3. Elsevier Science, Amsterdam

Ashenfelter O, Layard R (eds.) (1986) Handbook of Labor Economics. Volume 2. Elsevier Science, Amsterdam

Attfield CLF, Silverstone B (1997) Okun´s coefficient: A comment. The Review of Economics and Statistics, 79: 326–329

Belot M, Van Ours JC (2000) Does the Recent Success of Some OECD Countries in Lowering their Unemployment Rates Lie in the Clever Design of their Labour Market Reforms? Discussion paper no. 147, IZA, Bonn, Germany

Blanchard O, Katz LF (1997) What We Know and Do Not Know about the Natural Rate of Unemployment. Journal of Economic Perspectives, 11: 57–72

Blanchard O, Kiyotaki N (1987) Monopolistic Competition and the Effects of Aggregate Demand. American Economic Review, 77: 647–666

Blanchard O, Quah D (1989) The Dynamic Effects of Aggregate Demand and Supply Disturbances. American Economic Review, 79: 655–673

Blanchard OJ, Diamond P (1989) The Beveridge curve. Brooking Papers on Economic Activities, 1: 1–76

Blanchard OJ, Wolfers J (2000) The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence. Economic Journal, 110: C1–C33

Blanchflower DG, Oswald AJ (1994) The Wage Curve. Cambridge, Massachussetts, The MIT Press

Bleakley H, Fuhrer JC (1997) Shifts in the Beveridge Curve, Job Matching, and Labor Market Dynamics. New England Economic Review, 3–19

Börsch-Supan AH (1991) Panel Data Analysis of the Beveridge Curve: Is There a Macroeconomic Relation between the Rate of Unemployment and Vacancy Rate? Economica, 58: 279–297

Burdett K, Shi S, Wright R (2001) Pricing and Matching with Frictions. Journal of Political Economy, 109: 1060–1085

Cao M, Shi S (2000) Coordination, matching and wages. Canadian Journal of Economics, 33: 1009–1033

Card D (1995) The Wage Curve: A Review. Journal of Economic Literature, 33: 785–799

Card D, Hyslop D (1996) Does Inflation Grease the Wheels of the Labor Market. Working paper 5538, NBER

Dosi G, Marengo L, Fagiolo G (2004) Learning in Evolutionary Environment. In: Dopfer K (ed.) Evolutionary Principles of Economics, Cambridge, Cambridge University Press, Forthcoming

Dosi G, Nelson RR (1994) An introduction to evolutionary theories in economics. Journal of Evolutionary Economics, 4: 153–172

Dosi G, Winter SG (2002) Interpreting Economic Change: Evolution, Structures and Games. In: Augier M, March JG (eds.) The Economics of Choice, Change, and Organizations, Cheltenham, Edward Elgar Publishers

Epstein JM, Axtell R (1996) Growing Artificial Societies: Social Science from the Bottom-Up. Washington D.C., MIT Press

Evans GW (1989) Output and Unemployment Dynamics in the U.S. 1950-1985. Journal of Applied Econometrics, 4: 213–237

Fitoussi JP, Jestaz D, Phelps ES, Zoega G (2000) Roots of the Recent Recoveries: Labor Reforms or Provate Secto Forces? Brooking Papers on Economic Activities, 1: 237–291

Flaschel P, Kauermann G, Semmler W (2003) Testing Wage and Price Phillips Curves for the United States. Unpublished manuscript, Bielefeld University, Faculty of Economics

Forni M, Lippi M (1997) Aggregation and the Microfoundations of Dynamic Macroeconomics. Oxford, Clarendon Press

Hahn F, Solow RM (1997) A Critical Essay on Modern Macroeconomic Theory. Cambridge, MA, The MIT Press

Julien B, Kennes J, King I (2000) Bidding for Labor. Review of Economic Dynamics, 3: 619–649

Kirman AP (1992) Whom or what does the representative individual represent? Journal of Economic Perspectives, 6: 117–136

Laffond G, Lesourne J (2000) The genesis of expectations and of sunspot equilibria. Journal of Evolutionary Economics, 2: 211–231

Lagos R (2000) An Alternative Approach to Search Frictions. Journal of Political Economy, 108: 851–872

Layard R, Jackman R, Nickell S (1991) Unemployment. Oxford University Press, Oxford

Lesourne J (1992) The Economics of Order and Disorder. Oxford, Clarendon Press

Moosa IA (1997) A Cross-Country Comparison of Okun´s Coefficient. Journal of Comparative Economics, 24: 335–356

Mortensen DT (1986) Job Search and Labor Market Analysis. In: Ashenfelter O, Layard R (eds.) Handbook of Labor Economics, Volume 2, Amsterdam, Elsevier Science

Mortensen DT, Pissarides CA (1994) Job creation and job destruction in the theory of unemployment. Review of Economic Studies, 61: 397–415

Nickell S, Nunziata L, Ochell W, Quintini G (2001) The Beveridge curve, Unemployment and wages in the OECD from 1960s to 1990s. Working paper no. 502, CEPR, London, U.K.

Okun AM (1962) Potential GDP: Its Measurement and Significance. Proceedings of the Busisness and Economics Statistics, 98–103

Okun AM (1970) The political economy of prosperity. The Brookings Institution, Washington D.C.

Peters M (1991) Ex ante price offers in matching games non-steady states. Econometrica, 59: 1425–1454

Petrongolo B, Pissarides CA (2001) Looking into the Black Box: A Survey on the Matching Function. Journal of Economic Literature, 39: 390–431

Phelps ES (1972) Inflation Policy and Unemployment Theory. MacMillan, London

Phelps ES (1994) Structural Slumps. Cambridge, Harvard University Press

Phelps ES, Winter SG (1970) Optimal Price Policy Under Atomistic Competition. In: Phelps ES, Archibald GC, Alchian AA (eds.) Microeconomic Foundations of Employment and Inflation Theory, New York, Norton

Pissarides CA (2000) Equilibrium Unemployment Theory. Oxford, Blackwell

Prachowny MFJ (1993) Okun´s Law: Theoretical Foundations and Revised Estimates. The Review of Economics and Statistics, 75: 331–336

Schaik AV, Groot HLFD (1998) Unemployment and Endogenous Growth. Labour, 12: 189–219

Shapiro C, Stiglitz JE (1984) Equilibrium Unemployment as a Worker Discipline. American Economic Review, 74: 1257–78

Smith TE, Zenou Y (2003) A Discrete-Time Stochastic Model of Job Matching. Economic Dynamics, 6: 54–79

Sögner L, Stiassny A (2000) A Cross-Country Study on Okun´s Law. Working paper no. 13, Vienna University, Faculty of Economics and Business Administration

Solow RM (1998) What is Labour-Market Flexibility? What is it Good for? Proceedings of the British Academy, 97: 189–211

Tesfatsion L (1997) How Economists Can Get A Life. In: Arthur WB, Durlauf SN, Lane D (eds.) The Economy as an Evolving Complex System II, Santa Fe Institute, Santa Fe and Reading, MA, Addison-Wesley

Tesfatsion L (2001) Structure, Behavior, and Market Power in an Evolutionary Labor Market with Adaptive Search. Journal of Economic Dynamics and Control, 25: 419–457

Witt U (1985) Coordination of Individual Economic Activities as an Evolving Process of Self-Organization. Economie appliquée, 37: 569–595

An evolutionary model of international competition and growth

Carolina Castaldi[⋆]

ECIS, Eindhoven Technology University, The Netherlands and LEM, Sant'Anna School of Advanced Studies, Pisa, Italy (e-mail: castaldi@sssup.it)

Abstract. The aim of this paper is to investigate cross-country patterns of economic divergence in an evolutionary perspective. We propose a simple open economy evolutionary model of growth where the growth variables of each country are micro-founded on the dynamics of national firms. The model finds its antecedents in some of the evolutionary models of economic growth developed over the past years. We claim that evolutionary models are able to account for persistent differentiation in the growth performances of countries as a generic property. In fact, the model proposed here does so despite its quite simplified structure.

Key words: Growth – Technological Change – International Technological Diffusion – Evolutionary Models

JEL Classification: O30, O41

1 Introduction

Economists have devoted substantial theoretical and statistical efforts to understanding and modeling the process of international growth. Questions as to why

---

[⋆] I gratefully acknowledge financial support from the Robert Solow Post-Doctoral Fellowship of Saint-Gobain Center for Economic Studies, Paris, France. This work is a development upon my PhD dissertation at Sant'Anna School of Advanced Studies. Giovanni Dosi has provided the main inspiration for this paper. I have greatly benefited from comments and suggestions received at different stages from Giulio Bottazzi, William Brock, Giorgio Fagiolo, Roberto Gabriele, Marco Giarratana, André Lorentz, Stanley Metcalfe, Alessandro Nuvolari, Angelo Secchi, Mauro Sylos Labini and Bart Verspagen. I also wish to thank seminar participants at the ETE workshop 2002 in Jena, Germany and at the International Schumpeter Society Conference 2002 in Gainesville, FL, USA. The current version was improved thanks to useful comments by two anonymous referees. The usual disclaimers apply.

certain countries have achieved a remarkably higher level of economic development than others have made scholars of economic growth struggle for many years. The historical evidence shows persistent variability in the levels of per capita incomes and in the growth rates of countries worldwide. Even in the Post World War II period, commonly regarded as an era of growing uniformity, the hypothesis of global convergence, that is, convergence of the *whole* population of countries toward increasingly similar income levels, does not find support from the evidence (De Long, 1988; Easterly et al., 1992; Temple, 1999; Soete and Verspagen, 1993; Quah, 1996).

Indeed, a diverse body of 'endogenous growth' and 'new growth' models and empirics emerged in the 1980s and developed in the 1990s as an alternative to neoclassical growth theories with the aim of abandoning the assumption of exogeneity of technical progress in favor of including some mechanisms by which technological change is internally generated by the economic system. As presented in Romer (1994), the models developed in these years have been able to account for a set of stylized facts about the properties of technological change. One of the crucial achievements has been the acknowledgment of positive feedback mechanisms in production. More recently, 'new trade' models, such as the ones in Grossman and Helpman (1991), have enlarged the perspective onto open economy settings and studied how increasing returns and imperfect competition may endanger the gains from free trade.

Still, most new growth models have maintained the assumption of perfectly rational and perfectly informed agents, which justifies their concern with equilibrium, 'steady state' outcomes only. The resulting representation of technological change and of its relation with economic growth hides some of the 'dynamic' properties of a process which is in continuous evolution and far from smooth and instantenous in its diffusion. As a corollary, these models are well equipped to account for convergence to different types of steady states, whether single or multiple equilibria, but show relatively more difficulties in accounting for historical divergence.

A growing stream of evolutionary literature, finding its contemporary roots in Nelson and Winter (1982), has provided an alternative approach to model economic growth. The building blocks of evolutionary theories of economic change (as outlined recently in Coriat and Dosi, 1998 and Dosi and Winter, 2002) can be traced to a series of evolutionary models of international growth developed mostly in the 1990s. The most relevant ones are reviewed in Silverberg and Verspagen (1995) and Kwasnicki (2001). Broadly speaking, distinguishing features of these models as compared to new growth models may be found in terms of their ability to capture, *first*, the disequilibrium and path-dependent nature of the process of economic growth, and, *second*, empirically sounder representations of the micro-economic activities of technological change.

The model presented here finds inspiration in one of the few *open economy* evolutionary models, the one in Dosi et al. (1994), which is in turn an extension of Chiaromonte and Dosi (1993). The main driving forces remain learning and market selection. But the process of technological diffusion via imitation acquires properties specific to its international dimension and the 'distance' between technologies used by firms in different countries is higher than the distance between technologies used in the same country. In the international setting, the selection mechanism works in a world market and the definition of an 'international competitiveness' is the key to operationalize the link between technology and growth. A selection based on competitiveness is used in Verspagen (1993), where the micro-units are the sectors of production in the different countries.

The aim of this paper is to show how a rather simple open economy evolutionary model with homogenous settings is able to generate patterns of persistent cross-country divergence in income and growth despite the inclusion of a process of international technological diffusion which works towards promoting convergence. We do not aim here at performing any kind of calibration exercise or at reproducing specific historical time series. Our interest lies in accounting for some 'generic' statistical properties of the observed cross-country patterns of growth and their evolution in time. Such a goal, we believe, justifies a few simplifying assumptions and the choice of homogenous parameter settings for the model presented here. Given the simple structure of the model, we claim that the property of accounting for persistent divergence may be considered as a generic property of evolutionary models of growth which may thus challenge the awkward inability of most mainstream models to deal with almost any form of historically meaningful economic divergence.

2 A simple open economy evolutionary model of endogenous growth

2.1 The structure

Let us first briefly outline how the model is structured. We consider a system of open economies. A finite set of N firms is represented that produce a homogenous output. Firms are localized homogeneously in C countries and each firm $i = 1, ..., N$ is characterized by the country $c(i) = 1, ..., C$ in which its production is located. Labor is the only input in production, and so labor productivity indicates the efficiency of the technology used by each firm. Technological innovation is allowed in the form of capital-disembodied process innovation.

Firms systematically engage in R&D activities, to which they devote a fixed percentage of income. The search activities may result in innovation with the development of a new technique associated with a higher productivity level, or firms may end up imitating a better technique used by other firms worldwide. At the beginning of each period, firms draw the outcome of their search activities with a probability depending on the level of current and lagged investment. The imitation process depends upon a technological distance, adjusted by a geographical parameter.

Once all firms have updated their productivities, they set their prices. The effective price in a specific country of the good produced by a certain firm depends on costs related to trade. Price and trade costs together determine the 'competitiveness' of each firm. An evolutionary market selection mechanism establishes the market shares of all firms.

Demand comes from the wages earned in each country and is allocated to firms in relation to their market shares. Firms then set their production levels. At the end of each period, wages are updated in the labor market and aggregate variables can be calculated for each country.

From a methodological point of view, we develop an agent based model and we use Monte-Carlo techniques to study its predictions.

2.2 The activities of firms and the micro-economics of technological change

The set of micro stylized facts that we choose to build upon stems from the economics of innovation and the processes of technological change at the firm level (see Dosi, 1988).

The evidence shows that technological improvements are eventually achieved only after costly and risky R&D activities. In our model, firms produce a homogenous output using labor as the only input. The technology used is characterized by the labor productivity measure. Technological change is endogenously determined and firms engage in R&D in order to innovate or imitate existing techniques, the ultimate goal being an improvement in their productivity. We assume that firms invest a fixed and constant percentage of income in search activities; there is empirical evidence that firms follow this type of strategy and we can justify it with the idea that firms are required to keep investing in search because of the threat of advances by rival firms. Both innovation and imitation need investment for each period.

Search is costly and requires R&D investment. Search is also uncertain and thus modeled as a stochastic process with a certain probability of innovating or imitating. In fact, the probability of success crucially depends on the level of R&D investment[1] undertaken by the individual firm. The more invested in search, the higher the probability of the search being successful.

For each firm $i = 1, ..., N$ the amount of labor L_i to be invested in the two complementary search activities of innovation and imitation is decided at the beginning of each period $t = 1, ..., T$, based on the following rules:

$$L_{i,t} = \frac{RD_{i,t}}{w_{c(i),t-1}} \tag{1}$$

where RD_i is the total investment in search activities given by:

$$RD_{i,t} = r_1 * Y_{i,t-1} \tag{2}$$

where $Y_{i,t-1}$ is the income obtained in the previous period and $w_{c(i),t}$ is the wage rate valid in the country $c(i)$ in which firm i operates. The division of resources between innovation and imitation is set by the parameter r_2 for all firms $i = 1, ..., N$:

$$LINN_{i,t} = (1 - r_2) * L_i \tag{3}$$

$$LIM_{i,t} = r_2 * L_i \tag{4}$$

---

[1] In the case of developing countries, we take R&D expenditure to be investment in training and equipment.

where $LINN_{i,t}$ and $LIM_{i,t}$ are the numbers of workers employed in innovation and imitation activities respectively.

Innovation and imitation search are modeled as two stage stochastic processes, where the first stage defines the probability of innovation (imitation) and the second stage sets the actual random realization of innovation (imitation) itself.

As for innovation, each firm draws its own success with a probability that depends both on the current and past labor investment in innovation search:

$$\text{P(success in innovation)} = 1 - e^{-\alpha(LINN_{it}+LINN_{i,t-1})} \tag{5}$$

where the parameter α governs in probability the 'innovative productivity' of R&D investment.[2] Then the actual size of innovation s_i is drawn from a $Poisson\,(\lambda)$, and the resulting new productivity $\pi_{INN,i,t}$ for firm i is determined by:

$$\pi_{INN,i,t} = \pi_{i,t-1} * \left(1 + \frac{s_i}{100}\right) \tag{6}$$

It should also be noted that, in this model, innovation is in fact mainly incremental, but we allow for particularly 'big' jumps when letting the size of innovation be randomly chosen from a Poisson distribution. Finally, note that even if the opportunity in terms of percentage *change* in productivity were the same for every firm, firms would still generally be heterogenous with respect to their *level* of opportunity. Also, whenever firms successfully draw an increase in productivity, their level of investment in R&D in the next period is generally higher than that of firms that do not innovate. Therefore, in the next period, innovating firms enjoy a higher probability of gaining a percentage increase in their productivity. An analogous positive feedback mechanism applies for the imitation process.

For imitation each firm draws a binary success variable with probability given by:

$$\text{P(success in imitation)} = 1 - e^{-\beta(LIMM_{i,t}+LIMM_{i,t-1})} \tag{7}$$

If success is drawn, the firm will have the chance to imitate one of the existing techniques. We define a distance between technologies π_{i_1} and π_{i_2} located, respectively, in countries $j_1 = c(i_1)$ and $j_2 = c(i_2)$ as the following:

$$d\left(\pi_{i_1}, \pi_{i_2}\right) = \begin{cases} \max\{0, \pi_{i_2} - \pi_{i_1}\} & \text{if } j_1 = j_2 \\ \xi \max\{0, \pi_{i_2} - \pi_{i_1}\} & \text{if } j_1 \neq j_2 \end{cases} \tag{8}$$

where $\xi \geq 1$.[3] Note that we take the ξ parameter to be homogenous across countries, so that the only differentiation is between home and rest of the world.

Each firm will imitate the closest firm for which the distance is strictly positive. Call $\pi_{IMI,i,t}$ the imitated technique. The new production technique $\pi_{i,t}$ for firm

---

[2] Letting α_j depend on the country can be a way to account for the efficiency of different national innovation systems.

[3] One could envisage the parameter depending on the specific pair of countries compared, i.e. assuming ξ_{j_1,j_2} for each pair j_1, j_2. Here we take the simplifying assumption of a homogenous parameter.

i at time t is then chosen as the best alternative between the old one and the ones found through search.

$$\pi_{i,t} = \max\left\{\pi_{i,t-1}, \pi_{INN,i,t}, \pi_{IMI,i,t}\right\} \tag{9}$$

The way we model the search activities is meant to take into account the idea that the whole R&D search crucially draws on accumulated knowledge, so that dynamic increasing returns in production generally arise. Moreover, processes of learning play a key role for knowledge-driven progress.

We follow Nelson and Winter (1982) and more directly Chiaromonte and Dosi (1993) in distinguishing two types of learning.

First, a 'private' learning arises, which is mainly a learning-by-doing and pertains to the single firm. Experience cumulates with time as well with the volume of investment, so that the probability of both imitating and innovating depends on the current and lagged efforts (as in Eq.(5) and (7)). Engaging in R&D activities in every period allows firms to acquire competences and skills. This positive feedback explains why a firm that has consistently been successful in search activities in the previous periods enjoys a relatively higher probability of success in both innovation and imitation.

Second, there is 'collective' learning due to the technological spillovers that are granted to the other firms because of information diffusion, personnel mobility, etc. The latter effect ends up producing an enlargement of the public knowledge base. In principle, the extent of this collective learning may vary and the effect may work at a sectoral level and/or at a local (in the geographical sense) level. In this model, collective learning is represented by the knowledge spillovers that form the basis for technological diffusion and crucially enable imitation. In fact, imitation is feasible when technological knowledge is only partially appropriable by the firm that first developed it.

The size of spillovers depends upon the technological efficiency level achieved by firms. We model the imitation process so that the probability of imitating a technique decreases with the technological distance. Here the idea is that there are stronger information flows between firms with similar level of efficiency in production. This assumption is reasonably supported in the real world if we think that similar firms may indeed access the same type of labor force, undertake technological alliances, and so on.

The defined distance also includes the parameters ξ, which adjust for the distance of countries and are meant to take into account the geographical location of firms– possibly influenced also by the presence or absence of similar embedding institutions associated with a common country. A firm has a higher probability of imitating a better technique used in the same country, because the distance measure implied by the parameter ξ will make it harder to imitate techniques employed abroad. This formulation acknowledges the positive effect of spatial proximity in knowledge diffusion and the local nature of spillovers. Indeed, technological diffusion clearly depends on the properties of those micro-interactions likely to yield knowledge spillovers. There are costs, difficulties and delays associated with the transfer of information, and here the tacit component of technological knowledge is the one that limits a frictionless spread of technological knowledge. Increasing returns in production have been claimed to stem from spatial proximity in many models (see for example Arthur, 1994 and David, 1975) that deal with agglomeration economies and stress the localized nature of knowledge spillovers. It should also be noted that economies of agglomeration are very much related to innovation being cumulative. Innovative activity builds upon knowledge generated by previous research, so that the accumulation of innovative activity in a certain country/region facilitates the creation of new technological knowledge.

2.2.1 Market dynamics. Firms set the price for their goods with the simple rule:

$$price_{i,t} = \frac{w_{c(i),t-1}}{\pi_t}(1+m) \tag{10}$$

where m is a constant mark up.[4] Note also that the productivity is the (possibly) new productivity, while the wage is the wage given at the time that R&D investment decisions were taken. So, a one lag adjustment in the wage is assumed, which can be related to a weak idea of stickiness of prices.

Let us consider the evolutionary mechanism governing selection. The selection environment is the worldwide market and the ‘fitness’ of each firm is captured by its ‘competitiveness’. In turn, competitiveness of firm $i = 1, ..., N$ in country $j = 1, ..., C$ is defined as the inverse of the set price. The lower the price, the more competitive the firm. But the effective price $p_{ij,t}$ valid for goods produced by i and sold in country j at time t, is given by:

$$p_{ij,t} = \begin{cases} price_{i,t} & \text{if } c(i) = j \\ v * price_{i,t} & \text{if } c(i) \neq j \end{cases} \tag{11}$$

where $v > 1$. The v parameter is a weight that accounts for barriers in trade (including fees and/or transportation costs and of course tariffs, etc.) between countries.[5] In this case, the price of the good produced by firm i and sold in country j depends on the generally set price, $price_{i,t}$, and on the distance between the country $c(i)$ in which firm i is located and the country j in which the good is being sold. Note that the parameter v is homogenous across pairs of countries, so we assume homogenous and symmetric trade conditions.

Thus competitiveness is given by:

$$E_{ij,t} = \frac{1}{p_{ij,t}} \tag{12}$$

Therefore, thanks to the v parameters, the competitiveness is determined by how open the market in j is to the products coming from country i.

---

[4] In order to respect aggregate accounting equalities, for simplicity we set the profits generated by the mark up margins to be equal to the investment in R&D actually undertaken by all firms, i.e. we set the corresponding parameters r_1 and m equal to each other.

[5] We neglect exchange rates adjustments and we implicitly assume an ‘absorption approach’ to the balance of trade (Alexander, 1952).

The evolution of market shares of firms is governed by a replicator dynamics equation which has the following form:

$$ms_{ij,t} = ms_{ij,t-1} * \left(1 + c * \left(\frac{E_{ij,t} - EE_{j,t}}{EE_{j,t}}\right)\right) \tag{13}$$

where $EE_{j,t}$ is the average competitiveness of all firms operating in country j, calculated as a weighted mean, with weights provided by current market shares. We set $c = 1$ as the baseline value. The basic intuition behind Eq.(13) is that firms with above average competitiveness will expand in their relative importance and those below will shrink. This simplifying but heroic assumption captures in a highly stylized form the structure of strongly information imperfect markets with some degree of demand stickiness. In perfectly competitive markets, the firm setting the lowest price would have a market share of one and all the demand would eventually be satisfied by that company. But if information costs are accounted for, then it is reasonable to assume that a range of firms satisfies the market demand according to their relative competitiveness, so that a replicator dynamics mechanism can model the dynamics of shares in the product markets (more in Metcalfe, 1988 and Silverberg et al., 1988).

Note that two simplifying assumptions are made. First, product is homogenous. Second, the country of origin of the products does not affect their competitiveness, which is in turn the only relevant variable.

Market dynamics crucially interact with the processes of technological change. In fact, the ultimate outcome of successful R&D activities concerns the ability of the firm to gain or maintain its advantage in market competition. This is determined through the market shares dynamics governed by the replicator equation (Eq.13). This mechanism rewards those firms able to improve relatively to the average performance of all the other actors present in the world market.

Market shares determine the percentage of total demand that each firm satisfies. Domestic demand for goods in each country $j = 1, ..., C$ is equal to the total income of workers employed in domestic firms:

$$D_{j,t} = \sum_{\text{i st c(i)=j}} w_{c(i),t-1} * L_{i,t} \tag{14}$$

The production level of firm i at t, $Y_{i,t}$, is determined by the share of demand that the firm satisfies both in the domestic and in the foreign markets:

$$Y_{i,t} = \sum_{j=1}^{C} ms_{ij,t} * D_{j,t} \tag{15}$$

After determining their production levels firms set their demand for labor $L_{i,t}$:

$$L_{i,t} = \frac{1}{\pi_{i,t}} \sum_{j=1}^{C} \frac{ms_{ij,t} * D_{j,t}}{p_{ij,t}} \tag{16}$$

We assume that labor can move freely across firms within a country, but does not move across countries. For simplicity, we assume that labor employed in production and in R&D activities is the same. Therefore, we can consider a single labor market in which a common wage is determined. The wage rate changes according to the dynamics of domestic average productivity, the price index and the demand for labor:

$$w_{j,t} = w_{j,t-1}(1 + \Delta w_j) \tag{17}$$

where $\Delta w_j = e\Delta\pi_j + f\Delta p_j + g\Delta L_j$. $\Delta\pi_j$ is the percentage change in the average productivity of firms producing in country j, Δp_j represents the percentage change of a weighted index of prices available in country j, and ΔL_j is the percentage change in labor demand. By tuning the parameters e, f and g, we are able to specify a wide range of different labor market regimes. When $e = 1$, wage is indexed directly on labor productivity. At another extreme ($g = 1$), the market is competitive and the wage rate responds to changes in labor demand.

All aggregate country variables are obtained by summing up values at the firm level as in national accounts statistics. Let the national real income of country j at time t, $YY_{j,t}$, be:

$$YY_{j,t} = \sum_{\text{i st c(i)=j}} \frac{Y_{i,t}}{p_{i,j,t}} \tag{18}$$

while the level of per capita income is given by:

$$Z_{j,t} = \frac{YY_{j,t}}{L_{j,t}} \tag{19}$$

where $L_{j,t}$ is the total amount of labor force employed in firms of country j.

Exports $X_{j,t}$ and imports $M_{j,t}$ are given by:

$$X_{j,t} = \sum_{\text{i st c(i)=j}} \sum_{\text{jj}\neq\text{j}} ms_{i,jj,t} * D_{jj,t} \tag{20}$$

$$M_{j,t} = D_{j,t} - X_{j,t} \tag{21}$$

Finally, the growth rate of GDP is calculated as the percentage change in national income:

$$g_{j,t} = \frac{YY_{j,t}}{YY_{j,t-1}} - 1 \tag{22}$$

2.2.2 Interaction patterns. Different levels of interaction are specified in the design of the proposed model.

1. Imitation is mediated by the technological distance and controlled by the parameters β and ξ. It should be noted that this process does not come from a maximization rule, as firms simply select the closest productivity. When able to imitate, firms are not generally able to select the highest productivity available in the world, so that there is imperfection in the diffusion process. The parameter β adjusts the probability of drawing a success in the imitation search and can be used to control for the effectiveness of the search activity. ξ is the parameter used to discriminate domestic vs foreign firms in terms of imitability.

2. Competition in the commodities market works on the competitiveness variable and is tuned by the parameter c, capturing micro demand elasticity.
3. Trade relationships are modulated by the 'distance' parameter v, which in turn affects 'competitiveness' in foreign markets.

2.3 A first benchmark setting

2.3.1 Preliminary analysis. We start by analyzing a simple case. Let us begin from a perfect initial symmetry of all countries and firms with all parameters v and ξ set equal to 1 and no differences in initial endowments. Suppose the market competition parameter c also equals 1. As for innovation and imitation, they work as described in the previous section, but we first consider a fixed and constant innovation size: $s_i = s$ for $i = 1, ..., N$.

All firms are endowed with the same opportunities for technological change. Some of them will be able to have their initial productivity π_{INI} jump to $\pi_{INN,i,t} = \pi_{INI} * (1 + \frac{s}{100})$. The other firms will keep their initial productivity.

In the following period, those successful firms will maintain their technological superiority regardless of what happens to the remaining firms. We can see this as an inertia effect due to the assumption of fixed and constant innovation size. In the extreme case, firms that have innovated at time 1 do not increment again their productivities, while all other firms do. This complete "catching up" will cause market shares to stay the same and will prompt all firms to start again from a situation in which they all have the same labor productivity. The only changed element is that now the probabilities of innovation differ across firms. As noted above, firms that are able to increase their productivity at time t will manage to invest relatively more in R&D at time $t + 1$ and thus enjoy a higher probability of another increase in productivity. The path dependence effect will in general work against future complete catching up situations, given that the parameter α tuning the effect is high enough.

Thus, even in this symmetric, highly artificial case with fixed expected value of percentage productivity increments, initial events shape the future distribution of micro technological opportunities across firms. Now, if we assume that wages are indexed perfectly on productivity changes ($e = 1$), then the distribution of technological capabilities will be reflected directly in the cross-country distribution of growth performance. A delicate issue is the aggregation of firm production levels to calculate the country's growth rate, i.e. the modes in which ones goes from 'micro' to 'macro'. In this respect, one needs to to investigate whether the catching up in technological capabilities corresponds to a catching up in the growth performance.

2.3.2 The effect of distance. The formulation of the model entails two ways in which firms can be distant. First, firms have a distance in the technological space. Second, countries can be more or less close when trading with each other in the international market. We start by investigating the role of national barriers as related to the first notion of distance.

A 'technological distance' appears when probabilities of imitation depend on the differential in efficiency between firms. The parameter ξ accounts for positive

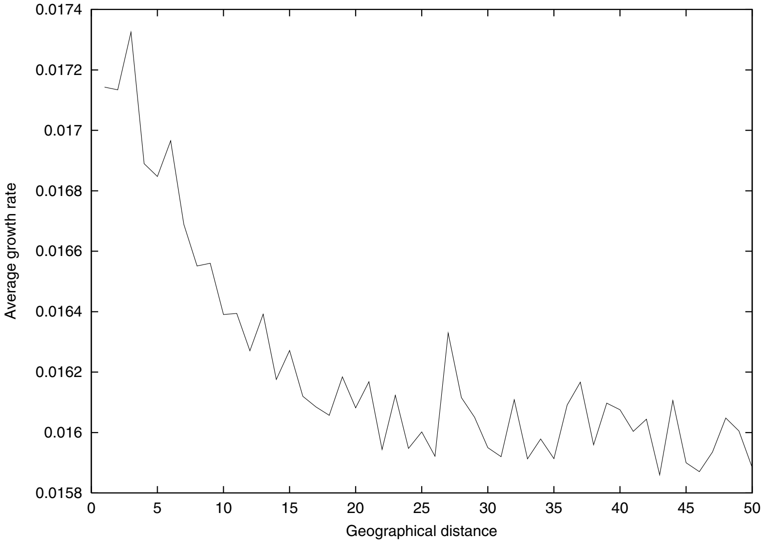


Fig. 1. Effect of the ξ parameter on the average growth performance

feedbacks in the process of innovation diffusion for firms located in the same country. After an innovation has been developed by a firm, it is more likely that it diffuses in the same geographical area. Different sorts of barriers to the flow of knowledge act as impediments against a perfect and smooth diffusion of new technologies to all countries. Information spillovers have a local nature (or national specificity) and are stronger when there is geographical proximity. Recent work (see Eaton and Kortum, 2001, Keller, 2000 and Caniëls and Verspagen, 2001 among others) has in fact analyzed the relevance of geographical barriers in the process of technological diffusion.

In a first run of simulations, we take the ξ parameters to be homogenous across countries, so that the only differentiation is between home and rest of the world. We want to use the model to study the effect of the ξ parameter on the growth performance of the system. Thus, let the average growth performance of country j in the T time steps, AGR_j, be measured by:

$$AGR_j = \frac{\log YY_{j,T} - \log YY_{j,1}}{T-1} \tag{23}$$

where $YY_{j,T}$ is the level of real GDP for country j at some ‘final’ time T. The overall growth performance of the system is then estimated in terms of the mean of all AGR_js, averaged on $M = 500$ runs of Monte-Carlo simulations.

As it appears for the “baseline” parameterizations[6] (cf. Fig. 1), the average growth performance decreases with the value of the parameter, supporting the previous discussion on the negative effect of distances between countries. The case

---

[6] The specific parameterization used sets the values: $N = 100$, $C = 10$, $Y_{i,1} = 1000$, $\pi_{i,1} = 0.5$, $T = 100$, $r_1 = 0.05$, $r_2 = 0.5$, $\alpha = 0.05$, $\beta = 0.05$, $d = 10$, $c = 1$.

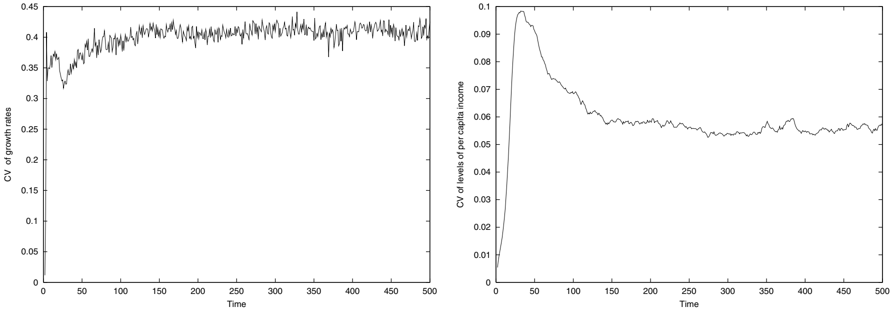


Fig. 2. Cross-country variability of growth rates (top panel) and of levels of per capita income (bottom panel).

$\xi = 1$ corresponds to the situation in which firms communicate with one another without bearing the costs associated with any sort of inter-country 'distance'. We can view this as a 'no gravity' model in which there are no national constraints to a smooth diffusion process regime. As soon as $\xi > 1$, country-specific externalities are inserted in the international system. When ξ is large, countries tend to be isolated and are very unlikely to imitate technologies developed by foreign firms. In this case, an innovation designed in a given country is hardly imitated outside the borders, so that the only source of productivity increments for firms are the self-produced innovation and the imitation of competing firms in the same country. This sharply reduces the chances of technological improvements for firms and, not surprisingly, negatively affects the overall performance of the system.

2.3.3 Patterns of divergence. In our model, technological change is the main force at work in determining national income. The diffusion process that we insert can be thought of as a mechanism promoting convergence between firms and thus between economies. In fact, imitation involves knowledge transfers in the form of spillovers from technologically superior firms to inferior ones. Clearly, firms that are positioned at the technological frontier are not profiting from spillovers because there are no better techniques for them to imitate. In this sense, imitation is a 'catch-up process' moving the system towards convergence.

Conversely, the innovation generated at the firm level works as a potential cause of heterogeneity and divergence in technological performance. Given the Schumpeterian insight that innovation creates technological gaps which then entail gaps in the economic performances of countries, one may see the innovation process as a force towards divergence.

We explore the evolution in time of a measure of cross-country variability of two economic variables of interest, namely (i) the levels of per capita income $Z_{j,t}$ and (ii) the growth rates of 'national' incomes $g_{j,t}$.

Figure 2 shows that the cross-country coefficient of variation of growth rates between countries is persistent and significant over time. A similar result is valid for the dynamics of the cross-country dispersion of per capita incomes (bottom panel of Fig. 2). After an initial phase of transition, the coefficient of variation settles to a steady value that indicates persistent differentiation of income level.

So, even when including a significant process of international technological diffusion and despite starting from a homogenous setting, the model generates steady differentiation in the distribution of growth rates and income across countries.

3 Some conclusions

We have proposed here an evolutionary model of endogenous growth for a system of open economies. Our model includes several features directly inspired by the models of growth in Nelson and Winter (1982), specifically in their focus on the micro activities of technological change performed at firm level, but extends the analysis to an open economy framework.

The results are insights coming from preliminary simulation runs of the model. More work remains to be done to establish carefully the robustness of the claimed results. Nevertheless, this first analysis offers quite interesting insights.

First, we investigate the properties of international technological diffusion and its bearing on the process of international growth. A technological distance between firms is defined and we find that the average growth performance of the system is negatively affected by increasing geographical barriers.

Second, despite the effectiveness of the process of technological diffusion, the model can generate persistent differentiation in the level of per capita income and in the growth rates of countries in the system. This property captures an important stylized fact in the empirical literature on growth. A crucial remark should be made here. We assumed the key parameters of our model of be homogenous across countries. If such a model is able to account for divergence patterns, the claim should *a fortiori* be true when one adds heterogenous parameters settings. Additionally, quite a few mechanisms in the model are exogenous.[7] Endogenizing them would reasonably add new sources of divergence in the system due to possible positive feedback mechanisms.

All this, we conjecture, sustains the claim that the ability of accounting for persistent international growth performance differentials is indeed a generic property of evolutionary models. This property stands out as a strong valued added of evolutionary models of growth as an alternative to mainstream endogenous growth models. Thus, further theoretical investigations in the evolutionary perspective may build on this property to produce novel insights over the properties of technological learning as an engine of international growth.

References

Alexander S (1952) The Effects of Devaluation on a Trade Balance. IMF Staff Paper

Arthur WB (1994) Increasing Returns and Path-Dependence in the Economy. University of Michigan Press, Ann Arbor

Caniëls MCJ, Verspagen B (2001) Barriers to knowledge spillovers and regional convergence in an evolutionary model. Journal of Evolutionary Economics, 11: 307–329

---

[7] For instance, the fraction of income invested in R&D is constant, when there is instead evidence that this investment decision may be better understood as an endogenous process.

Chiaromonte F, Dosi G (1993) Heterogeneity, competition and macroeconomic dynamics. Structural change and Economic Dynamics, 4: 39–63

Coriat B, Dosi G (1998) The Institutional Embeddedness of Economic Change. An Appraisal of the "Evolutionary" and the "Regulationist" Research Programs. In: Nielsen K, Johnson B (eds.) Institutions and Economic Change, Edward Elgar, Cheltenham, UK

David P (1975) Technical choice, innovation and economic growth. Cambridge University Press, Cambridge, MA

De Long BJ (1988) Productivity growth, convergence and welfare. American Economic Review, 78: 1138–54

Dosi G (1988) Sources, Procedures and Microeconomic Effects of Innovation. Journal of Economic Literature, 26: 1120–71

Dosi G, Fabiani S, Aversi R, Meacci M (1994) The dynamics of international differentiation: a multi-country evolutionary model. Industrial and Corporate Change, 3: 225–242

Dosi G, Winter SG (2002) Interpreting economic change: evolution, structures and games. In: Augier M, March J (eds.) The Economics of Choice, Change and Organizations: Essays in Memory of Richard M. Cyert, Edward Elgar, Cheltenham, UK

Easterly W, King R, Levine R, Rebelo S (1992) How do national policies affect long-run growth? A research agenda. World Bank, Discussion paper

Eaton J, Kortum SS (2001) Technology, trade and growth: a unified framework. European Economic Review, papers and proceedings, 45: 742–755

Grossman G, Helpman E (1991) Innovation and growth in the global economy. MIT Press, Cambridge, MA

Keller W (2000) Do Trade Patterns and Technology Flows Affect Productivity Growth? World Bank Economic Review, 14: 17–47

Kwasnicki W (2001) Comparative analysis of selected neo-schumpeterian models of industrial dynamics. presented at the 'Nelson and Winter conference', Aalborg, June 2001

Metcalfe JS (1988) The diffusion of innovation: an interpretative survey. In: Dosi G, Freeman C, Nelson R, Silverberg G, Soete L (eds.) Technical Change and Economic Theory, Pinter Publisher, London

Nelson RR, Winter SG (1982) An evolutionary theory of economic change. Harvard University Press, Cambridge, MA

Quah D (1996) Twin peaks: growth and convergence in models of distribution dynamics. Economic Journal, 106: 1045–1055

Romer PM (1994) The Origins of Endogenous Growth. Journal of Economic Perspectives, 8: 3–22

Silverberg G, Dosi G, Orsenigo L (1988) Innovation, diversity and diffusion: a self-organization model. The Economic Journal, 98: 1032–1054

Silverberg G, Verspagen B (1995) Evolutionary theorizing on economic growth. IIASA working paper 95-78, forthcoming in K. Dopfer (2004), The Evolutionary Principles of Economics, Cambridge University Press, Cambridge

Soete L, Verspagen B (1993) Technology and growth: the complex dynamics of catching up, falling behind and taking over. In: Szirmai A, van Ark B, Pilat D (eds.) Explaining Economic Growth, Elsevier Science Publishers, Amsterdam

Temple J (1999) The New Growth Evidence. Journal of Economic Literature, 37: 112–156

Verspagen B (1993) Uneven Growth between Interdependent Economies: an Evolutionary View on Technology Gaps, Trade, and Growth. Ashgate Pub., Avebury, Aldershot, UK

Innovation and growth in Germany over the past 150 years

Hariolf Grupp[1,2], Iciár Dominguez Lacasa[1], Monika Friedrich-Nishio[2], and Andre Jungmittag[3]

[1] Fraunhofer Institute for Systems and Innovation Research (ISI), Breslauer Str. 48, 76139 Karlsruhe, Germany
[2] Institute of Economic Policy Research (IWW) at the University of Karlsruhe, Post Box, 76128 Karlsruhe, Germany
[3] Chair for Macroeconomic Theory and Policy, University of Wuppertal, Gauss-Str. 20, 42119 Wuppertal, Germany

Abstract. This contribution starts from today's definitions of innovation indicators and traces their evolution back over the past 150 years. It is divided into a descriptive and an econometric part. The German innovation system has generally been very stable, even though it witnessed several political changes over the past century. This allows a comparison of the period 1850–1913 with 1951–1999. In the first period, the overall empirical results indicate a linear innovation relation between student numbers as well as public science expenditure, the number of patents granted, and economic demand. However, the second period suggests a different logic in the innovation process: demand drives total R&D expenditure, while patent output does not follow demand. The real domestic product does not seem to depend strongly on innovation activities.

JEL Classification: L16, L52, O11, O31, O47, N10

1 Introduction and structure of the contribution

If we want to understand our modern science and technology (S&T) system as an outcome of its evolution, we have to understand how it came about historically. This is an important task for historians of technology and economic historians, but not the prime concern of this paper. Rather, we want to understand what drives an innovation system over long periods, including external shocks which could be disruptive to the arrangement of driving factors. Can we provide evidence as to how such an innovation system reacted to earlier structural changes? Do we observe paradigm shifts or more persistent elements?

Germany is an ideal object of study as it has undergone many territorial and governance changes: it existed as a large array of individual states before 1871, was united in 1871 from Alsace-Lorraine (in the South-West) to East Prussia (in the North-East), reduced in 1918 after World War I and split into the GDR and West Germany from 1949 to 1990. You can hardly imagine more external shocks to an innovation system.

A suitable measurement concept is essential for such an endeavor. A practicable way to measure innovation and growth in the long term could be the elaboration of definitions and measurement methods historically, with the objective of recording the enormous changes characterizing innovation activities. However, our contribution takes the opposite point of departure: starting from today's definitions, an investigation of the comprehensive statistical material, including related indicators, is carried out, followed by an attempt to trace and complete these indicators back to before the foundation of the German Empire. This means that the presently achieved level of theory and methodology serves as a point of departure.

Consequently, this contribution tries to include a considerable number of quantitative variables in the form of *time series*. This analysis can be included in the field of *cliometrics*, the "new" kind of economic history, which is based on quantitative methods, including econometrics, aimed at reconstructing and interpreting the past (Bannock et al., 1998, p. 61). This method has been criticized by evolutionary economists (Freeman and Louçã, 2001, pp. 9; Walter, 1997, p. 75) since *indicators* cannot be *facts*; however, no fundamental difference exists between the description of facts and the interpretation, since every description already represents a certain interpretation which, moreover, depends on the definitions presently available (Lorenz, 1997, p. 32). This is aggravated by the fact that the theoretical constructs of innovation research are not clearly defined. Up to the present, rival and incompatible innovation theories still exist in several disciplines (Grupp, 1998). Linear models are widespread, presuming a sequential succession of innovation-oriented phases, the point of departure of which is an unpredictable serendipity in basic research or exogenous technical progress which falls like manna from heaven. It is evident that the definition of innovation, as it is used nowadays, cannot be considered as an anchor for the investigation period since 1850. Prior to the 1960s, innovation phenomena were described using other definitions: archives, libraries, and research institutions, as well as documents from management, personnel departments, or from production centers use terms that differ from those used according to present standards (like laboratory, try establishment, experimental factory).

According to today's view, the concept of a specific research process that leads to measurable innovation and that requires personnel and financial expenditure, is based on Bernal (1939), who distinguished the role of public research expenditure from that of civil research and – as things stood at the time – from that of the war industry. The first statistics on expenditure for "industrial research" by British companies are found in the annexes of his works. As reported by Freeman (1992, p. 3), the definitions used by Bernal during his lectures at the London School of Economics were brought to international committees (by Freeman himself, as well as by others). Here, in the 1960s, work was done on another standardization of definitions, which resulted in a first paper on the measurement of output of research and development (Freeman, 1969). Today, there is an established multidimensional array of innovation-related indicators (see, e.g. Grupp, 1998). Nevertheless, these still suffer from various respective shortcomings, so their combined use is recommended (loc. cit.).

Consequently, the empirical framework underlying this contribution will be determined using the current definitions and concepts. While these may have had other meanings in the past, this "*anachronism*" of long-lived indicators' definitions must be accepted (Lorenz, 1997, p. 364). The definitions found in the leading OECD manuals from the 1990s will be used. If a structural breakage is found in time series, this could point to sudden changes in the innovation system. Consequently, any structural breakages[1] located must always be interpreted and categorized in a qualitative way.

For Germany, the most relevant system shocks are those due to changing territory and thus population and governance. In this contribution, we consider, for example, the size of the Empire or the federal territory. Not only is the German Democratic Republic considered here, but also the Saarland, the Corridor, East Prussia, and others (refer also to Hoffmann, 1965, p. 2). On the basis of today's statistical procedures, we will introduce dummies for these shocks, so that the data series, a priori, do not have to be absolutely consistent with a territory. However, it must be pointed out that, in most cases, the omission of *smaller* districts (such as Alsace-Lorraine from 1871 to 1917) brings in its train less important errors of estimation than the *large variances* in the series of the whole territory of the German Reich (same paper, p. 3).

This contribution is divided into a description of the data (chapter 2) and an econometric part (Chapter 3), in an effort to analyze innovation and growth in Germany. The data chapter is extraordinarily long. The reason for this was mentioned above: this data base was reconstructed for the purpose of this analysis and is not documented elsewhere[2].

2 Basic data on the national innovation system in German states

In this chapter, the reconstructed data base on research and innovation in Germany is presented in full, explained by basic historical events and critically discussed because so many assumptions were required. Later on it will be argued that an econometric analysis is not meaningful between the wars (First and Second World War), but rather for 1850–1913 (first period) and 1951–1999 (second period). We indicate the two periods in the figures, but include the data regarding the period in between, which are not used in the analytical part, to facilitate an own judgement by the reader.

---

[1] Maddison (1982, p. 2 and 83) talks about "system shocks"; compare also Wagner (1984). Machlup (1957) found 25 variants in the economic literature of what "structural change" could mean. Gerschenkron (1943) points to the pervasive institutional powers that may overcome external shocks for decades.

[2] For German readers, see Grupp et al. (2002). A detailed English list of data sources is available from the authors on request.

2.1 Public expenditure for national science and technology

Traditionally, the development of science and technology is measured by the number of *scholars*. In this way, for example, Gascoigne (1992) submitted a historical demography of the scientific community between 1450 and 1900, by listing the nationality and age of all the scientists. According to this study, Italy was the leading scientific country at the beginning of modern times (in the late 15^{th} century), representing about half of all the scientists in the world. This remained almost unchanged during the entire Middle Ages, before that century; then exponential growth with a doubling period of approximately 50 years took place.

Detailed and complete statistics are available regarding the *scientific staff* in Germany since the foundation of the Empire, accessible via today's electronic means. Generally accessible statistical material about R&D personnel in Germany has only been recorded since the 1960s (in the framework of the Federal Research Report, which has been published since 1965).

Another traditional approach to the empirical definition of the importance of an innovation system is scientific expenditure (the sum of R&D funds and those for training, teaching, maintenance and the diffusion of knowledge). Whereas the evaluation of expenditure for pure educational and R&D institutions is rather simple, this is more complicated in the case of institutions engaged in *both* research *and* teaching. Quotas were adopted to cope with the individual fields of specialization, as well as with the individual types of universities. However, it is questionable whether these reflect the right proportions between the percentage of research and that of teaching at all *historical* points in time. In addition, not only is the historical consideration problematic, but also the consideration of the present time. Nevertheless, it is common statistical practice in all OECD countries to work with such quotas (Hetmeier, 1990; Irvine et al., 1990). Expenditures for training, maintenance, diffusion, marketing and other innovation-related activities are not included.

Pfetsch (1982) added up scientific expenditures between 1850 and 1975, so that rough estimates of the degree of R&D financing could be derived. However, these data records only include *public* expenditure, and so disregard the private sector. Consequently, industrial innovation indicators must be researched separately (see chapter 2.3). In order to avoid dealing with the difficulty of different currencies, the development of scientific expenditure can be best evaluated by the percentage of the *total expenditure* of public budgets.

Accordingly, scientific expenditure in the German regions prior to the foundation of the Empire was approximately 1% (see 1). Linked to the foundation of the Empire, this percentage reached more than 2%, but dropped to almost 1.7% between the 1880s and the First World War. This reduction should be interpreted with care, as the reduced share in total public budgets merely tells us that, in these times of pre-war rearmament and a booming economy, government outlays were ballooning (Ziegler, 2000).

The Republic of Weimar attained a doubling of scientific financing which, however, was lost again due to the world wide economic crisis. Recovering from hyperinflation, a second booming scientific phase was set off, along with "formidable creativity and experimentation enthusiasm" (Ambrosius, 2000). The Nazi arma-

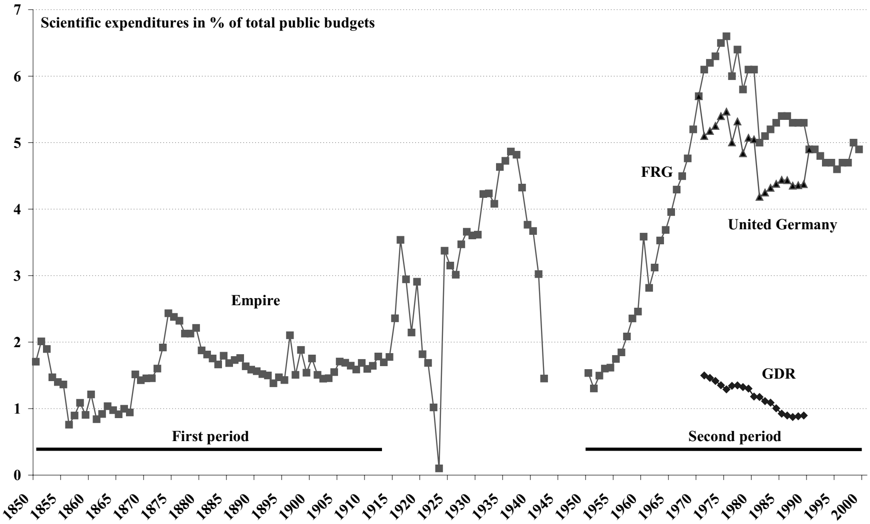


Fig. 1. Development of scientific expenditure in proportion to the total expenditure of public budgets
Source: Pfetsch (1982), Echterhoff-Severitt and Stegemann (1990), BMBF (2000).

ment research from circa 1935 onwards was financed with fiscal tricks and budget deficits. In West Germany, the support of science was pushed dramatically to reach a proportion of 6.5% of all public budgets by the 1970s (university expansion), but fell off to approximately 5.5% because of German re-unification. Finally, due to re-unification, the level dropped even further. These indications are based on the numbers of Empire or Federal institutions and those of regions and states. Surprisingly, the post-war-II expenditures start at around the same level as at the beginning of the last century (after World War I), which is at the same level as the endpoint of records in the Second World War period, and increases in a similar way after the war II as it did after World War I. This points to quite *stable* and *persistent* institutional structures underlying the financial totals.

The financial support of *Research* and *Development* is typical for post-war Germany. Until 1945, the financing share for R&D only played a subordinate role in total scientific expenditure. Although the research share[3] was 20 to 30% during the first period after the foundation of the Empire, it dropped to less than 20% by the beginning of the First World War. In addition, it is important to note that a great deal of scientific expenditure by the Empire was used for defense tasks shortly after the foundation of the Reich. During the Weimar Republic and the Third Empire, the R&D share of the total of scientific expenditure continued to fluctuate around 20% (industrial research not included). A quick increase in the R&D share of scientific expenditure was the case when research in certain areas was admitted again in the young Federal Republic, after the signing of the Treaty of Paris in 1955: at times it reached 70% and has only declined due to the recent re-

---

[3] More precisely, "research share" means the "R&D share" of the total expenditure for science and technology.

unification. Regarding the R&D expenditure of the *German Democratic Republic*, note that the individual statistics were centrally maintained and are comprehensive. However, the conditions which were applied do not fully comply with those used by OECD countries and often show exaggerated values. Following re-unification, the relevant statistics were revised and adapted to Western standards. However, the conversion problem of the East German bank's Mark (M) persists. Due to the non-convertibility of this currency, the reliable purchasing-power parity values of OECD countries cannot be applied.

In order to be certain, we applied a pessimistic and an optimistic variation to show a range of uncertainty due to conversion. The first possibility of conversion is based on the purchasing-power parity (PPP) of so-called baskets of commodities. In the second model, the subsidies included in GDR commodity prices are taken into consideration and deducted (Anonymous, 1986, p. 259-268). In both estimations, the national R&D expenditures of the German Democratic Republic could not quite equal the West German level (per head of the population), but the general downward trends (in 1) resemble each other somehow. This may come as a surprise to those who point to the inefficiencies of the communist part of Germany, but, again, the underlying institutional structures remained basically the same as before the war, requiring similar amounts of public support. Again, this points to persistent basic structures in the national innovation system. In the econometric model, we did not add the East German expenditure to the ones of West Germany to avoid conversion problems.

2.2 Development of scientific activities

It is impossible to achieve an insight into the development of non-codified and thus "tacit" experienced knowledge of the scientific staff. For this reason, the historical development of an innovation system is often shown by personnel statistics, or by statistics showing monetary expenditures. However, only expenditure is measured by this method, instead of the fruit of scientific activities. Efficiency measurements are particularly impossible. Consequently, modern innovation statistics make regular use of yield measures; regarding scientific work, statistics of publications are a typical output indicator. Analysis of the degree of *publication* activities have been maintained for centuries. However, it must be noted that the publication media chosen by scientists may differ from one faculty to another, as well as over time (Wagner-Döbler and Berg, 1996, p. 289). Only during the 19^{th} century did scientific magazines achieve the same degree of significance as books, the dominating publication media until then. From 1900 onwards the availability of data improved world wide. The growth rates of periodicals were evaluated by Mabe and Amin (2001).

Analyzing the situation in *Germany*, the Science Citation Index (SCI), which has been available as an online version as early as 1974 (see below), has a printed version listing the publications from 1945 until 1974. Although no indications are found regarding the authors' nationalities or the institutes' locations, the listing of periodicals is classified by the countries editing and printing them. The repeatedly written announcement by the SCI that records would be completed back to 1900 was withdrawn[4], so there is no hope for the early publication of a *century's inventory*. Although the statistics of publications seems to be an interesting output indicator, it cannot establish the first period and thus cannot be used for estimating the econometric model. However, as we are analyzing structural issues such as persistence and paradigm shift, we can use this indicator for an assessment of the GDR research, which is difficult in monetary form (see above in section 2.1).

From 1974 to 1990, SCI publications from West and East Germany can be compared electronically. In the 1970s, the share of *East German* publications among all German publications was approximately 16 to 17%. However, if one compares East and West Germany, both the proportional shares of population and the proportion of R&D staff is almost 30%, so that scientific publications from the German Democratic Republic are less well-represented in the US-based database. The proportion of East German publications had constantly diminished, to reach 13% by the end of the 1980s; and there is no answer to the question as to whether the *representation* in the database was even worse or if the output efficiency of East German research activities continued declining until the end of this state[5].

Measured by its publication output, the *profile* of GDR research resembles that of the former Federal Republic. This *structural similarity* could be the reason for such a strong diminution of publication activities on an all-German level following re-unification. Integration did not concern differently specialized East and West research systems, but research systems with the same principal orientation, which led to the deplorable "re-allocation and consolidation" in East Germany. Independently from a political evaluation of the organization of GDR research institutes, this *structural persistence* must be pointed out; obviously 40 years of division were not sufficient for a differentiated development of the basic specialization patterns of research in both parts of Germany. To a great extent, and in the sense of *path dependency*, research is still based on the (common) preferences which existed prior to the division. This unique historical situation may be understood as an unintended experiment: basic patterns of scientific structure change only slowly, even in times of great political system change (Hinze and Grupp, 1995, p. 65).

2.3 Industrial research and development in Germany

Since the foundation of the Empire, the economic growth of industrial countries, and that of Europe in particular, has increasingly been based on the innovation energy of the *knowledge-based industry*. "This is undeniably true for the impulses of growth immediately released by these industries, starting with carbon chemistry and electrical technology" (Wengenroth, 1997). There is hardly a clearer and more distinct way to describe the effects of industrial research on the culture and efficiency of innovation.

It is still difficult to prove the companies' increasing R&D expenditure for such an undeniable success. In particular, no complete data records are available re-

---

[4] Personal communication Garfield, 14 October 2000.

[5] Due to the delay in appearance of scientific publications following submission, no quantitative cutback in literature production by the researchers of German Democratic Republic institutes can be perceived until the end of 1990 (Weingart et al., 1991, p. 4).

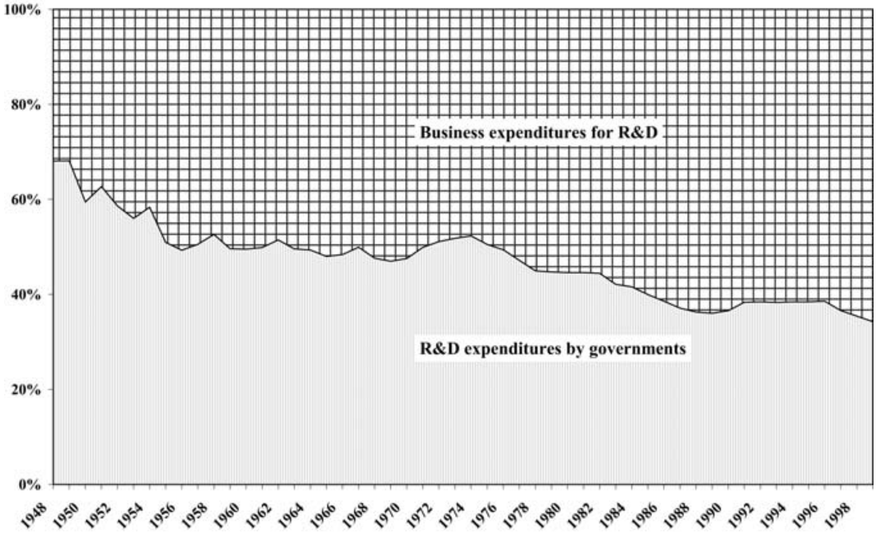


Fig. 2. Development of government and industrial R&D expenditure in relation to each other from 1948–2000

Source: BMBF (2000).

garding monetary expenditure or research personnel prior to the end of the Second World War, i.e. the data record established by Pfetsch (1982) regarding public scientific expenditure has no counterpart for industry. Today's statistics about R&D expenditure and personnel of the Federal Republic systematically start from the year 1962; certain presumptions allow the reconstruction of the corresponding indicators starting from 1948/49 (Fig. 2). According to this, industry has continuously increased its R&D budgets to a higher degree than government, the share of which is presently approximately 40%. R&D expenditure of the business sector will be included in the analytical model for the second period only. This is consistent with the assumption that these were small or negligible in the first period.

2.4 Development of technological activity in Germany

The observation of the development of innovation activity is important in itself in order to establish R&D results, mostly on a technological or application level. Adopted methods are statistics on *patent applications* (a figure representing successful innovation activity seen from the innovators' or applicants' subjective perspective) on the one hand, and, on the other hand, statistics on the number of *granted patents* (as a figure representing successful innovation activity, seen from the objective perspective of patent examiners). Statistics on patents make even more sense if one takes into consideration that only fragments of industrial R&D expenditure are known prior to the Second World War. Instead of inputs, industrial R&D activities can be measured by their patent outputs, and this even more precisely from a technological perspective than by monetary indicators. This also explains our interest in

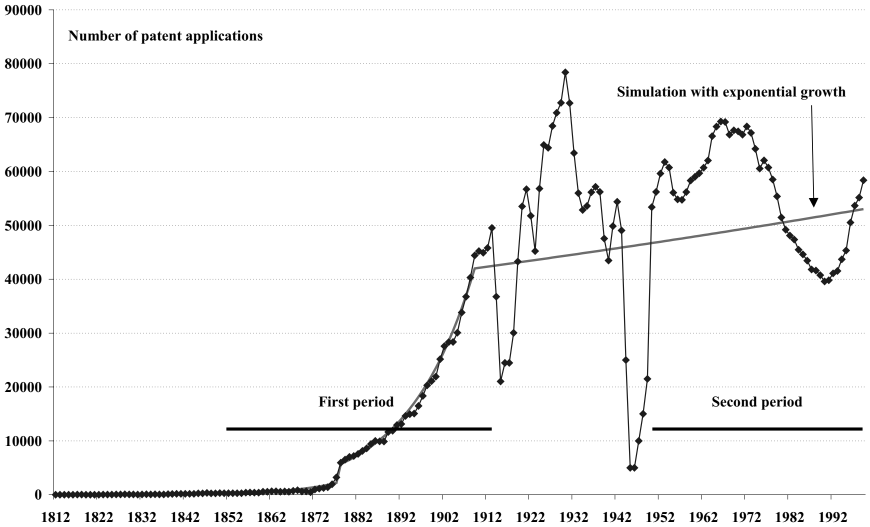


Fig. 3. Development of granted patents (from 1812 to 1877) and domestic patent applications from 1878 to the present.

Source: Federico (1964), German Office for Patents and Trademarks (several years).

both patent grants and patent applications: if no patent is granted after verification of the novelty, the inventive step and its commercial usefulness, for example due to a lack of novelty, the applying company nevertheless may have invested in R&D efforts – even if these led to an objectively already known result. Consequently, the “subjective” perspective of a successful invention is closely linked to the R&D performance which was in fact realized. Statistics on patent applications as a *proxy variable* for R&D expenditures may ignore whether the object of the invention was a world novelty or not. R&D expenditure also includes the costs of unsuccessful or belated inventions in comparison with competitors (imitations).

The periods to be considered are fully included in the statistics of patents. In some German regions, patents were applied for as early as 1820, starting from the South due to the influence of the Napoleonic legislation. From July 1st, 1877, a *patent act* for the German Empire standardized procedures. Thus, the creation of patent acts in Germany follows the scientific-technological innovation push of the 19th century, at the end of which Germany was one of the leading industrial nations. In about the middle of the century, the local, largely secluded markets were dissolved, and the German economy was integrated into the quickly expanding world economy (Ziegler, 2000, p. 198 and North, 2000, p. 13).

Since 1879, *patent statistics* have been available using *machine readable* methods. The electronic data records since 1970 are more informative than those of former periods, leading to an increased importance and use of these patent data records by modern studies in science and technology. But if one makes the effort to chain together different patent data records for the appropriate historical sequences including written sources, assembled patent statistics can be established for the whole period (see Dominguez-Lacasa et al., 2003). Considering global patent activities in Germany (Figure 3), the strongest growth at a low level takes place from 1820 to the foundation of the German Empire. The total growth rate for German regions is shown to be constant with the setback due to the war of 1870/71. Following the introduction of the countrywide German patent act, the number of applications and grants rises rapidly within a few years, and continues growing at a constant rate, which, although at a considerably higher level, is lower compared with that of the period preceding 1870. This growth, which lasted for almost one century, is abruptly stopped by the First World War, the annual patent production being halved. From approximately 1920 to 2000, an eventful development pattern nevertheless shows growth close to zero. During almost one century, the number of annual patent applications is approximately 50,000 to 60,000. German patent productivity per person reaches one of the highest rates, in comparison with the United States, Japan, and the European Union. Diverging from this rough rule, growth is observed during the Weimar Republic phase until the beginning of the Third Reich, followed by a very deep setback during the Second World War, which is distinctly more serious than that of the First World War, and a return to the secular quota by approximately 1960. Another boom follows until 1975, when a deep recession takes place which is only overcome in the mid 1990s.

No investigation has yet discovered whether these growth cycles have *only* economic causes. The economic boom around the foundation of the Empire is well-known (Ziegler, 2000, p. 201; Stolper et al., 1964); the same is true for the serious recession following the oil crisis in 1973, straight after the "economic miracle". The question remains as to whether the reduction of innovation activities at the beginning of the Third Reich was only due to economic reasons or to a modified practice of patenting (for example, by the stronger observance of secrecy due to the early war economy, by expulsion or migration of Jewish scientists). Further, the question is asked as to why the growing R&D budgets granted after the Second World War did not lead to an increase in patent activities. Obviously, this decrease of patent efficiency was not exclusively driven by R&D inputs.

The analysis so far includes all patent documents of national and international actors on German territory. By *international actor* is understood that either the inventors' residence or the applying company is located abroad. From 1881 to 1913, the share of foreign patent grants was extremely high, showing an average of 35%: until 1933, Germany's reputation as the leading scientific country attracted many young scientists from abroad, especially Americans who came to the German Empire in order to benefit from practice-oriented education for their degrees, and possibly even to experience some years of active industrial research (Erker, 1990; Smith Jr., 1990). After the First World War and the efforts to achieve self-sufficiency in the 1930s, the share of foreigners was reduced by almost 10% but remained a significant figure in spite of all war speculations. Since the reconstruction of the German patent administration following the Second World War, the share of foreign patent grants has consistently increased, reaching more than 60% in this so-called globalization era. In the analytical part of this contribution, we use domestic patent applications only in order to avoid any influences from changing migration policies.

The basic framework conditions for GDR activities in view of industrial property rights are fixed in the *patent law* of 6 September 1950 (Albrecht et al., 1991,

p. 4). Nevertheless, GDR patent activities according to Western legislation are hard to ascertain during the first years. This is linked to the various forms of recognition of the GDR as an autonomous state by different nations. Some GDR inventors operated from Federal Republic addresses. Compared with Western conditions, certain deviations in the patent law conditions of the *former GDR* were ruled by the socialist spirit of ownership. Consequently, the national patent applications at the former GDR Authority of Invention and Patent Administration (AfEP) can hardly be compared with those submitted in the West (Hinze and Grupp, 1995, pp. 42). Therefore, another method was chosen for the analysis summarized in the next paragraph, which is based on GDR patent activities in West Germany. With the help of this method, all the particularities related to patent law specifications are circumvented, enabling comparison with Western countries. GDR inventors were mostly interested in the economic sector of the former Federal Republic, so that the foreign applications submitted for this target market can be referred to (independently of whether the application was submitted to the German Patent and Trademark Office, to the European Patent Office, or to the International Patent Authority WIPO designating the Federal Republic of Germany).

A comparison of the *specialization* of GDR patent portfolios with those of West Germany is very interesting. According to a division of the whole technology area into 28 subfields, the specialization profile was constant over time. In particular, the eastern regions' patent profile of the 1990s (including East Berlin) corresponds largely with that of the GDR of the 1980s (Schmoch and Sass, 2000). In addition, there is an amazing correlation with that of West Germany. In spite of completely different economic conditions (Stolper et al., 1964), large fields of technology show a *correspondence* between East and West Germany until re-unification (Grupp and Schmoch, 1992, pp. 118). This was also found for the area of basic research (publication statistics) and can be explained by path dependencies and *persistent* structures in both parts of Germany despite their different political regimes (chapter 2.2).

When Germany was re-unified in 1990, two almost identically specialized technology systems came together. It was not possible to integrate the strength of one side and the weakness of the other one. Instead, the fields characterized by strength were the same on both sides of the "iron curtain" and the weaker fields were equally neglected. Any particular incentives to growth and innovation in the unified country are rather unlikely and thus, again, the German innovation system turns out to be very stable. This is an encouragement to undertake an econometric comparison of 1850–1913 with 1951–1999.

3 Statistical explanation of technical progress and welfare growth in Germany

In this section, we implement an econometric analysis using the economic and innovation indicators discussed in the previous sections. Our goal is to improve our understanding of the processes of innovation and growth. We compare the behavior of the variables in two periods, (1850–1913) and (1951–1999), by paying special attention to the effect of human capital, technology advances and expenditure in scientific activities to improve the standard of living. Within the scope of this contribution, we cannot go into the details of econometric analysis[6]

3.1 Variables and methodology

Our main goal is to identify the causal relationships underlying the innovation process on a macroeconomic level. Our approach is not to *define a priori* any variable as *endogenous*. This is because, with regard to econometrics, one criticism in the literature is that "there has been remarkably little attention paid to the problem of the endogeneity of the different variables used as regressors in modelling and testing growth theories" (Durlauf, 2001, p. 66). Instead, applying the state-of-the-art definitions of S&T indicators, we implement VAR tests and use the suggested causal relations to test a SURE (seemingly unrelated regression equations) model[7].

Model openness in empirical growth analysis is associated with a serious issue, the choice of variables (Durlauf, op. cit.). Therefore, new procedures that can assess the sensitivity of the estimates are required (e.g. impulse response functions, see below). In doing so, we add to the indicators that were the subject of analysis in the previous section, the usual variables for output that were not discussed above. Instead of the well-known human-capital variables such as schooling etc., which have proven to be neither very robust nor convincing (Weber, 1998, p. 49), we use the student numbers in higher education (for historical data sources see Titze et al., 1987). We have to admit that there seems to be no single privileged way to conduct empirical growth analysis, but to a certain degree this requires assumptions that cannot be falsified within the econometric procedure (Brock and Durlauf, 2001, p. 265).

The proxy specification is as follows (all are per capita and taken as logarithms):

- Human capital in Germany: the number of students in higher education (universities and technical schools) as a share of total population (variable LSTUDPK);
- Technology advances in Germany: domestic patents granted at the German Patent Office (LPATDPK);
- Government (and industry) participation in the innovation process: public (and private) expenditure in R&D activities in real terms (LEXPRPK);[8]
- Economic demand: Net National Product (in the first period) and Gross Domestic Product (in the second period) at constant market prices (1913 and 1991) per capita (LNSPRPK or LBIPRPK, resp.).

For each period, the first step checks the time series for the existence of unit roots. To address the causality issue, we then run the Granger causality tests

---

[6] See the full paper on econometric methodology by Jungmittag et al. (2004).

[7] See Zellner (1962, 1963); Zellner and Huang (1962), as well as Greene (2000, pp. 614–36) for a textbook presentation.

[8] To obtain public and private expenditure in R&D in real terms, we applied the price deflator of the Net National Product (in the first period) to the public science expenditure and that of the Gross Domestic Product (in the second period) to the public and private R&D expenditure at current prices.

(Granger, 1969) based on unrestricted vector autoregression models (VAR)[9]. These models normally contain some nuisance parameters which are eliminated in the third step by identifying an admissible restricted VAR model. This model serves as a starting point to identify a SURE model the equations of which only include in each case the significant variables and takes into account the contemporaneous correlations between the error terms of the individual equations.

3.2 Stationarity and causality in the first period (1850–1913)

To test for stationarity, we implement Augmented Dickey Fuller Unit Root Tests (ADF), which adequately take into account temporary additive and innovative outliers as well as structural breaks[10]. From the results it is clear that, in the first period (1855–1913), all variables are trend stationary[11]. This implies that we can model the variables as they are and that possible spurious correlations can be avoided by including deterministic trends. Yet, we need structural break dummies for the economic boom starting in 1872 (D72L) and ending around 1878 (D78L, D79L, D7278, resp.), this period also being relevant for the start-up of the national patent practice following the first national patent law. Student numbers dropped from 1870 (D70L) to 1871 (D71L) for a short period and increased strongly in 1895 (D95L).

To test for Granger causality in the first period (1850–1913), an unrestricted vector autoregression model (VAR) was first estimated. The many significant results are not reported here (compare Jungmittag et al., 2004). In order better to understand the dynamics of the model, we calculated the generalized impulse response function following Pesaran and Shin (1998). Impulse response functions measure the temporal profile of the impact of a shock at a certain point in time for the future values of the variables. From these sometimes seemingly contradictory profiles, we concluded that the model can be formulated more parsimoniously and therefore, in the next step, switched over to a restricted VAR model with one exogenous variable, namely the share of students which is not influenced by the other variables. The results, again not reported here, indicate that the model can be simplified even more and that more explicit parameters can be obtained. Therefore, in the final step, we further reduced the number of variables from the equations in a SURE model still

---

[9] To analyze the empirical relationship between two stationary variables, the Granger test verifies whether past values of one can help to predict current values of the other. There are different types of Granger-causality analyses (see Hamilton, 1994, p. 302–9). Since we have several variables, bivariate tests may give rise to confusing results. For instance, there might be an effect of public expenses on output that in fact is due to human capital. This effect might be erroneously allocated to public expenses if the variable human capital is not included in the equation. Accordingly (and even though we are now not primarily interested in the dynamic properties of the innovation process) we decide to test for causality in a multivariate context, estimating a Vector Autoregressive Model (VAR). In this case the equations include lags of all variables under consideration.

[10] To this end, the tests proposed by Cati et al. (1999) (taking into account outliers), Perron (1989); Perron and Vogelsang (1993); Lumsdaine and Papell (1997) as well as Ben-David et al. (2003) (taking into account structural breaks) are applied. All estimations and statistic tests are carried out with Microfit 4.1.

[11] For details, see Jungmittag et al. (2004). Values from 1850 to 1854 were taken to calculate lagged first differences.

Table 1. Domestic patents granted as dependent variable 1855–1913

Regressor	Coefficient	Joint significance	Sum of coefficients	Long-term coefficients
Constant	-12.6984			
	(0.000)			
LNSPRPK(-1)	-1.3104			
	(0.007)			
LNSPRPK(-2)	0.4511			
	(0.447)		0.8764	1.2665
LNSPRPK(-3)	0.5672	(0.000)	(0.000)	(0.000)
	(0.343)			
LNSPRPK(-4)	1.1685			
	(0.016)			
LPATDPK(-1)	0.3080			
	(0.001)			
LEXPRPK(-5)	0.2101			0.3036
	(0.046)			(0.025)
D71L	-0.4375			
	(0.004)			
D72L	0.7917			
	(0.000)			
D78L	0.3259			
	(0.003)			
	$R^2_{\text{adj.}}$ = 0.9846	DW = 1.7395		

Notes: Levels of significance in brackets.

ensuring that the common significance of variables with all their lags was given. As a result, it was found that the public expenditure on science is mainly influenced by per-capita income in the long term (statistical coefficients not reported here). Also highly significant is the share of students in the total population (lagged by two years). The lagged patent grants (by one to three years) are individually and jointly highly significant, but because of opposing signs, neither the short-term nor the long-term effect is statistically different from zero. Thus, public involvement in science is mainly influenced by increasing demand (or standard of living of the population) and the intellectual potential or human capital available.

Table 1 shows the corresponding relations for patent activity. Again, the highly significant positive influence of per-capita income is visible: in the long run, patent grants per-capita increase by 1.3% if per-capita income grows by 1%. Interestingly, public science expenditure also exerts a highly significant influence on patent granting with a lag of five years.

On the other hand, net domestic product per-capita is significantly influenced by patent grants and student numbers (joint significance). However, if one differentiates between short-term and long-term impacts, the patent variable remains positive but slightly below the weak significance level of 10%. The influence of student enrolment remains negative (see Table 2).

Table 2. Net domestic product as dependent variable 1855–1913

Regressor	Coefficient	Joint significance	Sum of coefficients	Long-term coefficient
Constant	2.3332			
	(0.005)			
LNSPRPK(-1)	0.2708			
	(0.025)			
LNSPRPK(-2)	0.1581			
	(0.224)			
LNSPRPK(-3)	-0.2432			
	(0.038)			
LPATDPK(-2)	-0.0400			
	(0.066)		0.0201	0.0247
LPATDPK(-3)	0.0601	(0.016)	(0.122)	(0.123)
	(0.007)			
LSTUDPK(-1)	-0.1533			
	(0.170)			
LSTUDPK(-2)	0.0105			
	(0.917)		-0.3120	-0.3832
LSTUDPK(-3)	0.1441	(0.000)	(0.000)	(0.000)
	(0.179)			
LSTUDPK(-4)	-0.3133			
	(0.002)			
T	0.0176			
	(0.000)			
D7278	0.0843			
	(0.000)			
	$R^2_{\mathrm{adj.}}$ = 0.9901	DW = 1.9001		

Notes: Levels of significance in brackets.

Altogether these results confirm the linear relationship between innovation activity, both on the side of inputs and outputs and increasing demand as a consequence of growing per-capita income in the period from 1855 to 1913. Additionally, public science expenditure increased innovation output considerably, although this was in a competitive relation to human capital.

3.3 Stationarity and causality in the second period (1951–1999)

The starting point for analyzing the second period between 1951 and 1999 was once again the assessment of structural breaks. For all time series, the structural breaks were definite, namely one around the year 1972 (oil price crisis) and the other for the “new” unification of Germany in 1991. These structural breaks, unlike the first period, did not only cause shifts in the levels of the variables involved, but also affected the trends themselves. This is particularly true for student enrolment and thus the expansion of the academic system after 1972. We use D72L and D91L to

Table 3. Public and private R&D expenditure as dependent variable 1956–1999

Regressor	Coefficient	Significance level	Long-term coefficient	Significance level
Constant	-1.5173	0.000		
LBIPRPK(-1)	0.3669	0.034	0.8490	0.000
LEXPRPK(-1)	0.5679	0.000		
T	0.0304	0.000		
D72L	0.3534	0.000		
D72L*T	-0.0213	0.000		
D91L	-0.1857	0.000		
	$R^2_{\text{adj.}}$ = 0.9987	DW = 1.7970		

Table 4. Student numbers per population as dependent variable 1956–1999

Regressor	Coefficient	Significance level	Long-term coefficient	Significance level
Constant	-4.1225	0.000		
LPATDPK(-1)	-0.0855	0.001	-0.1473	0.000
LSTUDPK(-1)	0.4199	0.000		
T	0.0264	0.000		
D72L	0.2493	0.000		
D91L	0.8681	0.000		
D91L*T	-0.0256	0.000		
	$R^2_{\text{adj.}}$ = 0.9984	DW = 1.3986		

denote the break in levels, and D72L*T and D91L*T to denote the trend changes. In the second period investigated, the time series for science expenditure and students are definitely trend stationary[12], whereas patent data seem to be ambivalent and only the gross domestic product is not stationary. However, in order to arrive at a comparable analysis to the first period, we analyze the variables as they are.

Again, the unrestricted VAR model and an analysis of the impulse response functions show that the full model can be simplified and accumulated considerably (in particular using the results of the restricted VAR model; see Jungmittag et al., 2004, for details). By reducing the non-significant variables from the equations in the SURE model, we arrive at the following relations for total R&D expenditure (Table 3). Exactly the same as in the first period in the 19^{th} century, R&D expenditure is mostly driven by per-capita income in post-war Germany. But in this period, there is no positive influence of human capital, which may be due to the fact that private R&D expenditure is included here as well. So it is interesting to learn more about the explanation of student enrolments per capita (Table 4).

The relative number of students in the population is explained by a highly significant negative influence of per-capita patent grants. This points to the fact

---

[12] Here, values from 1951 to 1955 were taken to calculate lagged first differences. For details of the ADF tests see Jungmittag et al. (2004).

Table 5. Domestic patents granted as dependent variable 1956–1999

Regressor	Coefficient	Significance level	Long-term coefficient	Significance level
Constant	18.5941	0.007		
LBIPRPK(-1)	-3.0050	0.001	-3.9870	0.002
LPATDPK(-1)	0.2463	0.058		
LEXPRPK(-1)	0.8250	0.010	1.0945	0.016
D72L	-0.6515	0.002		
D72L*T	0.0373	0.000		
D91L	-0.2468	0.018		
	$R^2_{adj.}$ = 0.6910	DW = 2.1815		

Table 6. Gross domestic product as dependent variable 1956–1999

Regressor	Coefficient	Significance level	Long-term coefficient	Significance level
Constant	4.2277	0.000		
LBIPRPK(-1)	0.5714	0.000		
LPATDPK(-1)	0.0250	0.204	0.0584	0.191
T	0.0180	0.000		
D72L	0.1969	0.002		
D72L*T	-0.0097	0.001		
D91L	-0.0609	0.000		
	$R^2_{\mathrm{adj.}}$ = 0.9963	DW = 1.7691		

that there is still a competitive situation between public and private resources of national innovation processes.

For the explanation of relative patent numbers in the short term, the negative influence of per-capita income is surprising (Table 5). However, considering our analysis of the impulse response function for the non-restricted VAR model, one can expect that a small positive effect will occur for longer time periods. In addition, there is a very strong positive influence of gross national R&D expenditure. In the long run, we can assume that there is a one-to-one relation between the increase in gross national R&D expenditure and patent grants. Finally, we look at the gross national product in Table 6. It is confirmed that all the innovation-related variables do not explain gross domestic product. This is also true for the long-term impact of patent grants.

Finally, our results for the second period point to the strong influence of demand on innovation performance, or, in other words, an increasing standard of living boosts innovation activities. What is quite different from the 19^{th} century is the fact that patent grants now depend on both public and private R&D expenditure and are no longer influenced by demand. This certainly reflects the continuing importance of public contributions to innovation in this period and also the intertwining of science (mostly supported by public sources) and technology (mostly supported by the private company sector) rather than a linear (sequential) relation, as was the case in the 19^{th} century. A different picture emerges, however, if the more traditional variables of capital and labor are added to the model (see Jungmittag et al., 1999).

4 Discussion and conclusions

The view into historical innovation reveals many interesting perspectives: Most astonishingly, the German innovation system was very *stable*, although it witnessed several political system changes in the past century. The total amount of government spending on science and innovation followed similar quantitative tracks after its formation in the 19^{th} century, the First World War and the Second World War. However, considerable differences are observed when regarding the strong role of enterprises on innovation after the Second World War which was – in pecuniary terms – not as visible before. In terms of the basic sectoral structures in science and technology, the strong and the weak sides were almost the same, whatever regime and territorial boundaries existed. This *persistence* of the innovation system points to a *resistant innovation culture* in and around Germany, which may not be influenced much by external shocks or incentives be it in monetary or institutional form. Even the isolation of the former GDR and its subjection under the communist regime could not change much. This sustainable culture imprint can only be analyzed and detected in historical time series.

The industrial research system in Germany was one of the first in the world to be formed and developed. Other countries followed that pattern more or less closely. The suggested range of indicators on a national level gives a detailed impression of both the extent and the contents of innovation activities during more than the past hundred years. The empirical base is now broadened to a large extent, so that there is no longer a serious *empirical gap*.

As the German innovation system has turned out to be stable even though it witnessed several political system changes in the past century, an econometric comparison of 1850–1913 with 1951–1999 is allowed. The econometric analysis is performed in three steps in order to arrive at an appropriate, and at the same time parsimonious, model. In the first step, the statistical characteristics of the single time series are clarified, with a special attention to structural breaks. In the second step, an unrestricted VAR model is estimated, while in the final step some nuisance parameters are eliminated (SURE model). In the first period, the period of the formation of a liberal, unified and large market with ongoing "scientification" of technology, the overall empirical results point to a linear relation. Per-capita income as a variable for economic demand, as well as student numbers (as a share of total population), explain public science expenditure. These – lagged by five years – explain, together with per-capita income, the number of patents granted and thus the growth of the technological potential in Germany. The net domestic product (per capita) is positively, but not significantly, influenced by patent numbers in a competitive situation with human capital (student numbers have a negative influence). Overall, in the early period, the results point to a strong influence of growing demand on innovation activities.

Yet, for the second period in the post-war twentieth century, a different logic in the innovation process is suggested: Now, the total R&D expenditure (public and private) are driven by per-capita income, while patent output does not follow demand but only R&D expenditure. The real domestic product seems not to depend on innovation activities; there seems to be only a very weak but positive influence of patent numbers among all possible variables. A different picture arises, however, if the more traditional variables of capital and labor are added to the model (see Jungmittag et al., 1999).

The future research agenda should include more such long studies of innovation systems. The basic findings for Germany should be compared to other countries that possibly suffered less from territorial and political changes. The data used in this article should exist in other countries as well and may be brought to the surface. Also, we need more sectoral studies in order to work out typologies of innovation development over long periods. Altogether the results achieved so far should encourage more such research based on the evolutionary understanding of long-term development.

References

Albrecht E, Dohnert O, Schneider M, Bourcevet H (1991) DDR-Forschung im internationalen Vergleich unter Zugrundelegung der Patentstatistik. Sektion Wissenschaftstheorie und Wissenschaftsorganisation

Ambrosius G (2000) Von Kriegswirtschaft zu Kriegswirtschaft 1914–1945. In: North, 282–350., C. H. Beck

Anonymous (1986) Das Kaufkraftverhältnis zwischen D-Mark und Mark in der DDR 1985. DIW-Wochenbericht 21(53): 259–268

Bannock G, Baxter RE, Davis E (1998) Dictionary of Economics, 6th edition. Penguin, London

Ben-David D, Lumsdaine RL, Papell DH (2003) Unit Roots, Postwar Slowdowns and Long-Run Growth: Evidence from Two Structural Breaks. Empirical Economics, 28: 303–319

Bernal JD (1939) The Social Function of Science. George Routledge & Sons, London

BMBF (2000) Bundesbericht Forschung. BMBF, Bonn, (see also earlier editions)

Brock WA, Durlauf SN (2001) Growth Empirics and Reality. The World Bank Economic Review 15(2): 229–271

Cati RC, Garcia MG, Perron P (1999) Unit Roots in the Presence of Abrupt Governmental Interventions with an Application to Bazilian Data. Journal of Applied Econometrics, 27–56

Dominguez-Lacasa I, Grupp H, Schmoch U (2003) Tracing technological change over long periods in Germany in chemistry using patent statistics. Scientometrics, 57(2): 175–195

Durlauf SN (2001) Manifesto for a growth econometrics. Journal of Econometrics, 100: 65–69

Echterhoff-Severitt HH, Stegemann WH (eds.) (1990) Forschung und Entwicklung in der DDR, Daten aus der Wissenschaftsstatistik 1971 bis 1989. SV-Gemeinnützige Gesellschaft für Wissenschaftsstatistik mbH im Stifterverband für die Deutsche Wissenschaft

Erker P (1990) Die Verwissenschaftlichung der Industrie. Zur Geschichte der Industrieforschung in den europäischen und amerikanischen Elektrokonzernen 1890–1930. Zeitschrift für Unternehmensgeschichte, 35: 73–94

Federico PJ (1964) Historical Patent Statistics. Journal of the Patent Office Society 46(2): 89–171

Freeman C (1969) Measurement of output of research and experimental development. Statistical Reports and Studies

Freeman C (1992) The Economics of Hope. Pinter Publishers, London

Freeman C, Louçã F (2001) As Time Goes By – From the Industrial Revolution to the Information Revolution. Oxford University Press, Oxford

Gascoigne R (1992) The Historical Demography of the Scientific Community, 1450–1900. Social Studies of Science

Gerschenkron A (1943) Bread and Democracy in Germany. Cornell University Press, Ithaca

Granger CWJ (1969) Investigating Causal Relationships by Econometric Models and Cross-spectral Methods. Econometrica 37: 24–36

Greene WH (2000) Econometric Analysis. Prentice Hall International, London

Grupp H (1998) Foundations of the Economics of Innovation Theory, Measurement and Practice. Edward Elgar, Cheltenham

Grupp H, Dominguez-Lacasa I, Friedrich-Nishio M (2002) Das deutsche Innovationssystem seit der Reichsgründung. Physica-Springer, Heidelberg

Grupp H, Schmoch U (1992) Wissenschaftsbindung der Technik. Physica-Verlag, Heidelberg

Hamilton JD (1994) Time Series Analysis. Princeton University Press, Princeton, N. J.

Hetmeier HW (1990) Öffentliche Ausgaben für Forschung und experimentelle Entwicklung 1987. Wirtschaft und Statistik 2: 123–129

Hinze S, Grupp H (1995) Ein Rückblick auf Wissenschaft und Technik in der ehemaligen DDR: Ostdeutschlands Forschungs- und Entwicklungspotential. In: Holland D, Kuhlmann S (eds.) Systemwandel und industrielle Innovation, 41–86, Physica-Verlag

Hoffmann WG (1965) Das Wachstum der Deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts. Springer, Berlin

Irvine J, Martin BR, Isard PA (1990) Investing in the Future. Edward Elgar, Aldershot

Jungmittag A, Blind K, Grupp H (1999) Innovation, Standardisation and the Long-term Production Function – A Cointegration Analysis for Germany 1960-1996. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 119: 206–222

Jungmittag A, Dominguez-Lacasa I, Grupp H (2004) Dynamische Beziehungen zwischen Innovationstätigkeit und Pro-Kopf-Einkommen in Deutschland von 1850 bis 1913 und 1951 bis 1999. Working paper. Available from the authors on request.

Lorenz C (1997) Konstruktion der Vergangenheit. Böhlau, Köln

Lumsdaine RL, Papell DH (1997) Multiple Trend Breaks and the Unit-Root Hypothesis. Review of Economics and Statistics 79: 212–218

Mabe M, Amin M (2001) Growth Dynamics of Scholarly and Scientific Journals. Scientometrics 51(1): 147–162

Machlup F (1957) Structure and Structural Change: Weaselwords and Jargon. Zeitschrift für Nationalökonomie, 280–298

Maddison A (1982) Phases of Capitalist Development. Oxford University Press, Oxford

North M (2000) Deutsche Wirtschaftsgeschichte. C. H. Beck, München

Perron P (1989) The Great Crash, the Oil Price Shock and the Unit Root Hypotheses. Econometrica 57: 1361–1401

Perron P, Vogelsang TJ (1993) The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis: Erratum. Econometrica 61: 248–249

Pesaran MH, Shin Y (1998) Generalized Impulse Response Analysis in Linear Multivariate Models. Economic Letters 58: 17–19

Pfetsch FR (1982) Datenhandbuch zur Wissenschaftsentwicklung. Zentrum für historische Sozialforschung, Köln

Schmoch U, Sass U (2000) Erfassung der technologischen Leistungsfähigkeit der östlichen Bundesländer mit Hilfe von Patentindikatoren. FhG-ISI, Karlsruhe

Smith Jr JK (1990) The Scientific Tradition in American Industrial Research. Technology & Culture 31: 121–131

Stolper G, Häuser K, Borchardt K (1964) Deutsche Wirtschaft seit 1870. J. C. B. Mohr, Tübingen

Titze H, Herrlitz HG, Müller-Benedict V, Nath A (1987) Datenhandbuch zur deutschen Bildungsgeschichte. Band I: Hochschulen. Teil 1: Das Hochschulstudium in Preußen und Deutschland 1820–1944. Vandenhoeck und Ruprecht, Göttingen

Wagner A (1984) Strukturbruch, Strukturwandel und Evolution in Volkswirtschaften. In: Schiemenz, Wagner (eds.) Angewandte Wirtschafts- und Sozialkybernetik, 333–350, Erich Schmidt

Wagner-Döbler R, Berg J (1996) Nineteenth-Century Mathematics in the Mirror of Its Literature: A Quantitative Approach. Historia Mathematica 23: 288–318

Walter R (1997) Evolutorische Wirtschaftsgeschichte. Wirtschaftswissenschaftliches Studium 26(2): 75–79

Weber AA (1998) Humankapital, Schulbildung und Wirtschaftswachstum: eine kritische Betrachtung der Literatur. In: von Weizsäcker RK (ed.) Bildung und Wirtschaftswachstum, 49–76, Duncker & Humblot

Weingart P, Strate J, Winterhager M (1991) Bibliometrisches Profil der DDR. University of Bielefeld

Wengenroth U (1997) Der Beitrag des Staates zu Forschung und Entwicklung in Deutschland seit der Reichsgründung (Manuscript). Manuscript

Zellner A (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias. Journal of the American Statistical Association 57: 500–509

Zellner A (1963) Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results. Journal of the American Statistical Association, 58: 977–992

Zellner A, Huang D (1962) Further Properties of Efficient Estimators of Seemingly Unrelated Regression Equations. International Economic Review, 3: 300–313

Ziegler D (2000) Das Zeitalter der Industrialisierung. In: North (ed.) Deutsche Wirtschaftsgeschichte, 102–281, C. H. Beck

Nonlinear dynamism of innovation and business cycles*

Masaaki Hirooka

Institute of Technoeconomics, 82, Kamifusa-cho, Koyama,Kita-ku, Kyoto, 603-8143, Japan
(e-mail: hirooka@rr.iij4u.or.jp)

Abstract. The aim of this paper is to describe the nonlinear dynamism of innovation and to clarify the role of innovation for economic development in terms of Kondratiev business cycles, especially the causal relation of the bubble economy and depressions with innovations. Any paradigm of technological innovation develops within a definite time span reaching maturity. This nonlinear nature clarifies many characteristic features of innovation. Schumpeter's innovation theory on business cycles is examined through this dynamism. Trunk innovation is defined as that which plays a decisive role in building infrastructures and inducing subsequent innovations. Every innovation has its own technological development period just before the innovation diffusion. The emergence of new markets can be estimated by chasing the ongoing technologies.

Keywords: Nonlinear dynamism of innovation – Infrastructure – Business cycles – Bubble and depression – Technology foresight

JEL Classification: E32, L16, O11, O14, O30

1 Introduction

This paper is based on awareness of the nonlinear nature of innovation and elucidates the dynamism of innovation as the origin of the economic development, focusing on Kondratiev business cycles.

The correlation of economic development with technological innovation has not been explicitly clarified with decisive evidence since Schumpeter. This paper throws light on this issue by using the novel concept of the nonlinearity of innovation. First of all, it is clarified that diffusion of innovation is a physical phenomenon with a

---

* Paper presented at the 9th Conference of the International J.A. Schumpeter Society, Gainesville, Florida, USA.
Previous affiliation was Ryutsu Kagaku University, Faculty of Information Science, Kobe, Japan.

definite diffusion coefficient penetrating into a market towards maturation. The locus of the diffusion directly corresponds to the transition of increasing value-added produced by innovation and is used as a measure of the contribution of innovation to the economy. By this means it is possible to examine the innovation theory of business cycles. The bubble economy and recession that follows are discussed on the basis of innovation dynamism and an interpretation is offered. It is interesting to watch the synchronizing behavior of business cycles and role of late comers.

The innovation that plays a decisive role in the formation of economic infrastructures is specifically named "trunk innovation," and includes resources such as coal and oil. This category is closely related to the notion of the techno-economic paradigm and to the institutional change of society. Thus, trunk innovations more strictly exhibit a clear-cut correlation between innovation and economic development and induce subsequent innovation. As an example of such induction effect, the evolution of retail businesses is dealt with.

An innovation paradigm is composed of a technological development period and a market diffusion period. That is, before the diffusion of innovation products to the market, there is a long latent period of technological development. This correlation makes it possible to estimate the emergence of new innovation markets, a kind of technology foresight. From the analysis of ongoing science and technologies, emerging industries are estimated.

This paper is prepared on the basis of data and discussion of our previous works: Hirooka (1992, 1994a,b, 1999, 2000 , 2002, 2003a,b), Hirooka and Hagiwara (1992).

2 Logistic dynamism of innovation

Since the first Industrial Revolution, the economy has developed through innovations creating economic infrastructures. The diffusion of innovation is described by a logistic equation, as pointed out by Griliches (1957) and many others, e.g., Mansfield (1961, 1963, 1969), Metcalfe (1970), Fisher and Pry (1971), Nakićenović and Grübler (1991), Modis (1992), Marchetti (1997, 1988, 1995, 1996), who have all confirmed this relationship. Some authors, e.g. David (1975), Davies (1979), Metcalfe (1981, 1984), and Stoneman (1984), however, have proposed alternative or modified models for the diffusion of innovation products. Actually, the diffusion of new products in the market is quite often retarded by various economic turbulences, such as recessions, wars and so on. Thus, it makes it rather difficult to evaluate which equation is valid. The author, however, has found that the diffusion of products proceeds according to a logistic equation in a sound economy, but is disturbed by economic turbulence. This is clarified when diffusion is analyzed according to the Fisher Pry plot.

The logistic equation is expressed by (1):

$$dy/dt = ay(y_o - y) \tag{1}$$

where y is product demand at time t, y_o is the ultimate market size, and a is constant.

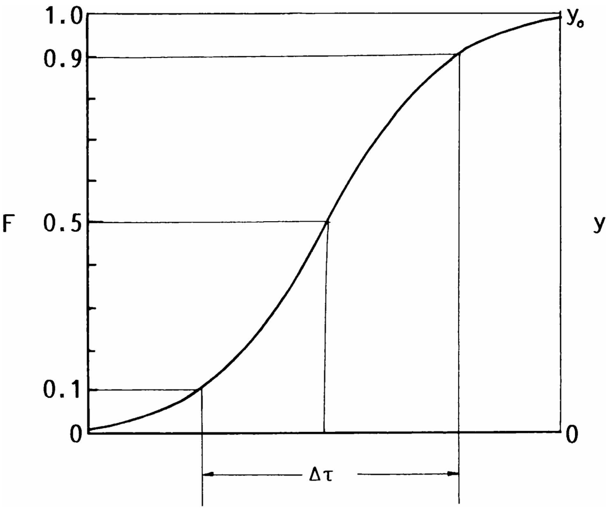


Fig. 1. Logistic equation and time span $\triangle\tau$

The solution of this nonlinear differential equation is (2):

$$y = y_o/[1 + C\exp(-ay_ot)] \tag{2}$$

If the logistic equation is expressed by the fraction $F = y/y_o$, the Equations (1), (2) are represented by (3) and (4):

$$dF/dt = \alpha F(1 - F) \tag{3}$$

$$F = 1/[1 + C\exp(-\alpha t)] \tag{4}$$

This equation is transformed by Fisher and Pry (1971) to form a linear relation over time t, which is expressed by Equation (5):

$$\ln F/(1 - F) = \alpha t - b \tag{5}$$

The ultimate market size y_o is determined by the flex point of the logistic curve, $y_o/2$, which is the secondary differential of (1), and the adaptability of the logistic equation is examined by the linearity of the Fisher Pry plot. The α is the diffusion coefficient of the product to market. If the time span between $F = 0.1$ and $F = 0.9$, $\triangle\tau$ is conveniently taken to express the spread of the logistic curve, this is a conventional expression of the time dependence of the product diffusion to market, as shown in Figure 1. This kind of treatment was also offered by Marchetti (1979, 1988).

The diffusion process was examined by the above procedure for 17 products, including five bulk chemicals, four engineering plastics, six electric appliances, crude steel, and automobiles. From the determination of the flex point, a linear correlation of the Fisher Pry plot was examined, as shown in Figures 2–5. The determination of the diffusion coefficient, α, was carried out, and the results are shown in Table 1. The data are for the Japanese market, except that of ethylene for the USA. These results clearly indicate that:

Table 1. Diffusion coefficients of innovation products[a]

Product	Diffusion coefficient α	Product	Diffusion coefficient α
Chemicals		Crude steel	0.28
Ethylene	0.39	Automobile	0.32
Polypropylene	0.49	Electric appliances	
Polyvinyl chloride	0.23	Refrigerator	0.65
Polystylene	0.37	Color TV	0.82
Nyron resin	0.24	Microwave oven	0.67
Polyacetal	0.27	VTR	0.73
Polycarbonate	0.24	Word processor	0.94
PPE	0.35	Facsimile	0.51

[a] Japanese market.

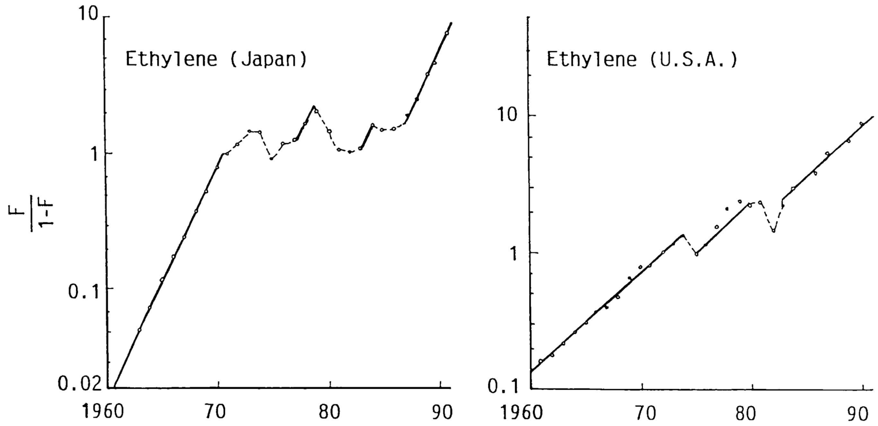


Fig. 2. Diffusion trajectories of ethylene (petrochemicals)

(1) The diffusion of new products obeys a simple logistic equation during sound economic conditions.
(2) The diffusion is easily disturbed by economic turbulences, such as recessions and wars, and sometimes the demand for products in periods of turbulence is greatly decreased and so diverges from the locus of the logistic equation.
(3) It is noteworthy that after the recession, the diffusion of the product resumes and takes the same slope of the logistic curve as before the recession. This strongly supports the fact that the diffusion of product has its own inherent trajectory with a definite diffusion coefficient.

The diffusion coefficients of products are classified in two groups, as pointed out by Davies (1979). Chemical products, crude steel, and automobiles are slow diffusing products, while electric appliances belong to the category of fast diffusing products. Electric appliances are consumer products the applications of which are recognized at a glance by the user, but future applications of chemicals are not easily determined by potential users. It should be noted that the growth rate of

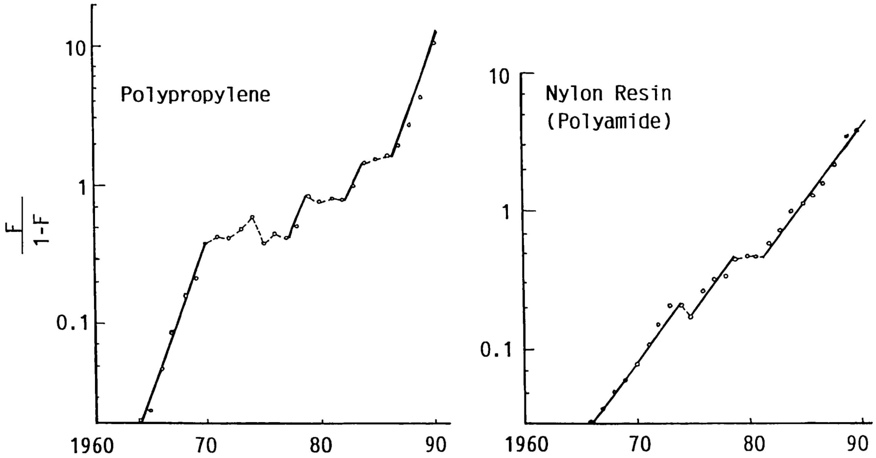


Fig. 3. Diffusion trajectories of plastics

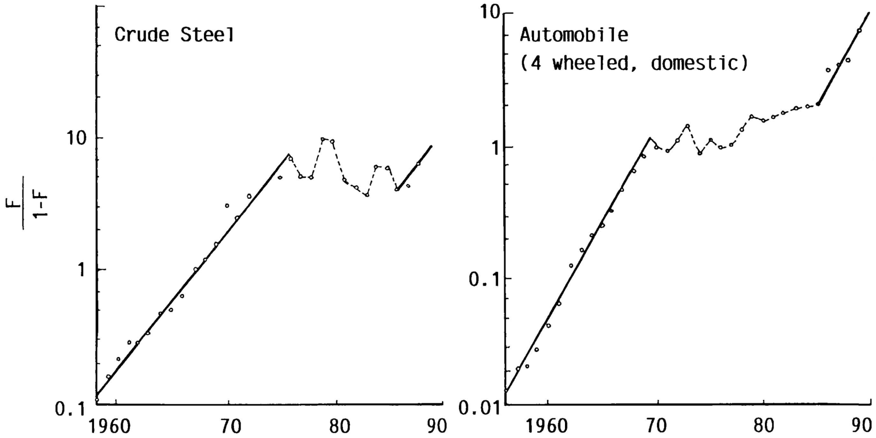


Fig. 4. Diffusion trajectories of crude steel and automobile

ethylene, being the main product of petrochemicals, is slower in the United States than in Japan. The Japanese petrochemical industry has lately started to introduce U.S. technologies so as to more easily find applications. This is because the United States discovered these applications early, which gave Japanese firms a follower's advantage. Another alternative explanation for such differences in diffusion rates could be a different susceptibility of the market: the Japanese could have the skill to find applications through quicker response.

Though there were recession delays of diffusion in both countries in the oil crises of 1973 and 1979, the delay in the United States was very short and the economy quickly recovered. This could be interpreted as meaning that Japan strongly depended on imported petroleum (naphtha) but the United States was able to use domestic natural gas. The Fisher Pry plot illustrates this kind of situation.

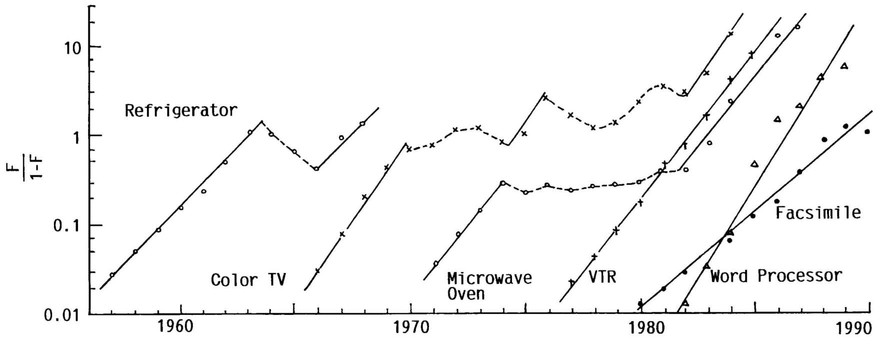


Fig. 5. Diffusion trajectories of electric appliances

3 Economic growth through innovations

3.1 Kondratiev business cycles and innovation clusters

J.A. Schumpeter (1912, 1934) ascribed economic development to technological innovations and pointed out that innovation is the only factor to raise the value-added of the economy. Schumpeter (1939) tried to reconstruct his theory of economic development through the concept of the Kondratiev cycle and ascribed the cause of the business cycles to the technological innovation. His explanation suggests that innovation takes place intermittently to form clusters and to stimulate the economy. He identified the first Kondratiev wave as having taken place by the time of the first Industrial Revolution, the second wave as having been bought about by steam locomotives and iron, the third wave as having been formed by the advent of electric power, chemicals, and automobiles, and the fourth wave as emerging in 1953. This conjecture was refuted by the empirical data, notably the fact that, while the peak was estimated in 1920, actual economic growth continued for more than ten years. Hansen (1941) amended the dating of the peak of the third wave from 1920 to 1929 to justify Schumpeter's hypothesis. The innovation theory for the cycle phenomenon, however, was not readily accepted, and other theories were postulated. There were alternative explanations for long business cycles which suggested that there is a cycle of scarcity and abundance under the following headings: capital accumulation theory (Mandel, 1972; Forester, 1979; Wallerstein, 1979; Gordon, 1978); labor theory (Freeman, 1977, 1982); relative price theory (Rostow, 1978); and war induction theory (Dickinson, 1940; Goldstein, 1971; Modelski, 1932). These theories have been widely summarized (e.g., in Freeman, 1983).

Mensch (1975) tried to clarify the correlation between innovation and economic development on the basis of the Kuznets business cycle. His important observation was that the emergence of technological innovations is not a continuous phenomenon but actually involves discontinuous clusters of innovations. He distinguished between inventions and innovations and defined the lead time as a time from a basic invention to the practical application. Its reciprocal value was adopted as the speed of the transfer processes from invention to innovation. He determined various speeds of innovations, and suggested that they emerge as a group in the economic depression stage of the business cycle, and interpreted this phenomenon as the "depression trigger effect": the depression exerts pressure for survival and so innovation is attempted as a means of escape. This argument is at least partially based on the fact that the speed of innovation is accelerated in the stage of depression.

Marchetti (1979, 1988) reprocessed Mensch's data using the idea that inventions and innovations are cultural pulses and as such should be analyzed using the population dynamics of Darwinian origin, coded into Volterra-Lotka equations. Thus, he successfully demonstrated that the Fisher Pry plot of cumulative numbers of inventions and innovations creates straight lines, indicating that these phenomena are expressed by logistic equations over time. Marchetti also examined the diffusion of many innovations in testing the fitness of the logistic equation under Fisher Pry notation, and found proof for his conjecture. He also compared these results with the position of the Kondratiev business cycle, but did not always obtain so clear a correlation.

On the other hand, Freeman, Clark and Soete (1982) summarized innovation systems to elucidate the Kondratiev cycles. According to their concept of "new technology systems", basic inventions emerge in the previous Kondratiev cycle. Thus, innovations create the recovery and boom stage of the cycle and reach stagnation followed by depression on the downswing. Freeman et al. strongly questioned Mensch's "depression trigger hypothesis," for two basic reasons. First, they questioned the data Mensch used, suggesting the samples were too ad hoc to present any adequate coverage of the major industrial fields involved in the long wave booms. They noted also that there is no adequate definition of basic innovation. Secondly, they doubted the plausibility of the depression trigger hypothesis to explain the cluster of innovations. They argued that R&D activity decreases in a depression, and their survey suggests that patent clusters exist during a long wave boom as well as in recessions. They argue that the clustering is of diffusion processes for these innovations but not the clustering of the innovations themselves.

Rosenberg and Frischtak (1984) discussed clusters of innovations, distinguishing between T-cluster and M-cluster. T-clusters are technically related, while M-clusters represent clusters of common stimulus of a generalized increase in demand, or other favorable macroeconomic conditions. Kleinknecht (1984) further examined the clustering phenomenon of innovations and empirically proved the idea.

Van Duijn (1983) thoroughly studied Mensch's treatment and recognized the importance of the time lag between inventions of technologies and their diffusion to market, examining the development course of the modern industrialized society. Especially, he took note that clusters of key technological innovations built new industries and became the trigger for long business cycles. He also tried to express the life cycles of new industries by S-shaped curves. New industries mature and the demand eventually levels off.

Those who ascribed the cause of Kondratiev business cycles to other factors than technological innovation also recognized the existence of industrial development during economic growth and the importance of technologies. For example, Rostow (1978) noticed that there were four clusters of technological innovations at the back of the economy: 1) spinning machines, iron making process using coke and steam engines, 2) railways, 3) steel making, 4) internal combustion engines, new chemicals and electric power. Mandel (1972) also reckoned four ages of 1) steam engines in handcraft manufacturing, 2) steam engines in the machinery industry, 3) electric power, internal combustion engines, and 4) electric and electronic appliances.

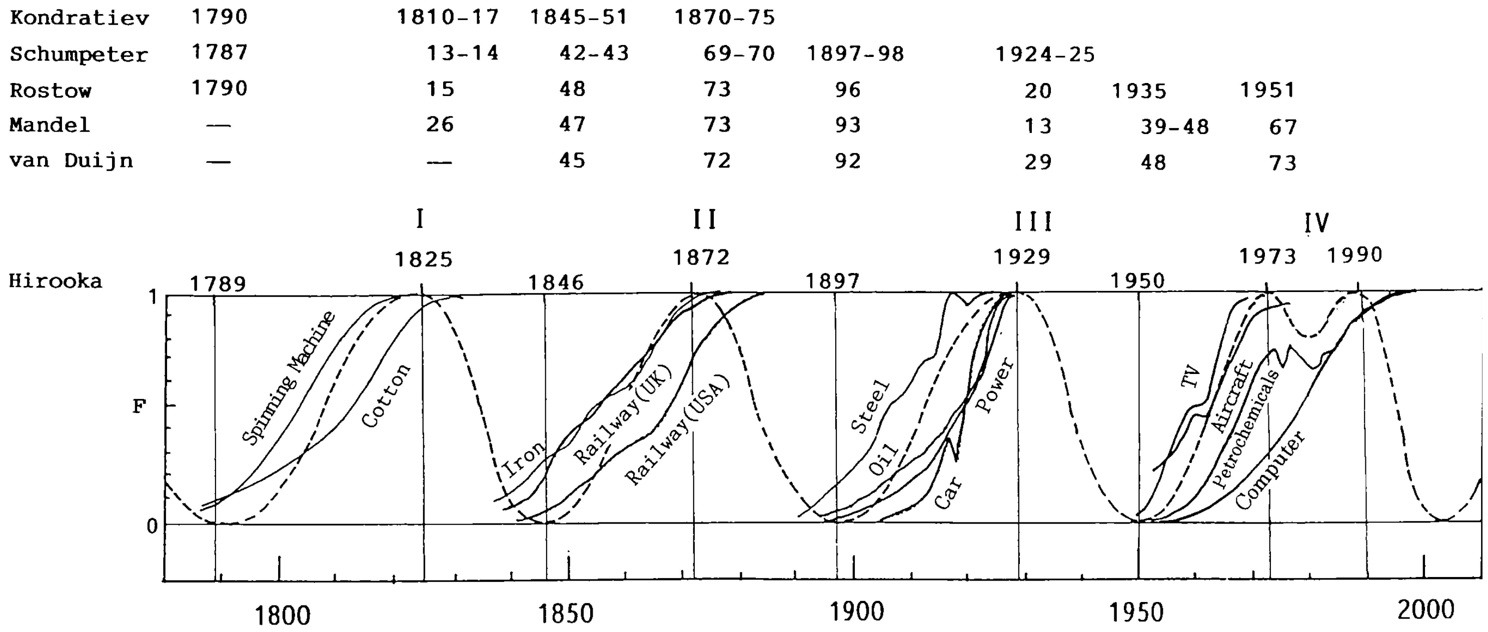
3.2 Determination of correlation between technological innovations and economic growth

The studies described above recognize the contribution of technological innovations within economic development and many studies have successfully expressed the diffusion of innovation products through a logistic equation, and discussed the relationship to business cycles. No one, however, has succeeded in getting direct evidence of the causal relationship between technological innovations and economic development as pointed out by Schumpeter. One of the reasons for this must be in the appropriate determination of the cycle. In fact, the Kuznets cycles used by Mensch deviate from the others, and there are various fluctuations within the cycles. The author carefully examined the position of the Kondratiev business cycles and obtained the most probable positioning as shown in Figure 6 .

The Kondratiev cycle was formulated by the determination of the degree of economic development by the use of GNP and other economic aggregates. On the other hand, technological innovations enhance value-added to the economy, and so the diffusion of innovation products corresponds to the progress of enhanced value-added by the creation of new markets. Therefore, the diffusion curve of innovation products directly reflects the degree of increasing value-added by the innovation per se. More important is that the diffusion curve has a nonlinear S-shape, which means that the diffusion tends to saturate within a definite time span. This trend makes it possible to locate the relevant innovation. Thus, the author tries to determine where the diffusion curves of innovation products are located on the Kondratiev waves. This is the central concept of this study. The locus of the diffusion of various technological innovations is depicted on the map of Kondratiev waves. The locus of innovation diffusion is normalized for the saturated market to unity. These innovations are those playing a crucial role in the formation of infrastructures in industrialized society. The results are illustrated in Figure 6: all of them cluster selectively at the upswing of the Kondratiev cycle.

The upswing of the first cycle is that of the first Industrial Revolution in the UK when the textile industry took off. The economic development there can be interpreted as the result of the development of the textile industry, which can be seen through the diffusion of mule spinning machines and cotton consumption along the upswing. The second cycle was caused by the development of railways and iron production. For the duration of the second upswing – from 1846 to 1872 – the extension of railway and iron production expanded more than ten times and the economy was stimulated by these innovations.

The diffusion trajectories at the upswing of the third cycle – between 1900 and 1929 – are those of the United States, reflecting the fact that the economic leadership shifted from the UK to the USA. Over this period the United States jumped from a developing country to an industrialized country by introducing



Kondratiev	1790	1810-17	1845-51	1870-75				
Schumpeter	1787	13-14	42-43	69-70	1897-98	1924-25		
Rostow	1790	15	48	73	96	20	1935	1951
Mandel	—	26	47	73	93	13	39-48	67
van Duijn	—	—	45	72	92	29	48	73

Fig. 6. Kondratiev's business cycles and diffusion of innovations; — Diffusion of innovation products, — — — Kondratiev's business cycles

various innovations, such as technologies in steel making, oil drilling, automobile making, and electric generating processes from Europe. The diffusion of these innovations led to the construction of the advanced state. After World War II, the United States again enjoyed the most prosperous era, accounting for one third of world GNP, by developing various innovations such as aircraft, electric appliances, and petrochemical products such as synthetic plastics, fibers, and rubbers. Japan jumped from a developing country to an industrialized one at this upswing with a substantial industrialization focusing on crude steel, automobile, petrochemicals, and electric appliances. This was quite similar to the United States at the beginning of the 20th century.

The fourth wave was heavily affected by oil crises but recovery took place in the latter half of 1980's based on the widespread diffusion of information and communication technology and other high-tech innovations.

These results clearly indicate that the diffusion timing of various innovations always gathered at the upswings of the Kondratiev business cycles. This strongly supports Schumpeter's hypothesis that innovation is the engine for the development of the economy. This causal relationship is derived from the fact that the contribution of innovations is directed by the location of diffusion curves along the upswings because that the diffusion curve itself represents the locus of increasing value-added created by new innovation markets.

3.3 Synchronization of world business cycles

The Industrial Revolution took place regionally in England, specifically Lancashire in the 1770's, and the United Kingdom continued to hold leadership across two business cycles. The big surge of the Industrial Revolution spread over Germany, France and the United States. The initiative at the third wave was taken by the United States, as was the fourth. Japan was a latecomer, joining the modern industrialization with the Meiji Restoration in 1868 after a long closed-door policy; this was 100 years behind the Revolution in England. How did Japan chase Western countries in terms of industrialization? It is interesting to follow this trend through the Kondratiev long cycles. Figure 7 illustrates the correlations among the United Kingdom, United States, and Japan on the same map of Kondratiev waves. At the second wave, the United States was a little late in railway construction and iron production compared with England. The behavior of the United States at the third wave was incredible in taking the leadership in all innovation products such as steel, electricity, and automobiles. The production of automobiles in England started in the same cycle but the actual market was not completed until the fourth cycle after World War II. Japan started her industrialization with the introduction of western technologies at the third wave, the first industrialization having been undertaken through cotton production and railways. The industrialization of cotton was carried out two waves later than England, while railway construction was one wave behind. Japan also commercialized to produce steel in the Meiji Era, but the maximum production was less than 7 million tons in 1940 and collapsed to 550 thousands tons in 1946 due to the War. The actual industrialization of Japan was achieved in the high growth period of the 1960's with the production of steel, automobiles, textiles, electricity,

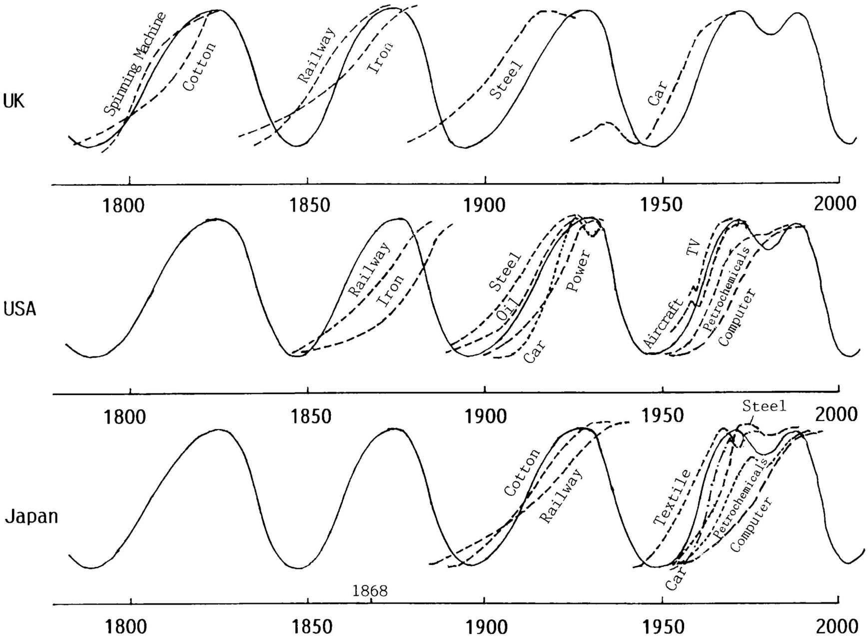


Fig. 7. Synchronization of business cycles in the world economy

and petrochemicals, and the catch-up to western countries finished around 1970. These events are illustrated in Figure 7, where it is interesting to know that the diffusion of innovation products in every country gathered at the upswings of the Kondratiev waves. This means that the business cycles have been synchronized all over the world and the main diffusion of products concentrated at the upswings, with latecomers industrializing at the upswing of later cycles. This also implies that the diffusion of innovation products is retarded over the recession and resumes at the next upswing. This trend is already observed in the diffusion of innovation products in Japan, as described in Figures 2 to 5. These correlations suggest that innovations induce economic growth, and economic prosperity provides a comfortable field to diffuse. This could be a reason of the gathering of innovations to form a cluster at the upswing. This could be a kind of chicken-or-egg question.

4 Turnover of bubble economy and depression

4.1 Discussion on the Great Depression

The Great Depression has been discussed by many economists but there has been no decisive explanation. The predominant cause has been ascribed to the failure of monetary policy and the escalation of speculative activity, especially the excessive fluidity of the money supply caused by speculation and the bubble economy. On the other hand, Schumpeter (1939) discussed the business cycle and ascribed the Great Depression to overlapping down-trends of the Kondratiev, Jugular, and

Kitchin cycles, while he attempted to explain each cycle as having been caused by innovation. Kuznets (1940) strongly criticized Schumpeter's complex theory of innovation cycles.

Among these discussions, the common characteristics of the Great Depression are as follows:

1) The depression emerged in the United States, the most prosperous country in the world.
2) The depression took place at the peak of prosperity.
3) The depression continued for a long time and affected various markets, such as stock market, manufacturing industries, and agriculture.

4.2 Industrialization and the Great Depression in the United States

The United States introduced various technologies from Europe in the latter half of the 19th century and jumped from a developing country to an industrialized country in the early 20th century. The United Kingdom enjoyed the second surge of the Industrial Revolution through the diffusion of railways and iron production. Later, the United States developed in such a way that the railways reached to the western coast and steamships actively made trade with Europe.

The United States introduced Bessemer furnace technology and steel production increased rapidly at the end of the 19th century. Hydraulic power generation was launched at Niagara Falls with the cooperation of Siemens of Germany in 1895. Oil began to be used for lighting in the form of kerosene in the middle of the 19th century, and production increased after large oil fields were found in California. Due to the development of the automobile, the demand for gasoline expanded and the oil industry developed. Though the motorcar was invented by Daimler and Benz in Germany, the United States took the leadership role in the car industry after Henry Ford incorporated the production line system. These innovations spread in the United States in the early 20th century, as shown in Figure 8a. At the same time, radio and various consumer durables such as washing machines and refrigerators came into wide use.

The Ford process drastically enhanced the productivity of automobile production and the price of the Model-T car of Ford Motor Company was decreased from $850 in 1908 to $250 in 1924. Electric power also improved the productivity of manufacturing industries and was supplied for domestic use as well. The trade with Europe was assisted by the diffusion of steam ships, with transport costs dropping to one tenth their former levels. These innovations drastically improved the productivity of industries and led to tremendous economic development. With this boom, stock prices skyrocketed. The stock price of General Motors rose from $21 in 1925 to $46 in 1928.

The rapid development of the economy brought about a great deal of investment demand, which continued for more than ten years. The rise of stock prices stimulated the trend to raise money through the stock market, and the leverage effect of such capital gains further accelerated investment. In 1928, dealings on credit abnormally increased and market prices rose heavily. In 1929, the stock prices skyrocketed, the rise for three months in the summer corresponding to the amount of the yearly rise in 1928. On 3 September, the stock price reached a peak of $469 on the average for 25 industrial companies and $381 for the Dow Jones DI industrial index and never recovered. On Black Thursday, 24 October, a sacrifice sale was carried out, heavily decreasing the stock price to $272 in the Dow Jones. The stock price decreased continuously to $41 in July, 1932 which was one tenth of the peak price.

4.3 The background and mechanism of the Great Depression

At the beginning of the 20th century, the United States was industrialized and the infrastructures of a modern economy had been laid through various innovations. These innovations activated the economy and GDP was drastically enhanced. Such intensification of economic development through innovation raised stock prices and induced a bubble economy. The diffusion of innovations, however, reached maturity. The production of automobiles leveled off at 3.7 million cars: 3.62 million in 1923, 3.73 in 1925, and 3.69 in 1926, respectively. The production of crude steel was 49.0 million tons in 1923, 49.7 in 1925, and 49.3 in 1927, respectively. The consumption of oil and electric power still increased throughout the bubble economy, but after the crash the demand dropped drastically. The demand for cars in 1929, however, jumped to 4.45 million, up 18% compared with the previous year, and the demand for crude steel also increased abnormally to 61.7 million tons. These increases were for an imaginary market which proliferated like a cancer cell. Such proliferation reached more than 20% over the normal level, as shown in Figure 8a. The market, however, shrank drastically (as shown), with car production being 2.78 million (down 38%), crude steel being 44.6 million (down 38%), with the level before the bubble only recovering ten years later. The crash can be interpreted to have been caused by the cognition gap between the reality of a nonlinear phenomenon of the market reaching maturation and the false imagination of the linear development of economic growth.

4.4 Economic development and the crash of the bubble in Japan

Every country has a chance to advance from the developing stage by introducing various innovations and can enjoy such high annual growth rate of the economy as more than 10% only once. The United States achieved this during the upswing of the third Kondratiev cycle, and Japan encountered the same performance at the upswing of the fourth cycle in the 1960's. The latter case is quite similar to the previous one in the United States: the Japanese economy grew by introducing various innovations from the United States and Europe, with an annual growth rate of more than 10% of in the 1960's. Economic infrastructures were set up through the diffusion of crude steel, automobiles, electric appliances, and petrochemicals, as shown in Figure 8b. At the beginning of the 1970's, the Japanese economy became mature but overheated. Stock prices rose from 3,000 Yen in March, 1972 to 5,000 Yen at the end of that year, and capital gains were enhanced. Japanese industries

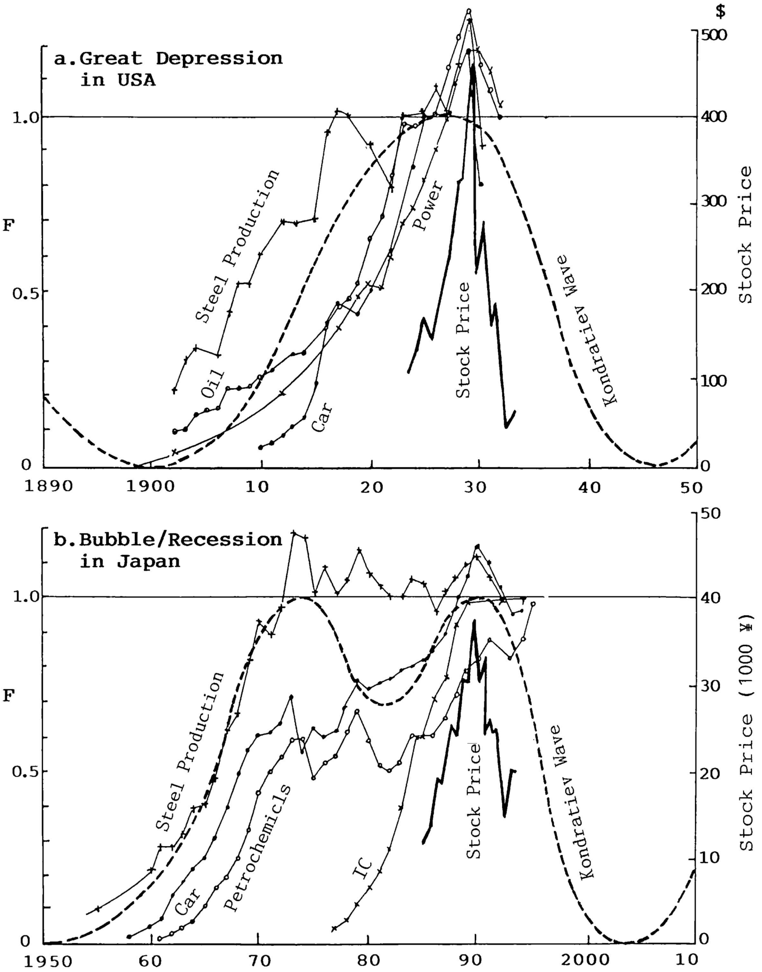


Fig. 8. Diffusion of innovation and bubble / depression turnover

began to export their products to western countries riding on the stream of globally expanding trade.

The worldwide prosperity of the 1960's collapsed with the oil crisis in the 1970's but recovered in the middle of the 1980's. After the Plaza Accord in 1985, the Japanese Yen was highly over-valued and the Japanese government took steps to stimulate domestic demand to cope with an increasingly difficult export business. Low interest rates and an abundant fluidity of money induced a bubble economy, with the stock market and land prices growing feverishly. This trend was quite similar to the state before the Great Depression in the United States in 1929, as shown in Figure 8.

The high economic growth in the United States at the beginning of the 20th century was caused by the diffusion of steel, petroleum, electric power, and automobiles, and at the top of the upswing the collapse occurred. The Japanese economy also took the same route, that is, the high growth of the economy was accomplished during the 1960's through the diffusion of crude steel, automobiles, electric appliances, and petrochemicals. Beyond the oil crisis, the Japanese economy again enjoyed growth through the contribution of electronics and information technology, and then entered a maturity period. Thus, the interval from the 1960's to the 1980's may be seen as the upswing of the Japanese economy in terms of innovation. In the latter half of the 1980's, many innovations reached maturity, with only electronics and petrochemicals had developed further. Thus, there was little room to absorb the flood of money, except through the stock market and land prices. The Japanese national bank, however, issued plenty of money to stimulate domestic demand, but succeeded only in raising the prices of stocks and land, including American real estate. Most firms raised money from equity financing by dealing in stocks and land and the boom escalated.

As seen in Figure 8b, the production of crude steel increased during the 1960's, reaching a mature market state of 100 million tons by 1973. Automobile production also grew to maturity in the domestic market in the early 1970's. The development of the petrochemical industry was interrupted by the oil shock, but resumed in the latter half of the 1980's. These data clearly indicate that, before the crash of the bubble economy in 1990, the market for various innovations had become mature and the huge demand in the bubble economy was imaginary.

4.5 Historical examination of turnover of bubble and depression

The above analysis indicates that common phenomena must take place along the Kondratiev cycles. We have discussed the turnover of the bubble and depression phases on the third and fourth cycles. Now let us check the previous cycles.

The first Industrial Revolution took place in the Lancashire District, UK beginning with the development of the cotton industry, but several innovations, such as steam engines, the iron making process (the coke blast furnace), and machine tools combined to make complex systems and give a comprehensive synergy effect to the stimulation of the economy. As shown in Figure 6, the first upswing was may be taken to have arisen from the rapid increase in cotton demand. The stock market for cotton gradually heated in the middle of the first upswing from 1789 to 1825. Trade with India and America was fostered by newly commissioned steamships, with the result that their stock prices increased. In 1822, economic prosperity progressed and speculation in the stock market accelerated. Stock prices skyrocketed from 1824 to 1825, but at the end of 1825, the bubble burst. Stock prices dropped to reach one fourth their value in four years. The depression continued for 20 years, until 1846.

The second Kondratiev wave was activated by the railways and iron production, and the upswing extended from 1846 to 1872. The first locomotive ran between Stockton and Darlington in 1825. The actual diffusion, however, began in 1845 after the long recession. There were many obstacles to the diffusion of railways and, in addition, the recession retarded the progress. However, the success of the first lines and a seal of Royal approval eventually swayed public opinion. Railway mania was ignited in 1845, with the extension of railways reaching 18,000 km in 1855 and

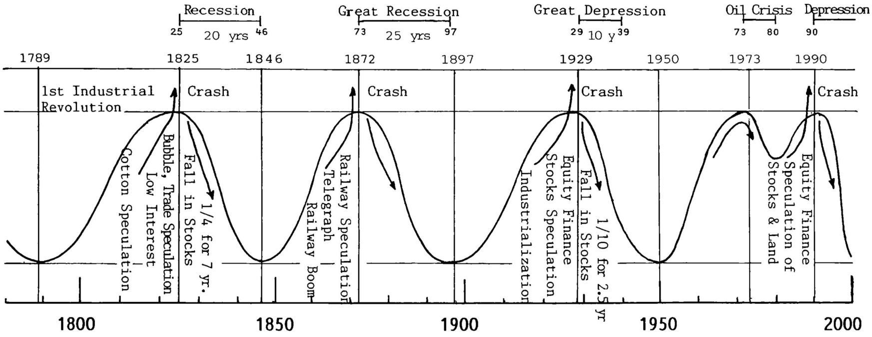


Fig. 9. Kondratiev business cycles and bubble / depression turnover

26,800 km in 1875. In the United States, railway construction began in 1830 but substantially took off after 1850. With the switch of the cotton trade from India to America, the construction of canals and railways escalated in the United States. In the latter half of the upswing of the second cycle, until 1872, speculation on railways overheated and the bubble collapsed in 1873. A great recession followed in both the UK and the United States, and continued for 25 years, until 1897. Unemployment in the UK reached 12.5% in 1879. We can see the same pattern as of the Great Depression in 1929 in the United States.

4.6 Innovation dynamism and business cycles

As we can see from the above studies, every wave has an upswing with clustered innovations, and the economy develops through value-added brought about by innovations. In the middle of an upswing, economic development is boosted and speculation in the stock market takes place to form a bubble economy. At the top of the upswing, the bubble collapses, bringing on a long recession. All of these figures are illustrated in Figure 9.

We may conclude that the modern industrialized society is built upon various technological innovations, and crashes and depressions, which subsequently take place, may be explained as the overheated economy induced through innovation diffusion. Before the innovation age, there were many cases of speculations causing disturbances to the economy, e.g., tulip speculation in Holland in 1636 and the stock speculation of South Sea Company early in the 18th century. Such speculation shortly ceased and the influence on the economy was not severe. The above discussed subjects induced by innovation diffusion, however, are quite different from such simple speculation; the influence was severe and long because the phenomenon induced by innovation dynamism was accompanied by a socioeconomic structural change which operated for a long time. Such a huge change in the techno-economic paradigm and institutional societal structure is not so easy to recover from.

Since the first Industrial Revolution, modern industrial society was built upon various technological innovations for 230 years and only innovation can enhance the value-added of the economy, as pointed out by Schumpeter. The GDP of advanced countries was drastically enhanced by technological innovations more than 100 times over the past 200 years, and 30 times over the last 100 years. This is important evidence that innovation plays a crucial role for economic growth. As evidenced by the fact that the economic growth rate exceeds 10% a year when a country changes from the developing country to an advanced one through innovation, the impact of innovation diffusion is extremely large and thus the effect of a bubble and crash is so severe that it takes a long time for recovery.

The identification of the Kondratiev cycles is not so easy because there are so often turbulences within a cycle. In the fourth wave, the oil crisis heavily affected the economic situation, deforming the shape of the cycle. If we focus on innovation cycles, however, the diffusion of innovation products can be easily identified even across short recessive retardations. A conclusion of this paper is that the behavior of innovation in the course of economic development can give us a more clear understanding of the techno-economic paradigm and societal institutional change. The innovation promotes economic development, but the diffusion of innovation is also affected by economic conditions. This suggests that innovations diffuse at the upswing to make a cluster and may be a cause of the clustering of innovations. This phenomenon is suggestive of Schmookler's theory of scissors for technology push and demand pull.

5 Network of infrastructures and business cycles

5.1 Definition of trunk innovations and formation of economic infrastructure networks – energy, transportation, and information and communications

Schumpeter, in his paper "The Theory of Economic Development", defined five groups of innovation: those consisting of (1) new products (goods), (2) new processes (production processes), (3) new markets, (4) new sources of supply of raw materials, and (5) new organization. Our discussion has concentrated on technological innovations defined by (1) and (2). The author, however, notices the importance of new resources (4) and defines innovations governing infrastructure and its network in economic development as a trunk innovation which includes coal, oil and other feedstock as resources. These raw materials are not invented as technologies, but just feedstock for the economic development. Thus, the definition of a trunk innovation is that which have a pervasive impact on the economy, forming infrastructures and networks beyond industries such as energy, motive forces, transportation, and communications. These types of infrastructure and network enhance economic development, increasing value-added through the synergy effect. This term "trunk innovation" is not the same as "general purpose technologies" defined by Bresnahm and Tritenberg (1994) or generic technologies.

The history of modern industrialization is that of the establishment of economic infrastructures. The above-defined trunk innovations which are related to energy, transportation and communications have decisively contributed to the construction of the economic infrastructure over the past 230 years. In the previous sections, we discussed Kondratiev business cycles constructed by various innovations. Among them, there have been many innovations in energy, transportation and communications. We would like to add canals, coal, and oil to the list – these are not technological innovations but are crucial contributions to economic development. Thus, if modern industrialization is discussed in terms of trunk innovations, the contribution of innovation is more readily interpreted. These innovations enhance economic development by a kind of synergy effect, as they have a pervasive influence on the economy. Figure 10 exhibits the correlation between these trunk innovations and business cycles.

Another important activity of trunk innovation is the induction activity of subsequent innovations. Here we deal with such inductive innovation in the case of revolutions of retail businesses.

In the long history of the humanity, the first breakthrough in economic activation was made through sailing ships. After the Renaissance 500 years ago, innovations in ship building and navigation, e.g., the invention of the compass, made it possible to cross oceans. The age of Great Voyages came about in the 14th century, while in 1522, Magellan succeeded in circumnavigating the world. With this as a turning point, world trade developed, e.g., with cotton and spices being imported from India. This can be said to be the first innovation of the world infrastructure network.

The first Industrial Revolution is described as having been caused by the invention of spinning machines and steam engines. The success of this revolution was, however, strongly supported by the canal transportation network, constructed from 1760 to 1820, just before the upswing of the first Kondratiev wave. The first canal was built from Liverpool to Manchester in 1760. In the 1770's, Grand Truck between Birmingham and the Severn River, Grand Junction between London and Birmingham, and many others were constructed, and throughout the UK, waterways were connected. In Europe and the United States many waterways were also constructed to establish transportation networks in the same age.

Energy and motive power played an important role in the first age of innovation in stimulating invention. The cotton industry first used the water wheel as the driving force for spinning machines, and the first steam engine was set up in a factory in 1789. Since then, the cotton industry achieved a drastic development. A. Darby adopted coke for blast furnaces in 1735, using coal instead of charcoal. The adoption of steam engines for blast furnaces was another innovation for ironworks. These interactions related to energy (motive power) and transportation infrastructures were an important background and seem to be the origin of the synergy effect of innovations.

The stems of the second cycle of the Kondratiev wave were railways and iron production, with the upswing occurring between 1846 and 1872. During this period, steam ships replaced sailing ships and transportation costs decreased one tenth. This innovation drastically improved conditions for international trade. Along the railway lines, electric wires were laid for communication between stations. This facility was expanded for public communications as the telegraph, the first modern communication system. Coal consumption also drastically increased to support the production of railways and ironworks.

The third Kondratiev wave took place with the United States as the leading country, with the most important innovations occurring as the upswing extended

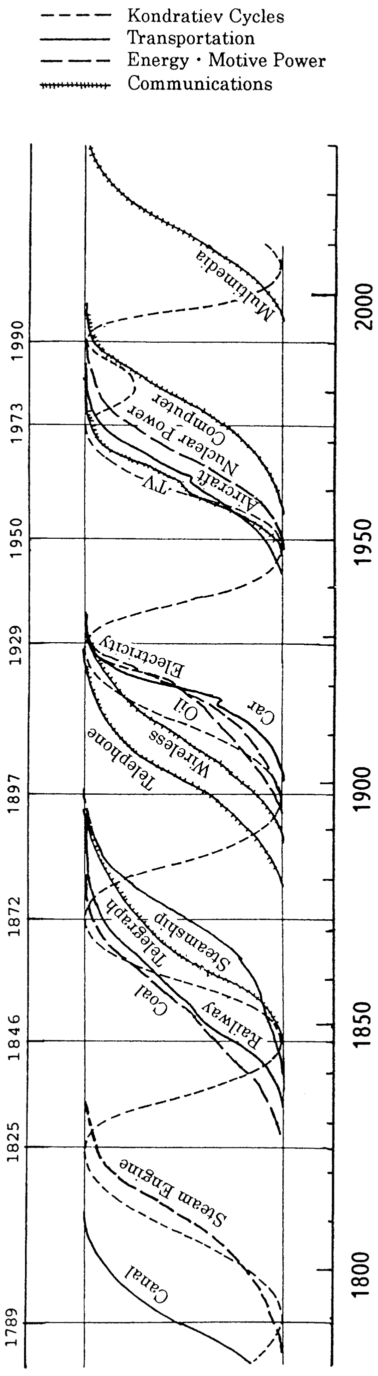


Fig. 10. Trunk innovations for infrastructures and business cycles

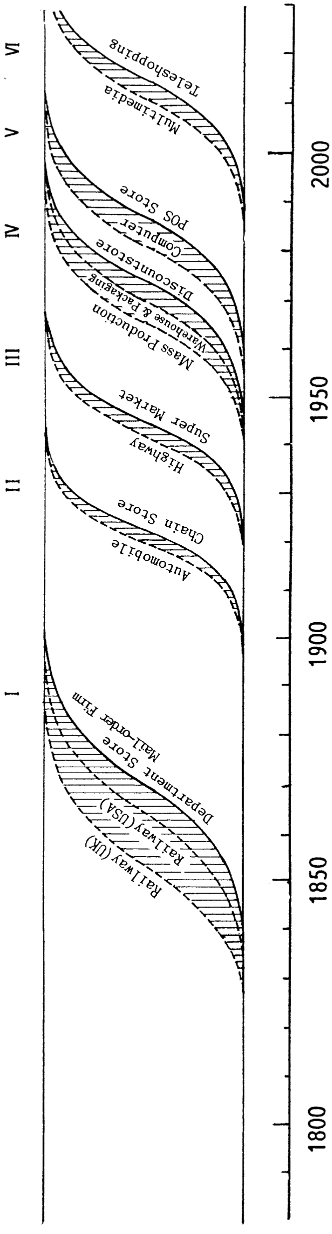


Fig. 11. Retail business evolving through trunk innovations

from 1900 to 1929. In this cluster, two important energies were introduced: oil and electric power which drastically enhanced industrial productivity and facilitated societal activities. Automobiles became practical and would become a strong competitor to the railway in the next generation. Innovation in communication technology included wireless radio technologies and the telephone. These innovations facilitated various communications and enhanced economic productivity.

We are now at the final stage of the fourth wave, the upswing located between 1950 and 1990. As for the energy resources sector, oil consumption has sharply increased and oil dependence has become predominant. Nuclear energy has been supplied as a transit to renewable clean energy, which is not yet established. Transportation has entered into a complex age, combining automobiles, trains, ships, and aircraft. Communication technologies have developed to create a so-called information society. Various information appliances have been developed, including televisions, tape recorders, facsimiles, video recorders, and so on. Most important are computers and semiconductor integrated circuits, which have had a huge impact in various fields. Industrial productivity has greatly improved through automation and the retail business has developed efficiencies through the use of POS systems and logistic innovation. Societal performance has been also improved by computerization. A multimedia revolution, however, has not been actually realized yet, and the introduction of Internet facilities suggests that the next upswing is about to take off.

5.2 Evolution of retail business and network of infrastructures

The impact of trunk innovations in the construction of economic infrastructures has been discussed above. As an example, the retail sector is introduced as having been induced by trunk innovations The evolution of retail business proceeded in stages: department stores, chain stores, supermarkets, discount stores, and convenience stores, in chronological order. These innovations took place just after the trunk innovations in infrastructure such as transportation and information technologies. These correlations are summarized in Figure 11.

Department stores originated with Bon Marché in Paris in 1852, with Au Printemps, Louvre, Samaritaine, and Lafayette emerging by 1890. Harrods and Marks & Spencer in London and many other department stores in Germany and Italy were built in the same period. In the United States, Macy's opened in New York in 1858, Wanamaker in Philadelphia in 1862, and Marshall Field in Chicago in 1868. The timing of these stores was along the upswing of the second Kondratiev cycle and just after the diffusion of railways in UK and in the United States, and as well coincided with the diffusion of steam ships. This correlation clearly indicates that the revolution in transportation induced the formation of department stores because the distribution cost became less, making it possible to collect various goods from a long distance. Along the railway lines, telegraph networks were constructed and the mail-order business became operable, allowing the development of companies such as Montgomery Ward in 1872 and Sears Roebuck in 1886.

Chain stores started to operate in the upswing of the third Kondratiev's cycle between 1900 and 1929. A typical chain store was the Great Atlantic and Pacific

Tea Company (A&P Co.), which was set up in 1825. At the beginning of 1900 A&P adopted a new business style called "economy store", chasing economies of scale and rationalization by integration. This kind of business was born through the changed social background strongly affected by the emergence of various infrastructures including automobiles, electric power, steel production and oil. The development of full-scale road systems allowed distribution systems to be effectively rearranged. Accordingly, urbanization was also promoted. Infrastructure development functioned to enhance the feasibility of chain stores. The actual development of chain stores closely followed the diffusion of automobiles. Chain stores, however, faced difficulties in the Great Depression.

Just as the crash began, a unique shop called "King Kallen Grocery" opened in 1930. This was the first supermarket, a big, self-service shop with a large parking area located in the suburbs. This sales method was quite in contrast to that of chain stores. The supermarket became popular and developed rapidly all over the United States after World War II, along with the extension of highways.

After World War II, the Kondratiev cycle entered the fourth stage, with the upswing occurring between 1950 and 1973. The most prosperous age started early in the 1960's as various innovation products flooded into markets. Various electric appliances, a variety of synthetic plastics, textiles and other durable consumer goods shaped huge markets. This was the age of mass production. "Discount stores" came to deal with various discount goods in large lots; the first, such as Kmart and Wal-Mart appeared in 1962. Kmart successfully expanded the business to more than 2,000 stores in the latter half of the 1970's. One of the feasibility factors for discount stores lay in the low warehouse costs achieved by introducing forklifts and pallets, so the labor cost was improved by 25%. The innovation of packaging through the use of synthetic plastics was another factor.

Information technologies drastically changed the retail business. The retail business adopted POS (point of sales) systems using barcodes, making possible stores with small sales floors, allowing goods for better sales to be selectively displayed, and improving logistics. Various retail businesses have also developed by introducing POS, EDI (Electronic Data Interchange) and various advanced information technologies. The earlier introduction of such information technologies resulted in better competitiveness. Wal-Mart, having only 200 stores, was far behind Kmart, which had more than 2,000, but began to introduce computers for inventory control in 1974. They decided to introduce POS systems in 1979, started experiments with EDI in 1981, expanded POS systems in 1983 and introduced satellite network systems in 1985. The POS system was completed over all stores in 1988. Wal-Mart completed automatic delivery systems in 1991. As a consequence, Wal-Mart exceeded Kmart in sales in 1990 and has developed smoothly since.

Seven-eleven Japan was set up in 1973, introducing the know-how of the American Seven-eleven, a division of Southland Co. Seven-eleven Japan introduced a computer ordering system and constructed diversified network systems for cooperative distribution in 1978. As a secondary stage facility, POS and EOB systems were introduced in 1982. Single article management and logistics systems were introduced by interactive POS and information management computers in 1985. Strategic logistic operations were controlled by introducing an ISDN network in

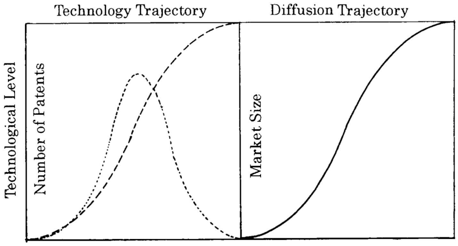


Fig. 12. Innovation paradigma – cascade of technology and diffusion

1991. These countermeasures using information technologies successfully provided Seven-Eleven Japan with a strong competitive edge.

These data clearly indicate that retail businesses have evolved through trunk innovations to construct network infrastructures.

6 Kondratiev 5th wave and estimation of the next generation

The above discussion dealt with diffusion phenomena of innovations.

Now, we would like to estimate the fifth wave of the Kondratiev business cycle. In order to estimate, we have to take a basis for extrapolation. Fortunately, we have been able to obtain an empirical relationship as shown in Figure 12. Before innovation products start to diffuse in the market, there is a long period of technological development. As discussed in Hirooka (2003a,b), the technological development period lasts for about 30 years, after which the diffusion of innovation products begins, reaching saturation in another 30 years. The locus of technological development also obeys a logistic equation. For example, taking the number of core inventions or basic patents during the development of a technology, the distribution of cumulative numbers can be expressed by a logistic equation. The trajectories of ongoing sciences and technologies can be easily recognized and we can estimate the future market by the extrapolation of the technology trajectories to the diffusion phase. The results indicate the timing of commercialization of these technologies.

Figure 13 exhibits results from such an extrapolation to estimate the fifth Kondratiev cycle using past empirical relationships. The upper figure indicates technology trajectories expressed by logistic curves and the lower figure is the actual diffusion trajectory. After the completion of the core technology, the diffusion of new products starts. This correlation holds with respect to past innovation paradigms, as shown in Figure 13.

As for the next generation of technologies, we choose several fields: advanced materials, biotechnology, and electronics. The first technology to take off is no doubt multimedia, through Internet systems. Next will be nano-catalysts, which will be commercialized in the near future. Materials controlled with nanometer size precision can then be commercialized. Superconductors will also be commercialized in the near future. Precision polymerization will offer next generation products having extremely high performance and superior functions. In biotechnology, genome engineering will take off, and regeneration engineering also offers very promising

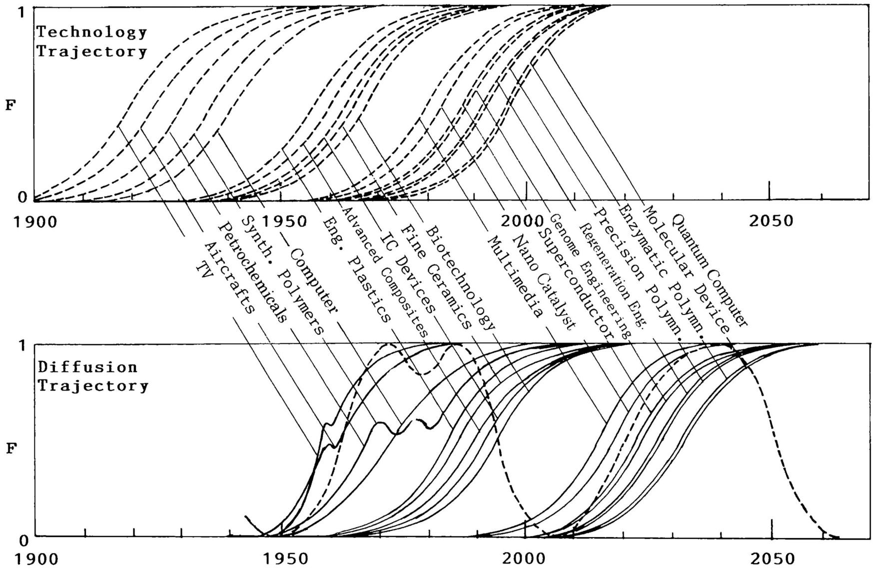


Fig. 13. Innovation clusters and 5th Kondratiev cycle

applications. These technologies have entered into the realm of venture business even now. In electronic technology, nano-devices and quantum computers can be expected. Single electron devices have already succeeded in making trial pieces using carbon nanotubes. Quantum computers have been already confirmed to be operable.

As a whole, these future technologies seem to make a cluster and this trend indicates that the upswing of the fifth cycle will take place between 2000 to 2040. We hope that this upswing will be large enough to sustain a prosperous economy because serious problems will be encountered, such as environmental issues and shortages of energy resources in not so far future.

7 Discussion and summary

This paper was presented at the 9th Conference of the International J.A. Schumpeter Society at Florida, USA on 28th March, 2002. Some papers and books have been published and referees have raised questions. This additional section is a reflection of such circumstances including comprehensive summary of this paper.

7.1 Characteristics of this paper

Eight characteristic results in this paper are summarized as follows:

(1) Inductive Analysis of Innovation: The whole analysis in this paper is inductively carried out on the basis of empirical data and fact finding.

(2) Physical Nature of Innovation Diffusion: Innovation diffusion is a kind of physical phenomenon having its own inherent diffusion coefficient.

(3) Nonlinear Nature of Innovation: The most important aspect of this paper lies in the recognition of the nonlinearity of innovation diffusion approaching a saturated level within a definite time span. This phenomenon makes it possible to locate the position of innovation diffusion.

(4) Causal Relation between Innovation and Economic Development: The diffusion of innovation is evidence that innovation gives value-added to the economy. The diffusion S-curve represents the progress of the value-added by innovation. Thus, the fact that the nonlinear diffusion S-curves of various innovations gather at the upswing of the Kondratiev wave directly indicates that the innovation cluster brings about economic development through enhancing value-added and proves Schumpeter's postulation on the innovation theory for the long business cycles.

(5) Wider Concept of Trunk Innovation: "Trunk innovation" is proposed to indicate a wider concept than technological innovation and is defined as indispensable innovations making economic infrastructures and networks and enhancing economic activity. Thus, energy feedstocks such as coal and oil are included as important ingredients. Trunk innovations are categorized into energy, resources, motive powers, transportation, and information & communications. Trunk innovations also induce many subsequent in novations, e.g., evolution of retail businesses.

(6) Synchronization of Business Cycles: The world economy is synchronized and innovations collectively diffuse at the upswing of the Kondratiev waves, making a cluster. Latecomers like Japan are industrialized at later upswings.

(7) Novel Interpretation of Great Depression: It is found that the bubble economy crashes at every peak of the Kondratiev wave and a long recession follows. The Great Depression is discussed on the basis of innovation diffusion and the crash of the bubble economy is interpreted as a phenomenon when people recognize the gap between the imaginary bubble economy and the actual saturation of the market. This is triggered by cognition of the gap between the nonlinear nature of innovation and a simple belief of the linear growth of the market. This analysis affords a new interpretation of the bubble economy and the Great Depression.

(8) Technology Foresight towards the fifth Wave: Before the diffusion of innovation products, there is a rather long period of technology development. We extrapolate trajectories of on-going science and technologies to estimate the next generation industries and suggest the occurrence of the Kondratiev fifth wave having a peak around 2040.

7.2 Discussion on recent references

Though after the presentation of this paper, we would like to make some comments on recent publications.

Freeman and Louçã (2001) and Louçã (2002) discuss the Kondratiev business cycle in terms of cliometrics. Perez (2002) produced a book entitled "Technological

Revolutions and Financial Capital – the Dynamics of Bubbles and Golden Ages". These books have quite stimulating contents about innovation dynamism and long business cycles. Both books describe technological innovations, focusing on the importance of development of techno-economic paradigm and institutional change.

Freeman and Louçã (2001) recognize four Kondratiev waves since the First Industrial Revolution and the fifth is ongoing. This book includes data regarding innovation diffusion and related economic trends which are informatively reflected in their discussion.

Perez (2002) deals with long surges of innovations to describe the S-shaped curve. In this aspect, her study is quite close to ours in terms of innovation dynamism. Her analysis coincides with ours at the following points:

(1) The development of innovation is described by an S-shaped curve approaching a maturity period due to the maturation of technology and saturation of the market. This means that Perez also recognizes the nonlinear nature of innovation.
(2) Economic development is brought about by innovation diffusion mostly to construct new industries
(3) Four innovation surges are identified since the first Industrial Revolution, and the fifth is starting.
(4) Within an innovation paradigm, there are periods of technological development and then market formation.
(5) Innovations are assimilated through the construction of economic infrastructures and networks.

I have learned much about from her great insight about innovation process, especially description of techno-economic paradigm and institutional change by construction of infrastructures and networks through assimilation of innovations with the economic society (e.g., Perez (1983) and Freeman (1984)). This kind of consideration is reflected as the introduction of concept of "trunk innovation" in this paper.

There are, however, some of different interpretation between Perez's and ours. One of the differences is about recognition of the Kondratiev wave. She criticizes the existence of Kondratiev waves in terms of the evaluation method. We, however, suggest that the Deployment Period perfectly coincides with our diffusion period of innovations, which strictly corresponds to the upswing of the Kondratiev wave.

Perez, as well as Freeman and Louçã, have doubts about the description of the Kondratiev wave by GNP or other economic aggregates. The Kondratiev wave is, however, defined as the result of economic analysis using GNP and other economic aggregates. As Schumpeter pointed out, a long business cycle is combined with Juglar cycles and fluctuates by other turbulences such as wars and oil crises. Therefore, it is rather difficult to expect a complete formation of Kondratiev waves due to such unavoidable fluctuations. Nevertheless, we can confirm the existence of Kondratiev business cycles. More importantly, this is not to confirm the strict existence of the long cycles but to acknowledge such long cycles and whether there is an explicit correlation between innovation clusters and such cycles. We should concentrate our effort on this point.

The second point about Perez's discussion is the drawing of the S-curve. Perez makes one S-curve throughout the innovation process. If the vertical axis is to be expressed by the amount of innovation products per annum as the dimension, the actual S-curve starts at the beginning of the Deployment Period because commercial products at the Installation Period are almost negligible compared with production at the Synergy Phase. Thus, the actual S-curve should be limitedly located around the Deployment Period.

The third point is the recognition of the Great Depression. According to our analysis, the bubble economy crashes at every peak of the Kondratiev wave. This kind of crash was observed in 1825, 1873, and 1929, after which a long recession follows for ten to twenty years. Such crash takes place at the end of the Deployment Period but not at the end of the Frenzy Phase. Crashes as identified by Perez seem to be only a shake-off process of venture business after a bandwagon effect, not enough to bring about a pervasive recession. The crash in 1929 should be located at the end of the third wave, not at the middle of the fourth. Both authors, Freeman and Louçã (2001) and Perez (2002), discard automobiles and the oil industry as the components of the third innovation surge and put them with those of the fourth. Although the production and consumption increase at the fourth wave according to the increase of GDP, these innovations certainly took off at the third wave and played decisive roles. The diffusion of these products surely started and soared with the upswing of third wave to reach a maturation level at that time just before the crash in 1929, and the highest production of cars never recovered before World War II. This trend at the third wave is also clear from the analysis of the car industry by Utterback (1994). The contribution of automobiles and oil to the third wave should be never neglected and should be considered as a main component of innovation surge because that these industries became triggers for the crash and the Great Depression.

As far as the recognition on the location of innovation surge is concerned; $i.e.$ timing, there is no discrepancy between these authors and us, and what remains could be an issue of interpretation. Especially, we should like to point out that definite surges of "innovation cluster" surely exist and are easy to locate and identify through the penetrative behavior of these innovation trajectories over various turbulences.

References

Bresnahan TF, Tritenberg M (1994) General purpose technologies: engines of growth? Journal of Econometrics 65(1): 441–452

David P (1975) Technical choice, innovation and economic growth. Cambridge University Press, Cambridge

Davies S (1979) The diffusion of process innovations. Cambridge University Press, Cambridge

Dickinson FG (1940) An aftercost of the world war to the United States. American Economic Review XXX(1): 326–339

Fischer JC, Pry RH (1971) A simple substitution model of technological change. Technological Forecasting and Social Change 3: 75–88

Forrester JW (1977) Growth cycles. De Economist 125(4): 525–543

Forrester JW (1979) Innovation and economic long wave. Management Review, 16–24

Freeman C (1983) Long waves in the world economy. Frances Pinter, London

Freeman C (1984) Keynes or Kondratiev? How can we get back to full employment? In: Marstrand P (ed) New technology and the future of work and skills. Frances Pinter, London

Freeman C, Clark J, Soete L (1982) Unemployment and technological innovation: a study of long waves in economic development. Frances Pinter, London

Freeman C, Louçã F (2002) As time goes by. Oxford University Press, Oxford

Goldstein JS (1985) Kondratieff waves as war cycles. International Studies Quarterly 29

Gordon DM (1978) American capitalism in crisis

Griliches Z (1957) Hybrid corn: an explanation in the economics of technological change. Econometrica 25(4): 501–522

Hansen AH (1941) Fiscal policy and business cycles. Norton, New York

Hirooka M (1992) Evaluation of technological trajectory and impact of technological innovation on the development of economy – examination of Schumpeter's concept by product life cycle. Paper Presented at the 4th Conference of the International J.A. Schumpeter Society, Kyoto

Hirooka M (1994a) Dynamism of technological innovation and economic development – a trial of Schumpeter's paradigm reconsidered. Paper Presented at the 5th Conference of the International J.A. Schumpeter Society, Münster, Germany

Hirooka M (1994b) Paradigm of technological innovation and business cycles – impact analysis of product cycles. Journal of Economics and Business Administration, Kobe University 169: 55–77 (in Japanese)

Hirooka M (1999) Dynamism of retail business and technological innovation. Journal of the University of Marketing and Distribution Sciences, in Parts of Information, Economics and Management Science 7(2): 73–86 (in Japanese)

Hirooka M (2000) Turnover of bubble economy and great depression – correlation with technological innovation. Journal of the University of Marketing and Distribution Sciences, in Parts of Information, Economics and Management Science 9(2): 1–19 (in Japanese)

Hirooka M (2002) Nonlinear dynamism of innovation and business cycles. Paper Presented at the 9th Conference of the International J.A.Schumpeter Society, Gainesville, Florida

Hirooka M (2003a) Nonlinear dynamism of innovation and knowledge transfer. Paper Presented at ASEAT & I&IR Conference, Manchester

Hirooka M (2003b) Technological innovation and economic growth – elucidation of nonlinear dynamism (in Japanese). Nihon Keizai Shinbunsha (Japan Economic News Paper Publishers Inc.), Tokyo

Hirooka M, Hagiwara T (1992) Characterization of diffusion trajectory of new products in the course of technological innovation. Kobe University Economic Review 38: 47–62

Kleinknecht A (1984) Innovation patterns in crisis and prosperity: Schumpeter's long cycle reconsidered. Doctoral Thesis, Free University of Amsterdam

Kondratiev ND (1926) Die langen Wellen der Konjuktur. Archiv für Sozialwissenschaft und Sozialpolitik 56: 573–606

Kuznets S (1940) Schumpeter's business cycles. American Economic Review XXX(2): 257–271

Louçã F (2002) Schumpeter and the new century: old lessons for new long cycles. Paper Presented at the 9th Conference of the International J.A. Schumpeter Society, Florida, Session C2

Mandel E (1972) Spätkapitalismus (Late capitalism). New Left Books, London

Mansfield E (1961) Technical change and the rate of imitation. Econometrica 29(4): 741–766

Mansfield E. (1963) Intrafirm rates of diffusion of an innovation. The Review of Economics and Statistics 45(4): 348–359

Mansfield E (1969) Industrial research and technological innovation. Longman, Upper Saddle River, NJ

Marchetti C (1980) Society as a learning system: discovery, invention, and innovation cycles revised. Technological Forecasting & Social Change 18: 267–278

Marchetti C, Nakićenović N (1979) The dynamics of energy systems and the logistic substitution model. Research Report RR-79-13, Luxeburg, Austria: International Institute for Applied Systems Analysis

Marchetti C (1988) Kondratiev's revisited – after one Kondratiev cycle. International IIASA Conference on Regularities of Scientific-Technological Progress and Long-Term Tendencies of Economic Development", Novosibirsk, USSR, March

Marchetti C, Mayer PS, Ausubel JH (1996) Human population dynamics revised with the logistic model. Technological Forecasting and Social Change 52: 1–30

Marchetti C, Meyer PS, Ausubel JH (1995) Population dynamics: how much can be modeled and predicted. Program for Human Environment, Rockefeller University

Mensch GO (1975) Stalemate in technology – innovations overcome the depression. Ballinger Publishing Co., Cambridge, MA

Metcalfe JS (1970) The diffusion of innovations in Lancashire textile industry. Manchester School of Economics and Social Studies 2: 145–162

Metcalfe JS (1981) Impulse and diffusion in the study of technical change. Futures 13: 347–359

Metcalfe JS (1994) Industrial life cycles and the theory of retardation: towards an ecology of industrial development. The Journal of Economics and Business Administration, Kobe University 170(6): 41–61

Modelski G (1987) Long cycles in world politics. University of Washington Press, Washington

Modis T (1992) Predictions. Simon & Schuster, New York

Mowery DC, Rosenberg N (1989) Technology and the pursuit of economic growth. Cambridge University Press, Cambridge

Nakićenović N, Grübler A (1991) Diffusion of technologies and social behavior. Springer, Berlin Heidelberg New York

Perez C (1983) Structural change and assimilation of new technologies in the economic and social system. Futures 15(4): 357–375

Perez C (2002) Technological evolutions and financial capital – the dynamics of bubbles and golden ages. Edward Elgar, Cheltenham, UK

Rosenberg N (1982) Inside the black box: technology and economics. Cambridge University Press, Cambridge, MA

Rosenberg N, Frischtak C (1984) Technological innovation and long waves. Cambridge Journal of Economics 8(1): 7–24

Rostow WW (1978) Getting from here to there. McGraw-Hill, New York

Schumpeter JA (1912) Theorie der Wirtschaftlichen Entwicklung, 2nd edn. Duncker & Humblot, Leipzig. English translation by Opie R (1934). Harvard University Press, Cambridge, MA

Schumpeter JA (1934) The theory of economic growth. Harvard University Press, Cambridge, MA

Schumpeter JA (1939) Business cycles. McGraw-Hill, New York

Stoneman P (1984) Theoretical approaches to the analysis of the diffusion of new technologies. In: MacDonald S, et al. (eds) The trouble with technology. Frances Pinter, London

Utterback JM (1994) Measuring dynamics of innovation. Harvard Business School, Cambridge, MA

van Duijn JJ (1983) The long wave in economic life. Allen & Unwin, London

Wallerstein IM (1979) The capitalist world-economy. Cambridge University Press, Cambridge, MA

The dynamic effects of general purpose technologies on Schumpeterian growth*

Iordanis Petsas

Department of Economics and Finance, 413 Brennan Hall, University of Scranton,
Scranton, PA 18510, USA (e-mail: iordanis.petsas@uofs.edu)

Abstract. General purpose technologies (GPTs) are drastic innovations characterized by pervasiveness in use and innovational complementarities. The dynamic effects of a GPT are analyzed within a quality-ladders model of scale-invariant Schumpeterian growth. The diffusion path of a GPT across a continuum of industries is governed by S-curve dynamics. The model generates a unique, saddle-path long-run equilibrium. Along the transition path, the measure of industries that adopt the new GPT increases, consumption per capita falls, and the interest rate rises. The growth rate of the stock market depends negatively on the rate of GPT diffusion and the magnitude of the GPT-ridden R&D productivity gains; and positively on the rate of population growth. It also follows a U-shaped path during the diffusion process of the new GPT. Finally, the model generates transitional growth cycles of per capita GNP.

Keywords: General purpose technologies – Schumpeterian growth – Scale effects – R&D races

JEL Classification: E3, O3, O4

1 Introduction

In any given economic "era" there are major technological innovations, such as electricity, the transistor, and the Internet, that have far-reaching and prolonged impact. These drastic innovations induce a series of secondary, incremental innovations. The introduction of the transistor, for example, triggered a sequence of

---

* I would like to thank Elias Dinopoulos for encouragement, and for constructive comments and suggestions. I would also like to thank David Figlio, Douglas Waldo, Steven Slutsky and participants in the 9th Biennial Congress of the International Schumpeter Society for useful discussions and suggestions. Any remaining errors are my own responsibility.

secondary innovations, such as the development of the integrated circuit and the microprocessor, which are themselves considered drastic innovations. These main technological innovations are used in a wide range of different sectors, inducing further innovations. For example, microprocessors are now used in many everyday products such as telephones, cars, personal computers, and so forth.

In general, drastic innovations have three key characteristics. The first feature refers to the generality of purpose, i.e., drastic innovations affect a wide range of industries and activities within industries. Consequently, Bresnahan and Trajtenberg (1995) christened these types of drastic innovations “General Purpose Technologies” (GPTs henceforth). Several empirical studies have documented the cross-industry pattern of diffusion for a number of GPTs.[1] In addition, a strand of empirical literature has established that the cross-industry diffusion pattern of GPTs is similar to the diffusion process of product-specific innovations and that it is governed by standard S-curve dynamics.[2] In other words, the internal-influence epidemic model can provide an empirically-relevant framework within which to analyze the dynamic effects of a GPT. During this diffusion process, these drastic innovations could generate growth fluctuations and even business cycles.

Second, the dynamic effects of these GPTs take a long period of time to materialize. For instance, David (1990) argues that it may take several decades before major technological innovations can have a significant impact on macroeconomic activity. Third, these GPTs act as “engines of growth”. As a better GPT becomes available, it gets adopted by an increasing number of user industries and fosters complementary advances that raise the industry’s productivity growth. As the use of a GPT spreads throughout the economy, its effects become significant at the aggregate level, thus affecting overall productivity growth. In his presidential address to the American Economic Association, Jorgenson (2001) documents the role of information technology in the resurgence of U.S. growth in the late 1990s.[3] There is plenty of evidence that the rise in structural productivity growth in the late 1990s can be traced to the introduction of personal computers and the acceleration in the

---

[1] For example, Helpman and Trajtenberg (1998b) provided evidence for the diffusion of the transistor. They state that transistors were first adopted by the hearing aids industry. Later, transistors were used in radios, followed by their adoption by the computer industry. These three industries are known as early adapters. The fourth sector to adopt the transistor was the automobile industry, followed by the telecommunications sector.

[2] Griliches (1957), for example, studied the diffusion of hybrid seed corn in 31 states and 132 crop-reporting areas among farmers. His empirical model generates an S-curve diffusion path. Andersen (1999) confirmed the S-shaped growth path for the diffusion of entrepreneurial activity, using corporate and individual patents granted in the U.S. between 1890 and 1990. Jovanovic and Rousseau (2001) provided more evidence for an S-shaped curve diffusion process by matching the spread of electricity with that of personal computer use by consumers.

[3] At the aggregate level, information technology is identified with the output of computers, communications equipment, and software. These products appear in the GDP as investments by businesses, households, and governments along with net exports to the rest of the world.

price reduction of semiconductors, which constituted the necessary building blocks for the information technology revolution.[4]

The growth effects of GPTs have been analyzed formally by Helpman and Trajtenberg (1998a). In their model, GPTs require complementary inputs before they can be applied profitably in the production process. Complementary inputs developed for previous GPTs are not suited for use with a newly arrived GPT. The sequential arrival of GPTs generates business cycles. A typical cycle consists of two phases, a phase in which firms produce final goods with the old GPT and components are developed for the new GPT, and a second phase in which final goods producers switch to the new GPT and the development of components for that GPT continues. Output declines in the first phase of a cycle as workers switch from production to research to invent new inputs and increases again in the second phase once the new technology is implemented.[5]

Following the Helpman and Trajtenberg (1998a) model, Aghion and Howitt (1998b) explored the macroeconomic effects of GPTs. They derived a simple version of the model from the basic Schumpeterian model of endogenous growth by adding a second stage to the innovation process, a stage of component-building, and they endogenized the arrival times of successive GPTs.[6] Their model results in similar per capita GNP growth cycles due to the adoption of the new GPT.

In this paper, I analyze formally the effects of a GPT within a state-of-the-art model of Schumpeterian growth without scale effects. Schumpeterian (R&D-based) growth is a type of growth that is generated through the endogenous introduction of new goods or processes based on Schumpeter's (1934) process of creative destruction, as opposed to physical or human-capital accumulation.[7]

Earlier models of Schumpeterian growth assumed that the growth rate of technological change depends positively on the level of R&D resources devoted to innovation at each instant in time. As population growth causes the size of the economy (scale) to increase exponentially over time, R&D resources also grow exponentially, as does the long-run growth rate of per capita real output. In other words, long-run Schumpeterian growth in these models exhibits scale effects. Two influential papers by Jones (1995a,b) provided time series evidence for the absence

---

[4] Another study from OECD documents that U.S. investment in information processing equipment and software increased from 29% in 1987 to 52% in 1999. The diffusion of information and communication equipment accelerated after 1995 as a new wave of information and communication equipment, based on applications such as the World Wide Web and the browser, spread rapidly throughout the economy.

[5] There is a growing literature with this approach. See, for example, Helpman and Rangel (1998), Aghion and Howitt (1998b), and the volume edited by Helpman (1998).

[6] Eriksson and Lindh (2000) explored a variation of the Helpman and Trajtenberg (1998a) model in which technological development occurs partly by discrete replacements of obsolete technologies and the timing of technology shifts is endogenized.

[7] There are two classes of scale invariant Schumpeterian growth models; endogenous and exogenous. Endogenous [exogenous] Schumpeterian growth models are those in which *long-run* growth can [cannot] be affected by permanent policy changes.

of these scale effects. This evidence led theorists to construct Schumpeterian growth models that exclude scale effects.[8]

My approach to modeling the GPTs has the following features. First, the model abstracts from scale effects and generates long-run growth, which is consistent with the time-series evidence presented by Jones (1995a). Second, I take into consideration the above mentioned evidence on long diffusion lags associated with the adoption of a new GPT. I therefore analyze both the transitional dynamics and long-run effects of a new GPT. Third, I assume that a GPT is beneficial to all firms in each industry. Thus, when a GPT is implemented in an industry, it affects the productivity of R&D workers, the size of all future innovations in that industry and its growth rate. Finally, I assume that, although a GPT's rate of diffusion is exogenous, its diffusion path across a continuum of industries is governed by S-curve dynamics.

I incorporate the presence of a GPT into the standard quality-ladder framework of Schumpeterian growth without scale effects that was developed by Dinopoulos and Segerstrom (1999). In the model, there is positive population growth and one factor of production, labor. Final consumption goods are produced by a continuum of structurally identical industries. Labor in each industry can be allocated between two economic activities, manufacturing of high-quality goods and R&D services that are used to discover new products of higher quality.

The arrival of innovations in each industry is governed by a memoryless Poisson process whose intensity depends positively on R&D investments and negatively on the rate of difficulty of conducting R&D. Following Dinopoulos and Segerstrom (1999), I assume that R&D becomes more difficult over time in each industry. Specifically, I assume that the productivity of R&D workers declines as the size of the market (measured by the level of population) increases. This assumption captures the notion that it is more difficult to introduce new products and replace old ones in a larger market.[9]

The main purpose of this paper is to explore the effects of GPTs on Schumpeterian growth. Thus, it is imperative to develop a Schumpeterian growth model

---

[8] Dinopoulos and Thompson (1999) provided a survey of the empirical evidence on scale and growth, and describe recent attempts to develop models that generate growth without scale effects.

[9] Several authors have developed microfoundations for this assumption. Young (1998), Dinopoulos and Thompson (1998), and Aghion and Howitt (1998a, chapter 12) have combined tastes for horizontal and vertical product differentiation to generate models in which absolute levels of R&D drive productivity growth at the firm-level, but aggregate R&D in larger economies is diffused over a larger number of product lines or industries. At the steady state, the number of varieties is proportional to the level of population. As population grows, the number of varieties increases and aggregate R&D is diffused over a larger number of product lines or industries, making R&D more difficult.

Dinopoulos and Syropoulos (2000) have provided microfoundations for this specification in a model of Schumpeterian growth, where the discovery of higher quality products is modeled as an R&D contest (as opposed to an R&D race) in which challengers engage in R&D and incumbent firms allocate resources to rent-protecting activities. Rent-protecting activities are defined as costly attempts of incumbent firms to safeguard the monopoly rents from their past innovations. These activities can delay the innovation of better products by reducing the flow of knowledge spillovers from incumbents to potential challengers, and/or increase the costs of copying existing products. Their model postulates that R&D may become more difficult as the size of the economy grows because incumbent firms may allocate more resources to rent-protecting activities.

that is scale effect free and to analyze the behavior of GPTs within this framework. After removing the scale effects property from the model, I can discuss its transitional and long-run properties and implications. I use Mulligan and Sala-i-Martin's (1992) time-elimination method to study the transitional dynamics of the model.

This analysis generates several novel findings. First, there exists a unique globally stable-saddle-path along which the measure of industries that adopt the new GPT increases, per capita consumption expenditure decreases, the market interest rate increases, and the innovation rate of those industries that have adopted the new GPT decreases at a higher rate than that of those that have not adopted the new GPT. Second, the model exhibits transitional growth cycles of per capita GNP.

In previous GPT-driven growth models, GPTs also generate transitional growth cycles of per capita GNP. However, their results are not robust to the introduction of positive population growth. The introduction of positive population growth in Helpman and Trajtenberg (1998a) and Aghion and Howitt's (1998b) models of GPTs, will make these growth cycles shorter and shorter as the size of the economy increases, and in the long-run the GPT-induced cycles disappear. In the present model, the fall in output comes from the reduction in per capita consumption expenditure on final goods and the rise in the per capita R&D investment. As the size of the economy increases, the duration of the per capita GNP cycle remains the same. When all industries have adopted the new GPT and the diffusion process has been completed, the economy experiences a higher per capita income constant growth rate.

In the absence of a new GPT, the economy does not exhibit zero long-run growth as in previous models of GPTs (see Helpman and Trajtenberg, 1998a; Aghion and Howitt, 1998b). That is, the long-run growth rate depends positively on the rate of innovation (which equals per capita R&D) and thus any policy that affects per capita R&D investment has long-run growth effects.[10] In addition, the removal of scale effects allows one to analyze the effects of changes in the rate of growth of population that is absent from earlier models.

I also analyze the effects of a new GPT on the stock market. The growth rate of the stock market depends negatively on the rate of GPT diffusion process and the magnitude of the GPT-ridden R&D productivity gains, and positively on the rate of population growth. It also follows a U-shaped path during the diffusion process of the new GPT (Proposition 4). During the transition from the old to the new GPT, there are two types of industries in the economy: one that has adopted the new GPT and one that has not yet adopted it. The former type of industry is more innovative in terms of discovering higher quality products than the latter type. In the initial stages of a GPT's diffusion, the aggregate stock value decreases, since most of the industries belong to the latter type. As more industries switch to the new GPT, the aggregate stock value rises. This result is consistent with

---

[10] The evidence on the empirical validity of endogenous versus exogenous Schumpeterian growth models without scale effects is still limited. However, Zachariadis (2003) found strong support for the Schumpeterian endogenous growth framework without scale effects by using U.S. manufacturing industry data for the period 1963–1988. The manufacturing sector accounted for more than ninety percent of R&D expenditures in the U.S. until the late eighties.

that of previous GPT-driven growth models.[11] In addition, an increase in the GPT diffusion rate increases the economy-wide resources devoted to R&D. Thus, the probability that the incumbent firm will be replaced by a follower firm increases. In other words, when the GPT diffusion process accelerates, the decrease in per capita consumption expenditure is more severe, and per capita R&D investment increases. This last result provides a novel link between the GPT adoption and higher risk for incumbent firms and captures the effects of creative destruction on the stock market valuation of monopoly profits.[12]

However, the mechanism identified in the present model that links the growth rate of the stock market with the GPT differs from that of previous GPT-driven growth models. In Helpman and Trajtenberg's (1998a) model, for example, during the first phase, the components of both the best practice GPT and of the previous one have positive value. When the economy is in the second phase of a typical cycle, only components of the best practice GPT are valuable because at that time it is known that no component of the older technologies will ever be used. Thus, the introduction of a new GPT brings a sharp decline in the real value of the stock market during a substantial part of phase one, but it picks up toward the end of the phase. In the second phase, the stock market rises.

The effect of the GPT diffusion on the aggregate investment during the adoption process is ambiguous (Proposition 5). In the initial stages of the diffusion process, only a limited number of industries adopt the new GPT. These industries are called the early adopters. As more industries adopt the new GPT, aggregate R&D investment increases.

The rest of the paper is organized as follows. Section 2 develops the structure of the model. Section 3 analyzes the long-run properties of the model and Section 4 deals with the transitional dynamics. Section 5 summarizes the model's key findings and suggests possible extensions. The algebraic details and proofs of propositions are relegated to the Appendix.

---

[11] Jovanovic and Rousseau (2001) documented empirically how technology has affected the U.S. economy over the past century, using 114 years of U.S. stock market data. Their estimates reveal evidence that entries to the stock market as a percentage of firms listed in each year, were proportionately largest between 1915 and 1929, and that these levels were not again approached until the mid-1980s. About half of American households and most businesses were connected to electricity in 1920, and about one half of the households and most businesses today own or use computers. Both expansions, therefore, coincide with periods during which electricity and information technology saw widespread adoption. During times of rapid technological change, the new entrants of the stock market will grab the most value from previous entrants because the incumbents will find hard to keep up. The downward trend in the starting values of the vintages reflects a slowing down in the growth of the stock market.

[12] The first OPEC shock may also explain a part of the drop in the stock market in the early 1970s, as well as a part of the productivity slowdown. Hobijn and Jovanovic (2001) argued that there are several problems associated with the oil-shock explanation. One problem is that a rise in oil prices should have lowered current profits more than future profits, because of the greater ease of finding substitutes for oil in the long-run, perhaps current output more than future output and, therefore, should have produced a rise in the ratio of market capitalization to GDP, not a fall. This scenario also implies a constant entry in the stock market, something that contradicts their evidence. Another problem that is associated with the oil-price-shock explanation for the stock-market drop is that the energy-intensive sectors did not experience the largest drop in value in 1973–1974. Their evidence supports that the information-technology-intensive sectors experienced the largest drop in 1973–1974.

2 The model

2.1 Industry structure

I consider an economy with a continuum of industries, indexed by $\theta \in [0, 1]$. In each industry θ, firms are distinguished by the quality j of the products they produce. Higher values of j denote higher quality, and j is restricted to taking on integer values. At time $t = 0$, the state-of-the-art quality product in each industry is $j = 0$, that is, some firm in each industry knows how to produce a $j = 0$ quality product and no firm knows how to produce any higher quality product. To learn how to produce higher quality products, firms in each industry engage in R&D races. In general, when the state-of-the-art quality in an industry is j, the next winner of an R&D race becomes the sole producer of a $j + 1$ quality product. Thus, over time, products improve as innovations push each industry up its "quality ladder", as in Grossman and Helpman (1991).

2.2 Diffusion of a new GPT

The diffusion path of a new GPT is modeled as follows: The economy has achieved a steady-state equilibrium, manufacturing final consumption goods with an old GPT. I begin the analysis at time $t = t_0$, when a new GPT arrives unexpectedly. Firms in each industry start adopting the new GPT at an exogenous rate.[13]

I use the epidemic model to describe the diffusion of a new GPT across the continuum of industries.[14] Its form can be described by the following differential equation,

$$\frac{\dot{\omega}}{\omega} = \delta(1 - \omega), \tag{1}$$

where $\dot{\omega} = \partial\omega/\partial t$ denotes the rate of change in the fraction of industries that use the new GPT and $\delta > 0$ is the rate of diffusion. Equation (1) states that the number of new adoptions during the time interval dt, $\dot{\omega}$, is equal to the number of remaining potential adopters, $(1 - \omega)$, multiplied by the probability of adoption, which is the product of the fraction of industries that have already adopted the new GPT, ω, and the parameter δ, which depends upon factors such as the attractiveness of the innovation and the frequency of adoption, both of which are assumed to be exogenous.

---

[13] Aghion and Howitt (1998b) model the spread of GPTs using a continuum of sectors. In their model, the innovation process involves three stages. First, the GPT is discovered. Then each sector discovers a "template" on which research can be based. Finally, that sector implements the GPT when its research results in a successful innovation. They have computed paths of the fraction of sectors experimenting with the new GPT and the fraction using the new GPT and found that the time path of the later follows a logistic curve (S-curve).

[14] See Thirtly and Ruttan (1987, pp. 77–89) for various applications of the epidemic model to the diffusion of technology.

The solution to Equation (1) expresses the measure of industries that have adopted the new GPT as a function of time and yields the equation of the sigmoid (S-shaped) logistic curve:

$$\omega = \frac{1}{[1 + e^{-(\gamma + \delta t)}]}, \tag{2}$$

where γ is the constant of integration. Notice that for $t \to \infty$, Equation (2) implies that all industries have adopted the new GPT.[15]

2.3 Households

The economy is populated by a continuum of identical dynastic families that provide labor services in exchange for wages, and save by holding assets of firms engaged in R&D. Each individual member of a household is endowed with one unit of labor, which is inelastically supplied. The number of members in each family grows over time at the exogenous rate $g_N > 0$. I normalize the measure of families in the economy at time 0 to equal unity. Then the population of workers in the economy at time t is $N(t) = e^{g_N t}$.

Each household is modeled as a dynastic family,[16] which maximizes the discounted utility

$$U = \int_0^{\infty} e^{-(\rho - g_N)t} \log\, u(t) dt, \tag{3}$$

where $\rho > 0$ is the constant subjective discount rate. In order for U to be bounded, I assume that the effective discount rate is positive (i.e., $\rho - g_N > 0$). Expression $\log u(t)$ captures the per capita utility at time t, which is defined as follows:

$$\log u(t) \equiv \int_0^1 \log[\sum_j \lambda(\theta)^j q(j, \theta, t)] d\theta \ . \tag{4}$$

In Equation (4), $q(j, \theta, t)$ denotes the quantity consumed of a final product of quality j in industry $\theta \in [0, 1]$ at time t. Parameter $\lambda(\theta)$ measures the size of quality improvements and is equal to

$$\lambda(\theta) = \begin{cases} \lambda_1 & if\ \theta \in [0,\ \omega] \\ \\ \lambda_0 & if\ \theta \in [\omega,\ 1], \end{cases} \tag{5}$$

---

[15] When $t \to -\infty$, then $\omega = 0$. If one assumes that the new GPT arrives at time $t = 0$, then $\omega > 0$. That is, the new GPT is introduced in the economy by a given fraction of industries ω (i.e., the industry or industries that developed this particular GPT).

[16] Barro and Sala-i-Martin (1995, Ch.2) provide more details on this formulation of the household's behavior within the context of the Ramsey model of growth.

where $\lambda_1 > \lambda_0 > 1$. At each point in time t, each household allocates its income to maximize Equation (4) given the prevailing market prices. Solving this optimal control problem yields a unit elastic demand function for the product in each industry with the lowest quality-adjusted price

$$q(j, \theta, t) = \frac{c(t)N(t)}{p(j, \theta, t)}, \tag{6}$$

where $c(t)$ is per capita consumption expenditure, and $p(j, \theta, t)$ is the market price of the good considered. The quantity demanded of all other goods is zero.

Given this static demand behavior, the intertemporal maximization problem of the representative household is equivalent to

$$\max_{c(t)} \int_0^\infty e^{-(\rho - g_N)t} \log c(t) dt \ , \tag{7}$$

subject to the intertemporal budget constraint $\dot{a}(t) = r(t)a(t) + w(t) - c(t) - g_N a$, where $a(t)$ denotes the per capita financial assets, $w(t)$ is the wage income of the representative household member, and $r(t)$ is the instantaneous rate of return. The solution to this maximization problem obeys the well-known differential equation

$$\frac{\dot{c}(t)}{c(t)} = r(t) - \rho \ , \tag{8}$$

According to Equation (8), per capita consumption expenditure increase over time if the instantaneous interest rate exceeded the consumer's subjective discount rate ρ.

2.4 Product markets

Every firm in each industry θ uses labor $L(\theta, t)$ as the sole input in its production, according to the following production function

$$Q(\theta, t) = \frac{L(\theta, t)}{\alpha_Q} \ , \tag{9}$$

where α_Q is the unit labor requirement. The monopolist engages in limit pricing, i.e., it charges a price equal to unit cost of manufacturing a product times the quality increment

$$P = \lambda(\theta)\alpha_Q w \ . \tag{10}$$

At each instant in time, the incumbent monopolist produces the state-of-the-art quality product and earns a flow of profits

$$\pi(\theta, t) = \left(\frac{\lambda(\theta) - 1}{\lambda(\theta)}\right) c(t)N(t) \ . \tag{11}$$

2.5 R&D races

Labor is the only input used to do R&D in any industry. Each firm in each industry θ produces R&D services by employing labor $L_R(\theta, t)$ under the constant returns to scale production function[17]

$$R(\theta, t) = \frac{\mu(\theta)}{\alpha_R} L_R(\theta, t) \ , \tag{12}$$

In Equation (12), $\alpha_R/\mu(\theta)$ is the unit-labor requirement in the production of R&D services and $\mu(\theta)$ is equal to

$$\mu(\theta) = \begin{cases} \mu_1 = \mu & \text{if} \quad \theta \in [0, \omega] \\ \\ \mu_0 = 1 & \text{if} \quad \theta \in [\omega, 1], \end{cases} \tag{13}$$

where $\mu > 1$. A firm k that engages in R&D discovers the next higher-quality product with instantaneous probability $I_k dt$, where dt is an infinitesimal interval of time and

$$I_k(\theta, t) = \frac{R_k(\theta, t)}{X(t)} \ . \tag{14}$$

$R_k(\theta, t)$ is firm k's R&D outlays and $X(t)$ captures the difficulty of R&D in a typical industry. I assume that the returns to R&D investments are independently distributed across challengers, across industries, and over time. Therefore, the industry-wide probability of innovation can be obtained from Equation (14) by summing up the levels of R&D across all challengers. That is,

$$I(\theta, t) = \sum_k I_k(\theta, t) = \frac{R(\theta, t)}{X(t)} \ , \tag{15}$$

where and $R(\theta, t)$ denotes total R&D services in industry θ. Variable $I(\theta, t)$ is the effective R&D.[18] The arrival of innovations follows a memoryless Poisson process with intensity I_1 for the industries that have adopted the new GPT, and I_0 for industries that have not adopted the new GPT.

Early models of Schumpeterian growth considered $X(t)$ to be constant over time. This implied that the rates of innovation and the long-run growth increase exponentially as the scale of the economy grows exponentially. This scale-effects property is inconsistent with post-war time-series evidence presented in Jones (1995a).

---

[17] The empirical evidence on returns to scale of R&D expenditure is inconclusive. Diminishing returns would make the analysis of the transitional dynamics more complicated. Segerstrom and Zolnierek (1999) among others developed a model in which they allow for diminishing returns to R&D effort at the firm level and industry leaders have R&D cost advantages over follower firms. In their model, when there are diminishing returns to R&D and the government does not intervene, both industry leaders and follower firms invest in R&D.

[18] The variable $I(\theta, t)$ is the intensity of the Poisson process that governs the arrivals of innovations in industry θ.

A recent body of theoretical literature has developed models of Schumpeterian growth without scale effects.[19] Two approaches have offered possible solutions to the scale-effects problem. The first generates exogenous long-run Schumpeterian growth models.[20] The second approach generates models that exhibit endogenous long-run Schumpeterian growth.[21] Here I adopt the second approach and remove the scale-effects property by assuming that the level of R&D difficulty is proportional to the market size measured by the level of population,

$$X(t) = kN(t) \; , \tag{16}$$

where $k > 0$ is a parameter.[22]

Consumer savings are channeled to firms engaging in R&D through the stock market. The assumption of a continuum of industries allows consumers to diversify the industry-specific risk completely and earn the market interest rate. At each instant in time, each challenger issues a flow of securities that promise to pay the flow of monopoly profits defined in (11) if the firm wins the R&D race and zero otherwise. Consider now the stock-market valuation of the incumbent firm in each industry. Let $V(t)$ denote the expected discounted profits of a successful innovator at time t when the monopolist charges a price p for the state-of-the-art quality product. Because each quality leader is targeted by challengers who engage in R&D to discover the next higher-quality product, a shareholder faces a capital loss $V(t)$ if further innovation occurs. The event that the next innovation will arrive occurs with instantaneous probability Idt, whereas the event that no innovation will arrive occurs with instantaneous probability $1 - Idt$. Over a time interval dt, the shareholder of an incumbent's stock receives a dividend $\pi(t)dt$ and the value of the incumbent appreciates by $dV(t) = [\partial V(t) \big/ \partial t]dt = \dot{V}(t)dt$. The absence of profitable arbitrage opportunities requires the expected rate of return on stock issued by a successful innovator to be equal to the riskless rate of return r; that is,

$$\frac{\dot{V}(\theta,t)}{V(\theta,t)}[1 - I(\theta,t)dt]dt + \frac{\pi(\theta,t)}{V(\theta,t)}dt - \frac{[V(\theta,t) - 0]}{V(\theta,t)}I(\theta,t)dt = rdt \; . \tag{17}$$

---

[19] See Dinopoulos and Thompson (1999) and Dinopoulos and Sener (2003) for an overview of these models.

[20] Jones (1995b), and Segerstrom (1998) have removed scale effects by assuming that R&D becomes more difficult over time because "the most obvious ideas are discovered first." The model that results from their specification is called the *temporary effects of growth* (TEG) model. In these models, the growth rate does not depend on any measure of scale. Increases in the steady-state level of R&D raise technology and income per capita at any point in time, but they do not raise the growth rate.

[21] Young (1998), Aghion and Howitt (1998a, chapter 12), Dinopoulos and Thompson (1998), Peretto (1998), and Peretto and Smulders (1998) removed the scale effects property by essentially the same mechanism as the one developed by exogenous Schumpeterian growth models. They introduced the concept of localized intertemporal R&D spillovers. Dinopoulos and Syropoulos (2000) proposed a novel mechanism based on the notion of innovation-blocking activities that removes the scale-effects property and generates endogenous long-run Schumpeterian growth. This model offers a novel explanation to the observation that the difficulty of conducting R&D has been increasing over time.

[22] Informational, organizational, marketing, and transportation costs can readily account for this difficulty. Arroyo, et al. (1995) have proposed this specification under the name of the *permanent effects of growth* (PEG) model, and have provided time-series evidence for its empirical relevance.

Taking limits in Equation (17) as $dt \to 0$ and rearranging terms appropriately gives the following expression for the value of monopoly profits

$$V(\theta, t) = \frac{\pi(\theta, t)}{r(t) + I(\theta, t) - \frac{\dot{V}(\theta,t)}{V(\theta,t)}}. \tag{18}$$

Consider now the maximization problem of a typical challenger k. This firm chooses the level of R&D investment $R_k(\theta, t)$ to maximize the expected discounted profits

$$V(\theta, t)\frac{R_k(\theta, t)}{X(t)}dt - w\frac{\alpha_R}{\mu(\theta)}R_k(\theta, t)dt \ , \tag{19}$$

where $I_k dt = [R_k(\theta, t)/X(t)]dt$ is the instantaneous probability it will discover the next higher-quality product and $w\alpha_R R_k(\theta, t)/\mu(\theta)$ is the R&D cost of challenger k.

Free entry into each R&D race drives the expected discounted profits of each challenger down to zero and yields the following equilibrium condition:

$$V(\theta, t) = \frac{w\alpha_R k N(t)}{\mu(\theta)} \ . \tag{20}$$

2.6 Labor market

All workers are employed by firms in either production or R&D activities. Taking into account that each industry leader charges the same price p, and that consumers only buy goods from industry leaders in equilibrium, it follows from (9) that total employment of labor in production is $\int_0^1 Q(\theta, t)d\theta$. Solving (12) for each industry leader's R&D employment $L_R(\theta, t)$ and then integrating across industries, total R&D employment by industry leaders is $\int_0^1 [R(\theta, t)\alpha_R/\mu(\theta)]d\theta$. Thus, the full employment of labor condition for the economy at time t is

$$N(t) = \int_0^1 Q(\theta, t)\alpha_Q d\theta + \int_0^1 \frac{\alpha_R R(\theta, t)}{\mu(\theta)}d\theta \ . \tag{21}$$

Equation (21) completes the description of the model.

3 Long-run equilibrium

The dynamic behavior of the economy is governed by two equations that determine the evolution of the per capita consumption expenditure, c, and the number of industries that adopt the new GPT, ω. To facilitate the interpretation and understanding of my results, I begin by deriving expressions for long-run per capita real output and long-run growth. Following the standard practice of Schumpeterian growth models, one can obtain the following deterministic expression for sub-utility $u(t)$, which is appropriately weighted consumption index and corresponds to real per capita income[23]

$$\log u(t) = \log c - \log \alpha_Q + I(\theta, t)t \log \lambda(\theta) - \log \lambda(\theta) \ . \tag{22}$$

The economy's long-run Schumpeterian growth rate is defined as the rate of growth of sub-utility $u(t)$, $g_u = \dot{u}(t)/u(t)$. By differentiating Equation (22) with respect to time, I obtain:

$$g_u = \frac{\dot{u}(t)}{u(t)} = I(\theta, t) \log \lambda(\theta) \ , \tag{23}$$

which is a standard expression for long-run growth in quality-ladders growth models. Because the size of each innovation becomes larger (i.e., $\lambda_1 > \lambda_0$) after all industries have adopted the new GPT (i.e., the diffusion process has been completed), long-run growth, g_u, can be affected not only through changes in the rate of innovation, but also through the diffusion of the new GPT.

At this point it is useful to choose labor as the numeraire of the model and set

$$w \equiv 1 \ . \tag{24}$$

Using Equations (6), (10), (12), (15), (16), and (21), and taking into account (24), I obtain the *resource condition*

$$1 = c\left(\frac{\omega}{\lambda_1} + \frac{(1-\omega)}{\lambda_0}\right) + k\alpha_R\left(\frac{\omega}{\mu} I_1 + (1-\omega) I_0\right) \ , \tag{25}$$

which defines a negative linear relationship between per capita consumption expenditure, c, and the effective R&D, I. The above resource condition holds at each instant in time because by assumption factor markets clear instantaneously.

I now derive the differential equation that determines the growth rate of per capita consumption expenditure, $\dot{c}/c$, as a function of its level and the rate of innovation. Equation (20) holds at each instant in time, and yields $\dot{V}(\theta, t) \Big/ V(\theta, t) = \dot{X}(t) \Big/ X(t) = g_N$. In other words, the values of expected discounted profits, $V(t)$, and the level of R&D difficulty, $X(t)$, grow at the constant rate of population growth, g_N. Using Equations (18) and (20), I obtain

$$c = \frac{\lambda(\theta)\alpha_R k}{(\lambda(\theta) - 1)\mu(\theta)}][\rho + I(\theta, t) - g_N] \ , \tag{26}$$

which defines a positive linear relationship between per capita consumption expenditure, c, and the effective R&D, I. It also implies the familiar condition that $r = \rho$, which means that the market interest rate must be equal to the subjective discount rate in the steady-state equilibrium. This property is shared by all Schumpeterian models in which growth is generated by the introduction of final consumption (as opposed to intermediate production) goods.

---

[23] See Dinopoulos (1994) for an overview on Schumpeterian growth theory.

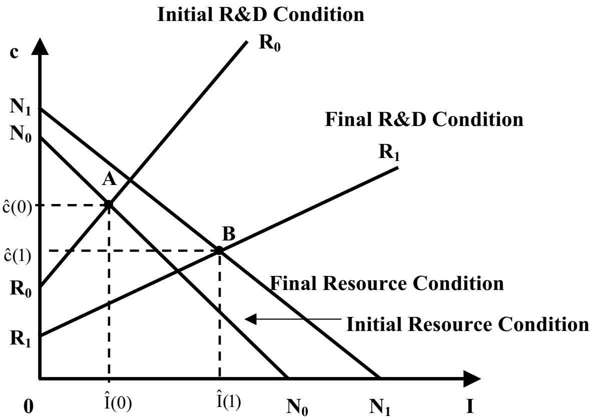


Fig. 1. Steady-state equilibria: Point A: No industry has adopted the new GPT. Point B: All industries have adopted the new GPT

Let a hat "^" over variables denote their market value in steady-state equilibrium. The resource condition (25) and the equilibrium R&D condition (26) determine simultaneously the long-run equilibrium values of per capita consumption expenditure, $\hat{c}$, and the rates of innovation, $\hat{I}_0$ and $\hat{I}_1$. Figure 1 illustrates the two steady-state equilibria: the initial steady-state (point A) in which no industry has adopted the new GPT (i.e., $\omega = 0$) and the final steady-state (point B) in which all industries have adopted the new GPT (i.e., $\omega = 1$). When $\omega = 0$ the balanced-growth resource condition is

$$1 = \frac{c}{\lambda_0} + k\alpha_R I_0 \ , \tag{27}$$

and the balanced-growth R&D condition is given by Equation (26) (when $\omega = 0$). The vertical axis measures consumption expenditure per capita, c, and the horizontal axis measures the rate of innovation, I. The resource condition is reflected by the negatively-sloped line N_0N_0 and the R&D equilibrium condition is represented by the positively-sloped line R_0R_0. Their unique intersection at point A determines the long-run values $\hat{c}(0)$ and $\hat{I}_0(0)$, where $\hat{c}(0)$ denotes the per capita consumption expenditure evaluated at $\omega = 0$ and $\hat{I}_0(0)$ denotes the innovation rate for industries that have adopted the new GPT evaluated at $\omega = 0$. Therefore, I arrive at:

Proposition 1. *For a given $\omega \in [0,1]$, where ω is the measure of industries with a new GPT, there exists a unique steady-state equilibrium such that the long-run Schumpeterian growth, $\hat{g}_u$, is endogenous and does not exhibit scale effects: it depends positively on policies that affect the size of innovations, λ, the labor productivity in R&D services, $\mu(\theta)/\alpha_R$, and the rate of population growth, g_N; it depends negatively on the consumer's subjective discount rate, ρ. At each steady-state equilibrium, consumption expenditure per capita, $\hat{c}$, is constant, the interest rate, $\hat{r}(t)$, is equal to the constant subjective discount rate, ρ, and the aggregate stock value, $\bar{V}$, increases at the same rate as the constant rate of population growth, g_N.*

Proof. See Appendix.

The removal of scale effects from the long-run growth rate, g_u, depends on the assumption that the level of R&D difficulty is proportional to market size. At the steady-state equilibrium, the level of R&D difficulty, $X(t)$, increases exponentially at the rate of population growth g_N, i.e., $\dot{X}(t)\Big/X(t) = g_N$, as can be seen from Equation (16). The absence of a new GPT does not result in zero long-run growth rate, as in the Helpman and Trajtenberg (1998a) and Aghion and Howitt (1998b) models. That is, the long-run growth rate depends positively on per capita R&D and, thus, any policy that affects this variable has long-run growth effects. The following proposition describes the long-run properties of the economy:

Proposition 2. *If ω is governed by S-curve dynamics, there are only two steady-state equilibria: the initial steady-state equilibrium arises before the adoption of the new GPT, where $\omega = 0$, and the final steady-state equilibrium is reached after the diffusion process of the new GPT has been completed, where $\omega = 1$. At the final steady-state equilibrium: aggregate investment is higher, $\hat{I}(1) > \hat{I}(0)$, long-run growth rate is higher, $\hat{g}_u(1) > \hat{g}_u(0)$, per capita consumption expenditure is lower, $\hat{c}(1) < \hat{c}(0)$, per capita stock market valuation of the incumbent in each industry is lower, $\hat{V}(1)\Big/N < \hat{V}(0)\Big/N$, relative to the initial steady-state equilibrium. In both steady states the market interest rate is equal to the subjective discount rate, $\hat{r} = \rho$.*

Proof. See Appendix.

These comparative steady-state properties can be illustrated with the help of Figure 1. Before the introduction of the new GPT, the economy is in a steady state (point A), where $\omega = 0$, with per capita consumption expenditure $\hat{c}(0)$, and with innovation rate $\hat{I}_0$. An increase in the measure of industries that adopt the new GPT makes the R&D condition in Figure 1 shift downward from R_0R_0 (where $\omega = 0$) to R_1R_1 (where $\omega = 1$) and the resource condition shift upward from N_0N_0 to N_1N_1, resulting in higher long-run rate of innovation and in lower long-run consumption expenditure per capita. In other words, when all industries have adopted the new GPT, the long-run Schumpeterian growth rate increases. The new steady state is at point B, where $\omega = 1$, with per capita consumption expenditure $\hat{c}(1)$, and innovation rate $\hat{I}_1$.

4 Transitional dynamics

I analyze the transitional dynamics of the model by adapting the time-elimination method described by Mulligan and Sala-i-Martin (1992).[24] The time-elimination method enables one to construct a system of two differential equations that govern the evolution of c and ω. Since Equation (26) holds at each instant in time (when the subjective discount rate, ρ, is replaced by the interest rate, r), I can solve for the rates of innovation for the two types of industries, I_0 and I_1. After substituting

---

[24] See also Mulligan and Sala-i-Martin (1991) for more details on this method.

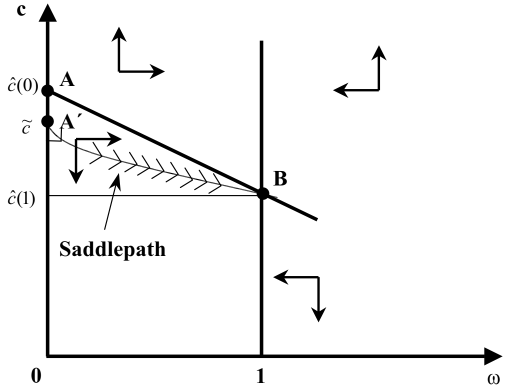


Fig. 2. Stability of the balanced-growth equilibrium

these rates into the resource condition (25), which holds at each instant in time, I can solve for the market interest rate along any path and obtain

$$r = \frac{\mu(c-1)}{k\alpha_R[\mu - \omega(\mu-1)]} + g_N \; . \tag{28}$$

Substituting (28) into (8) yields the following differential equation:

$$\frac{\dot{c}}{c} = r - \rho = \frac{\mu(c-1)}{k\alpha_R[\mu - \omega(\mu-1)]} + g_N - \rho \; . \tag{29}$$

Equations (29) and (1) determine the evolution of the two endogenous variables of the model, per capita consumption expenditure, c, and the number of industries that have adopted the new GPT, ω.

Since the right-hand side of Equation (29) is decreasing in ω, $\dot{c} = 0$ defines the downward-sloping curve in Figure 2. Starting from any point on this curve, an increase in ω leads to $\dot{c} > 0$ and a decrease in ω leads to $\dot{c} < 0$. The right-hand side of Equation (1) is independent of c, and therefore the $\dot{\omega} = 0$ locus is a vertical line. Starting from any point on this line, decrease in ω leads to $\dot{\omega} > 0$. The area to the left of the vertical line (i.e., locus $\dot{\omega} = 0$) identifies a region in which the potential number of adopters is greater than one. Therefore, this region is not feasible. There exists a downward-sloping saddle path going through the unique balanced-growth equilibrium point B. Thus, I arrive at:

Proposition 3. *Assume that* $\delta > (g_N - \rho)$. *Then, there exists a unique negative-sloping globally stable-saddle-path going through the final unique balanced-growth equilibrium point* B. *Along the saddle path, the measure of industries that adopt the new GPT,* ω, *increases, the per capita consumption expenditure, c, decreases, the market interest rate,* r, *increases, the innovation rate of the industries that have adopted the new GPT,* I_1, *decreases at a higher rate than that of those that have not adopted the new GPT,* I_0. *In addition, there exist transitional growth cycles of per capita GNP.*

Proof. See Appendix.

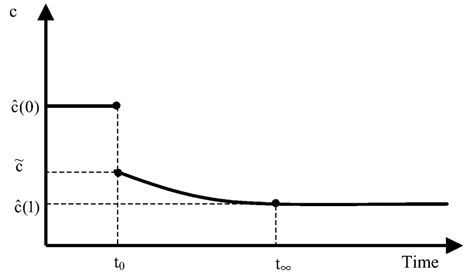


Fig. 3. Time path of the per captia consumption expenditure after a GPT arrives in the economy

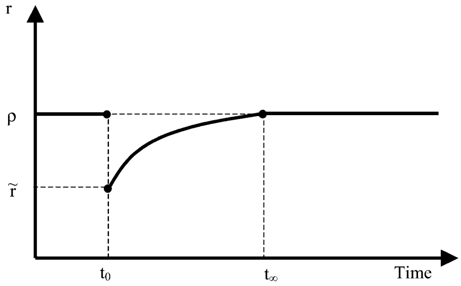


Fig. 4. Time path of the market interest rate after a GPT arrives in the economy

The analysis is predicated on the assumption of perfect foresight.[25] When the new GPT arrives, per capita consumption expenditure, c, jumps down instantaneously to $\tilde{c}$ (point A' in Fig. 2). This per capita consumption expenditure jump lowers the interest rate to $\tilde{r}$ (Fig. 4) since there are more savings available. The downward jumps on the per capita consumption expenditure and on the interest rate imply an upward jump in the innovation rates of both types of industries; those that have adopted the new GPT and those that have not adopted the new GPT ($\tilde{I}_1$ and $\tilde{I}_0$ in Fig. 5).

Figure 1 illustrates that the R&D line R_0R_0 will shift downwards and the resource line N_0N_0 will shift upwards with the arrival of the new GPT resulting in lower per capita consumption expenditure. Going back to Figure 2, the instantaneous decrease in c is reflected by a movement from point A to point A'. The decrease in per capita consumption expenditure leads to a decrease in the market interest rate r (from Eq. (28), which always hold). When the market interest rate r is lower than the subjective rate ρ, per capita consumption expenditure decreases even further, until the market interest rate approaches the subjective discount rate at the new steady state (point B in Figs. 1 and 2). During the transition dynamics (i.e., as the equilibrium moves from point A' to point B in Fig. 2), the interest

---

[25] There also exists a degenerate equilibrium at which the adoption of the new GPT is not completed. Suppose that when a new GPT arrives, every potential consumer expects that no one will decrease his consumption expenditure in order to finance innovation. As a result, it does not pay to decrease consumption expenditure of a single consumer, because the new GPT will never be fully adopted. In this event, the pessimistic expectations are self-fulfilling, and no new GPTs are fully adopted. I do not discuss these types of equilibria in what follows.

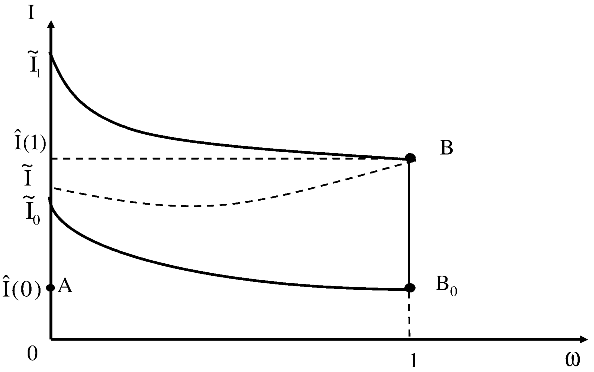


Fig. 5. Evolution of the aggregate investment during the diffusion path: the initial steady-state equilibrium is at point A. There is an upward jump in the aggregate investment with the introduction of the new GPT. One possible path of the aggregate investment, along the diffusion path, is depicted in the figure by the dotted curve $\tilde{I}B$

rate increases leading to more savings and a decrease in per capita consumption expenditure. At point B in Figure 2, all industries have adopted the new GPT.

Along the transition path, the aggregate investment may increase or decrease. One possible path of the aggregate investment is shown in Figure 5 by the dotted curve. There is an upward jump in the innovation rate of industries that have not adopted the new GPT (from point B_0 to point B in Fig. 5).

Figures 3 and 4 show the time paths of per capita consumption expenditure and the market interest rate (where t_0 indicates the time when the new GPT arrives in the economy and t_∞ indicates the time when all industries in the economy have adopted the new GPT). Figure 6 shows the effect of a GPT on the Schumpeterian growth rate. The adoption of the new GPT entails cyclical growth patterns.[26] The growth rate decreases in the initial stages of the adoption of the new GPT. There exist transitional growth cycles.

4.1 Stock market behavior

The fact that the adoption of a new GPT affects positively the productivity of R&D together with free entry into each R&D race is a key factor in explaining the behavior of the stock market. The probability of discovering the next higher quality product in each industry increases with the adoption of the new GPT and so does the probability that the incumbent in each industry will be replaced by a follower firm (i.e., the hazard rate). This link between the GPT adoption and higher risk for incumbent firms captures the effects of creative destruction on the stock market valuation of monopoly profits. In other words, during the diffusion of a GPT, per capita consumption declines, the market interest rate rises, and the hazard

---

[26] Earlier contributions on this issue include the macroeconomic model of Cheng and Dinopoulos (1996) in which Schumpeterian waves obtain as a unique non-steady-state equilibrium solution and the current flow of monopoly profits follows a cyclical evolution.

rate increases. These changes lower the per capita expected discounted profits of the successful innovator and drive down its per capita stock market valuation.[27] Furthermore, the larger the productivity gains associated with the new GPT, the larger the slump of the stock market. For example, it may be that the productivity gains generated by the introduction of the new GPT are large not because the new GPT is technologically very advanced at that initial stage, but because the previous GPTs are particularly inadequate for the needs of these sectors.[28] However, the size of the slump in the stock market is more severe, when the new GPT is diffused at a higher rate.

The aggregate stock value is given by the following equation:

$$\bar{V} = \left[\omega \frac{V_1}{N(t)} + (1-\omega)\frac{V_0}{N(t)}\right] N(t) \, , \tag{30}$$

where V_1 and V_0 are given by Equation (20) after taking account Equation (24). Thus, the growth rate of the aggregate stock value is given by:

$$g_V = \frac{\dot{\bar{V}}}{\bar{V}} = g_N - \frac{(\mu-1)\omega\delta(1-\omega)}{[\omega + (1-\omega)\mu]} \, . \tag{31}$$

At the initial steady-state, where $\omega = 0$ and at the final steady-state, where $\omega = 1$, the growth rate of the aggregate stock value is equal to the rate of the population growth.

That is,

$$\hat{g}_V = g_N \, . \tag{32}$$

The effects of a GPT on the stock market valuation of monopoly profits are summarized in the following proposition:

Proposition 4. *The growth rate of the stock market, g_v, depends negatively on the rate of GPT diffusion process, δ, and the magnitude of the GPT-ridden R&D productivity gains, μ, and positively on the rate of population growth, g_N. It also follows a U-shaped path relative to the population growth rate during the diffusion process of the new GPT.*

Proof. See Appendix.

These comparative properties, which differentiate the model from several others in its class, can be illustrated with the help of Figure 7, which shows the growth

---

[27] Hobijn and Jovanovic (2001) argue that U.S. stock market decline in the early 1970s was due to the arrival of information technology and the fact that the stock-market incumbents were not ready to implement it. They state “Instead, new firms would bring in the new technology after the mid-1980s. Investors foresaw this in the early 1970s and stock prices fell right away.” The U.S. stock market value relative to GDP plummeted to 0.4 in 1973, just after Intel developed the microprocessor in late 1971. The decrease of the stock market value relative to GDP did not recover until the mid-1980s, and then rose sharply. Leading OECD countries also experienced similar movements in their stock markets, following a U-shaped path.

[28] This was clearly the case for early computers, where even though valves had been getting smaller for over a decade prior to the arrival of the transistor, the transistor was still an order of magnitude smaller.

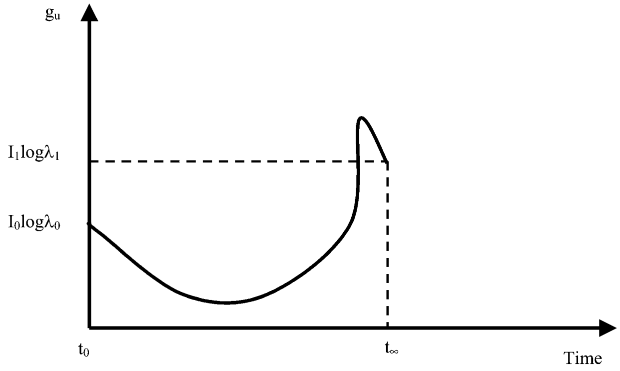


Fig. 6. The effects of a GPT on the Schumpeterian growth rate: when all industries have adopted the new GPT, the economy experiences higher steady-state Schumpeterian growth. There also exist transitional growth cycles of per capita GNP

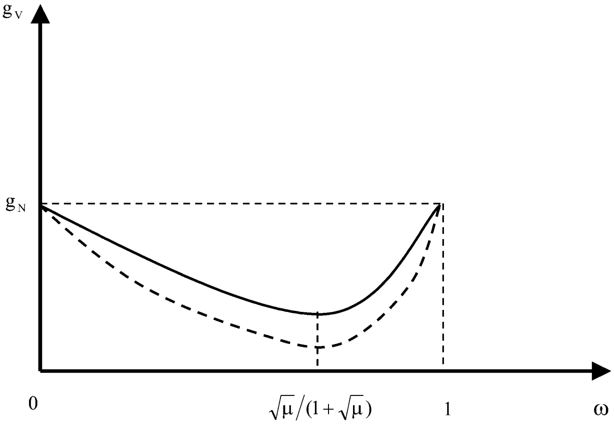


Fig. 7. The effects of a GPT on the stock market: The growth rate of the stock market depends on the rate of GPT diffusion process, the magnitude of the GPT-ridden R&D productivity gains, and the rate of population growth. It also follows a U-shapepd path relative to the population growth

rate of stock market as a function of the measure of industries that have adopted the new GPT. The initial adoption of the new GPT decreases the growth rate of the stock market below the rate of the population growth. In the later stages of the adoption of the new GPT, the growth rate of the stock market increases. When the diffusion process of the new GPT has been completed, the growth rate of the stock market is equal to the rate of the population growth. That is, it follows a U-shaped path relative to the population growth rate during the diffusion process of the new GPT.

This last result can be seen from the second term of the right hand side in Equation (31).[29] The free entry condition in each R&D race (Eq. 20) implies that the per capita stock value in any industry θ, $V_\theta/N(t)$, is constant over time. It jumps down instantaneously with the adoption of the new GPT, and it remains constant thereafter. The aggregate stock value, which increases exponentially at the rate of the population growth, jumps down with the arrival of the new GPT, and then increases again at the population growth rate. The slump in the aggregate stock value is due to the realization of the R&D productivity gains associated with the new GPT. The higher these R&D productivity gains are, the higher is the jump in the per capita industry and aggregate stock value at the time of the adoption of the new GPT.[30]

An increase in the GPT diffusion rate, δ, increases the economywide resources devoted to R&D. Thus, the probability that the incumbent firm will be replaced by a follower firm increases. This can be seen from Equation (18), which gives the value of monopoly profits. In other words, when the GPT diffusion process accelerates, the decrease in per capita consumption expenditure is more severe, and the per capita R&D investment increases. In this case, the U-shaped path of the growth rate of the stock market sags (this is shown by the dotted-shaped curve in Fig. 7). That is, the slope of the curve representing the growth of the stock market gets steeper at the initial stages of the diffusion process of the new GPT and gets flatter at the final stages of the process.[31]

An increase in the productivity gains generated by the new GPT, μ, lowers the cost of discovering the next higher quality product. This, in turn, will affect negatively the stock market valuation of the incumbent firm (see Eq. (20)).

An increase in the rate of population growth, g_N, shifts the U-shaped curve in Figure 7 upwards and increases the growth rate of the stock market.

4.2 Aggregate investment

Proposition 5. *The effect of the GPT diffusion on the aggregate investment during the adoption process is ambiguous.*

Proof. See Appendix.

---

[29] The numerator in Equation (31), which is positive and reflects the slope of a truncated S-curve, is equal to $\dot{\omega}$ times a positive fraction that depends on the magnitude of the GPT-ridden R&D productivity gains and on the number of the industries that have adopted the new GPT.

[30] This can be seen from Equation (31), where the first term on the right-hand side gets smaller when each industry adopts the new GPT relative to the second term of the right-hand side of the same equation.

[31] Hobijn and Jovanovic (2001) provide evidence that the drop in the stock market in the early 1970s was due to the arrival of information technology. Their evidence supports that the information-technology-intensive sectors experienced the largest drop in 1973–1974, reducing the role of the first OPEC shock in explaining the decrease in the stock market. This result is also consistent with the empirical evidence provided by Jovanovic and Rousseau (2001) on how the U.S. economy is affected by new technologies. They show by using 114 years of U.S. stock market data that the growth of the stock market slows down due to the fact that the new entrants of the stock market will find hard to keep up.

The initial steady-state equilibrium is at point A in Figure 5. There is an upward jump in the aggregate investment with the introduction of the new GPT (from $\hat{I}$ to $\tilde{I}$). Along the diffusion path, both innovation rates (I_0 and I_1) decrease until the economy reaches the final steady-state equilibrium point B, where the aggregate investment is higher relative to the initial steady-state equilibrium point A. There is an upward jump in the innovation rate of the industries that have not adopted the new GPT at the final steady state (from point B_0 to B). One possible picture of how the aggregate investment behaves along the diffusion of the new GPT is shown in Figure 5. Along the transition path, the aggregate investment decreases and then increases. In the initial stages of the diffusion process, only a limited number of industries adopt the new GPT (see Eq. (1)). These industries are called the early adopters. As more industries adopt the new GPT, the aggregate investment increases.

5 Concluding remarks

Previous models that have analyzed GPTs exhibit the scale effects property. The present paper analyzed the effects of a GPT on short-run and long-run Schumpeterian growth without scale effects. The absence of growth scale effects and the modeling of the diffusion process through S-curve dynamics generate several novel and interesting results.

First, the long-run growth rate of the economy depends positively on the magnitude of quality innovations. Any policy that affects this magnitude has long-run growth effects. However, the absence of the arrival of a new GPT in the economy does not reduce the long-run growth rate to zero, as in the previous GPTs-based growth models. All the previous R&D-based models that analyze the effects of GPTs exhibit scale effects.

The assumption that the diffusion of the new GPT follows an S-curve generates two steady-state equilibria: one is the initial steady-state before the adoption of the new GPT begins and the other is the final steady-state after the diffusion process of the new GPT has been completed. At the final steady-state relative to the initial steady-state the long-run growth rate is higher, the aggregate investment is higher, the per capita consumption expenditure is lower, and the market interest rate is equal to the subjective discount rate.

The growth rate of the stock market depends negatively on the rate of GPT diffusion process and the magnitude of the GPT-ridden R&D productivity gains, and positively on the rate of population growth. It also follows a U-shaped path relative to the population growth rate during the diffusion process of the new GPT. This is consistent with the empirical evidence provided by Jovanovic and Rousseau (2001), who empirically document that, during times of rapid technological change, the growth of the stock market slows, since the new entrants will grab the most value from previous entrants (because the incumbents will find hard to keep up). Hobijn and Jovanovic (2001) also provide evidence that the drop in the stock market in the early 1970s was due to the arrival of information technology. Their evidence supports the notion that the information-technology-intensive sectors experienced the largest drop in 1973–1974, reducing the role of the first OPEC shock in explaining the decrease in the stock market.

One could also develop a dynamic general equilibrium model to study the effects of a GPT diffusion on a global economy that exhibits endogenous Schumpeterian growth. As in this model, the adoption of a GPT by a particular industry can generate an increase in the productivity of R&D workers, and the magnitude of all future innovations and its diffusion across industries can be governed by S-curve dynamics. The diffusion of the GPT within an industry from one country to the other can occur with a time lag. Under this framework, it would be interesting to analyze the long-run and transitional dynamic effects of a new GPT on trade patterns, product cycles and (transitional) divergence in per capita growth rates between the two countries. This is a fruitful direction for future research.

Appendix

A.1. Proposition 1

Equations (25) and (27) define a unique steady-state equilibrium.

The long-run Schumpeterian growth rate is endogenous and does not exhibit scale effects. This follows from Equation (23).

Solving the expression for I in (25) (by substituting Eq. (25) into Eq. (26) in the main text), substituting it into (23) and differentiating the resulting expression with respect to the appropriate parameter.

At each steady-state, per capita consumption expenditure is constant. That is $\hat{c} = 0$. Then Equation (8) implies that the interest rate, $\hat{r}$, is equal to the subjective discount rate, ρ.

The aggregate stock value is given by: $\bar{V} = [\omega \frac{V_1}{N(t)} + (1 - \omega) \frac{V_0}{N(t)}] N(t)$, where V_1 and V_0 are given by Equation (20) after taking account Equation (24). After substitution of these values into the aggregate stock value and taking logs and derivatives with respect to time, I obtain the growth rate of the aggregate stock value:

$$g_V = \frac{\dot{\bar{V}}}{\bar{V}} = g_N - \frac{(\mu - 1)\omega\delta(1 - \omega)}{[\omega + (1 - \omega)\mu]} \, . \tag{A.1}$$

At the initial steady-state, where $\omega = 0$ and at the final steady-state, where $\omega = 1$, the growth rate of the aggregate stock value is equal to the rate of the population growth. That is $\hat{g}_V = g_N$.

This completes the proof of Proposition 1.

A.2. Proposition 2

Equations (1) and (29) define two loci: one where $\omega = 0$ and one where $\omega = 1$.

Evaluating the aggregate investment (which is given by $I = I_0(1 - \omega) + I_1\omega$) at the two steady-states implies that it is higher at the final steady state relative to the initial steady state.

Substituting Equations (6) and (10) into Equation (4) in the main text (after taking account Eq. (24)), I obtain $\log u(t) = \log c(t) - \log \alpha_Q + \int_0^1 \log \lambda(\theta)^{j(\theta)} d\theta - \int_0^1 \log \lambda(\theta) d\theta$. I invoke the properties of the Poisson distribution to argue that the expected number of improvements is $I_1 t$ (Feller, 1968, p.159) and obtain:

$$\log u(t) = \log c(t) - \log \alpha_Q + \omega(I_1 t - 1) \log \lambda_1 + (1-\omega)(I_0 t - 1) \log \lambda_0 \ . \quad \text{(A.2)}$$

The economy's long-run Schumpeterian growth is defined as the rate of growth of sub-utility $u(t)$, $g_u = \dot{u}(t)/u(t)$. By taking logs and differentiating Equation (A.2) with respect to time and substituting Equation (29) from the main text, I obtain:

$$\begin{aligned} g_u = {} & \frac{\mu(c-1)}{k\alpha_R[\mu - \omega(\mu-1)]} + g_N - \rho + \dot{\omega}[(I_1 t - 1)\log \lambda_1 - (I_0 t - 1)\log \lambda_0] \\ & + \omega(I_1 + t\,\dot{I}_1)\log \lambda_1 + (1-\omega)(I_0 + t\,\dot{I}_0)\log \lambda_0 \ . \end{aligned} \quad \text{(A.3)}$$

Evaluating Equation (A.3) at the two steady-states implies that the long-run Schumpeterian growth rate is higher at the final steady-state than at the initial steady-state.

By setting Equation (29) equal to zero, I obtain the per capita consumption expenditure as a function of the measure of industries that have adopted the new GPT, $c(\omega) = \frac{(\rho - g_N)k\alpha_R[\mu - \omega(\mu-1)]}{\mu} + 1$. Evaluating the per capita consumption expenditure at the two steady-states implies that $\hat{c}(1) < \hat{c}(0)$.

At each steady-state, per capita consumption expenditure is constant. That is $\hat{\dot{c}} = 0$. Then Equation (8) implies that the interest rate, $\hat{r}$, is equal to the subjective discount rate, ρ.

Equations (20) and (13) in the main text imply that $\frac{V(1)}{N} = \frac{V_1}{N} < \frac{V(0)}{N} = \frac{V_0}{N}$.

This completes the proof of Proposition 2.

A.3. Proposition 3

In order to prove that there exists locally a negative sloping saddle path, I use a polynomial of order one to linearize the nonlinear differential Equations (1) and (29) around their steady-state values ($\hat{c} = \frac{(\rho - g_N)k\alpha_R}{\mu} + 1, \hat{\omega} = 1$). Equations (31) and (1) in the main text can be written as $\dot{c} = a_1 c + b_1 \omega$ and $\dot{\omega} = b_2 \omega$, respectively (where $a_1 = \frac{[(\rho - g_N)k\alpha_R + \mu]}{k\alpha_R}$, $b_1 = \frac{[(\rho - g_N)k\alpha_R + \mu][(\rho - g_N)k\alpha_R(\mu - 1)]}{\mu k \alpha_R}$ and $b_2 = -\delta$).

Suppose $c = Ae^{rt}$, $\omega = Be^{rt}$ are the particular solutions to this homogeneous system $(\dot{c}, \dot{\omega})$. I can substitute these proposed solutions into this system and I can write it in matrix notation as follows:

$$\begin{bmatrix} a_1 - r & b_1 \\ 0 & b_2 - r \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} . \quad \text{(A.4)}$$

In order that the solution to (A.4) be other than $A = B = 0$, the coefficient matrix in (A.4) must be singular. Thus, I obtain a quadratic equation in $r : r^2 - r(a_1 + b_2) + a_1 b_2 = 0$. Computing the roots of this quadratic equation, I obtain:

$$\begin{aligned} r_1 &= \frac{(a_1 + b_2) + \sqrt{(a_1 + b_2)^2 - 4a_1 b_2}}{2} \\ r_2 &= \frac{(a_1 + b_2) - \sqrt{(a_1 + b_2)^2 - 4a_1 b_2}}{2}\,. \end{aligned} \tag{A.5}$$

Since $a_1 > 0$ and $b_2 < 0$, the roots are real. Because the argument of the square root function exceeds $(a_1 + b_2)^2$, the smaller of the roots is negative. The larger root is positive. Hence, the roots are real and of opposite sign; the stationary point is a saddlepoint.

In order to prove that there exists globally a negative sloping saddle path, I use the time elimination method. The slope of the policy function can be obtained by taking the ratio of the two differential equations that govern the dynamic behavior of the economy:

$$\frac{\partial c}{\partial \omega} = c'(\omega) = \frac{\dot{c}}{\dot{\omega}} = \frac{\frac{[c^2(\omega)\mu - c(\omega)\mu]}{k\alpha_R[\mu - \omega(\mu - 1)]} + c(\omega)(g_N - \rho)}{\delta\omega - \delta\omega^2}\,. \tag{A.6}$$

Time does not appear in the above equation. To solve this equation numerically, there must be one boundary condition; that is, one point, (c, ω), that lies on the stable arm. Although the initial pair, $[c(0), \omega(0)]$, is unknown, the policy function goes through the steady state $(\hat{c}, \hat{\omega})$.

The slope of the policy function at the steady state is $c'(\hat{\omega}) = \frac{\dot{\hat{c}}}{\dot{\hat{\omega}}} = \frac{0}{0}$, which is indeterminate. Applying the L'Hôpital's rule to this slope and evaluating it at the steady state values, I obtain:

$$c'(\hat{\omega}) = \frac{[1 - c(\hat{\omega})](\mu - 1)c(\hat{\omega})}{\{\mu[2c(\hat{\omega}) - 1] + (\delta + g_N - \rho)k\alpha_R\}}\,. \tag{A.7}$$

From the phase diagram in Figure 2, I can sign the following expressions: $[2c(\hat{\omega}) - 1] > 0$ and $[1 - c(\hat{\omega})] < 0$. If $\delta > (g_N - \rho)$, then the expression in (A.7) is negative. Thus, if the rate of diffusion is high enough (higher than the effective discount rate), the slope of the stable arm is negative.

Along the saddle path, the measure of industries that adopt the new GPT evolves according to Equation (1). As ω increases, Equation (A.7) implies that per capita consumption expenditure decreases. Along the saddle path, $\dot{c}/c < 0$, which implies that the market interest rate, r, is lower than the subjective interest rate, ρ (from Eq. (8) in the main text). Once the diffusion process is about to complete and the economy approaches the new steady state, the market interest rate increases to become equal with the constant subjective interest rate at the new steady state.

In order to analyze the behavior of the two innovation rates along the diffusion process, I differentiate equations them with respect to ω. By taking the difference between these innovation rates, I obtain:

$$\frac{\partial I_1}{\partial \omega} - \frac{\partial I_0}{\partial \omega} = \left(\frac{\mu(\lambda_1 - 1)}{k\alpha_R \lambda_1} - \frac{(\lambda_0 - 1)}{k\alpha_R \lambda_0} \right) \frac{\partial c}{\partial \omega} < 0 \, . \tag{A.8}$$

That is, along the diffusion process, the innovation rate for the industries that have adopted the new GPT decreases more than that of the industries that have not adopted the new GPT.

In terms of the growth rate, differentiating Equation (A.3) with respect to time and taking the limit when the economy approaches the two steady-states, I obtain: $\lim_{\omega \to 0}[\frac{\partial g_u}{\partial t}] = \frac{\dot{c}}{k\alpha_R} + \ddot{I}_0 t \log \lambda_0 + 2\dot{I}_0 \log \lambda_0 < 0$ and $\lim_{\omega \to 1}[\frac{\partial g_u}{\partial t}] = \frac{\mu \dot{c}}{k\alpha_R} + \ddot{I}_1 t \log \lambda_1 + 2\dot{I}_1 \log \lambda_1 < 0$. Since the signs of these equations are negative, the growth rate decreases both in the initial stage and towards the final stage of the diffusion process. Since the quantity of the first equation is smaller than that of the second equation in absolute value, the growth rate decreases more towards the final stage than in the initial stage.

A.4. Proposition 4

The growth rate of the stock market is given by Equation (A.1). From Equation (A.1), it is obvious that the growth rate of the stock market depends on the rate of GPT diffusion process, δ, the magnitude of the GPT-ridden R&D productivity gains, μ, and the rate of population growth, g_N.

By differentiating Equation (A.1) with respect to ω, I obtain: $\frac{\partial g_V}{\partial \omega} = \frac{\delta(1-\mu)(k\alpha_R)^2[(1-\omega)^2\mu - \omega^2]}{[\omega k\alpha_R + (1-\omega)k\alpha_R \mu]^2}$. The sing of this expression depends on the sign of the expression $[(1-\omega)^2\mu - \omega^2]$. When $\omega < \sqrt{\mu}/(1+\sqrt{\mu})$, the growth rate of the aggregate stock value decreases. When $\omega > \sqrt{\mu}/(1+\sqrt{\mu})$, the growth rate of the aggregate stock value increases. That is, along the diffusion process, the growth rate of the aggregate stock value follows a U-shaped path.

Differentiating Equation (A.1) with respect to the diffusion rate, δ, I obtain:

$$\frac{\partial g_V}{\partial \delta} = \frac{\omega(1-\omega)(1-\mu)}{[1 + (1-\omega)\mu]} < 0 \, . \tag{A.9}$$

Equation (A.9) implies that when the diffusion rate increases, the growth rate of the aggregate stock market decreases.

Differentiating expression $\frac{\partial g_V}{\partial \omega}$ with respect to the diffusion rate, δ, I obtain the expression $\frac{\partial(\partial g_V/\partial \omega)}{\partial \delta} = \frac{(1-\mu)[(1-\omega)^2\mu - \omega^2]}{[\omega + (1-\omega)\mu]^2}$, which sign depends on the sign of the expression $[(1-\omega)^2\mu - \omega^2]$. When the rate of the diffusion process of the new GPT increases, the slope of the growth of the stock market increases (i.e., gets steeper) for $\omega < \sqrt{\mu}/(1+\sqrt{\mu})$ and decreases (i.e., gets for $\omega > \sqrt{\mu}/(1+\sqrt{\mu})$.

By differentiating Equation (A.1) with respect to μ, I obtain:

$$\frac{\partial g_V}{\partial \mu} = -\frac{\omega\delta(1-\omega)}{[\omega + (1-\omega)\mu]^2} < 0 \, . \tag{A.10}$$

The sign of Equation (A.10) implies that the growth rate of the stock market decreases when the productivity gains from the new GPT are larger.

By differentiating Equation (A.1) with respect to g_N, I obtain:

$$\frac{\partial g_v}{\partial g_N} = 1 > 0\ . \tag{A.11}$$

The sign of Equation (A.11) implies that the growth rate of the stock market increases at the rate of population growth.

A.5. Proposition 5

Differentiating the equation that expresses the aggregate investment with respect to the measure of industries that adopt the new GPT, I obtain:

$$\frac{\partial I}{\partial \omega} = \frac{\partial I_0}{\partial \omega}(1-\omega) + \frac{\partial I_1}{\partial \omega}\omega + (I_1 - I_0)\ . \tag{A.12}$$

The first two terms in Equation (A.12) are negative. The third term is positive. Thus, the aggregate investment will decrease or increase during the diffusion process depending on the magnitude of these signs.

The first two terms are the slopes of the solid curves depicted in Figure 7. These slopes depend on the rate of diffusion of the new GPT. By contrast, the third term of Equation (A.12) does not depend on the rate of diffusion of the new GPT. Thus, when the rate of diffusion of the new GPT increases and more industries switch to the new GPT faster, the negative effect dominates the positive effect and the aggregate investment decreases.

$$\lim_{\omega \to 1} I_1(\omega) = \frac{\mu(\lambda_1 - 1)}{k\alpha_R \lambda_1} c(\omega) + g_N - \rho\ . \tag{A.13}$$

$$\lim_{\omega \to 1} I_0(\omega) = \frac{(\lambda_0 - 1)}{k\alpha_R \lambda_0} c(\omega) + g_N - \rho\ . \tag{A.14}$$

Equations (A.13) and (A.14) imply that when the measure of industries that adopt the new GPT approaches the final steady-state, the innovation rates of the industries that have adopted the new GPT and the innovation rate of industries that have not adopted the new GPT are not the same. That implies that the innovation rate of the industries that have not adopted the new GPT jumps upward (i.e., when the last industry switches to the new GPT, I_0 becomes I_1).

References

Aghion P, Howitt P (1998a) Endogenous growth theory. MIT Press, Cambridge, MA

Aghion P, Howitt P (1998b) On the macroeconomic effects of major technological change. In: Helpman E, (eds) General purpose technologies and economic growth. MIT Press, Cambridge, MA

Andersen B, (1999) The hunt for s-shaped growth paths in technological innovation: a patent study. Journal of Evolutionary Economics 9: 487–526

Arroyo C, Dinopoulos E, Donald S (1995) Schumpeterian growth and capital accumulation: theory and evidence. Mimeo

Barro R, Sala-I-Martin X (1995) Economic growth. McGraw-Hill, New York

Bresnahan T, Trajtenberg M (1995) General purpose technologies: Engines of growth? Journal of Econometrics 65: 83–108

Cheng L, Dinopoulos E (1996) A multisectoral general equilibrium model of Schumpeterian growth and fluctuations. Journal of Economic Dynamics & Control 20: 905–923

David P (1990) The dynamo and the computer: a historical perspective on the modern productivity paradox. American Economic Review 80: 355–361

Dinopoulos E (1994) Schumpeterian growth theory: an overview. Osaka City University Economic Review 29: 1–21

Dinopoulos E, Segerstrom P (1999) A schumpeterian model of protection and relative wages. American Economic Review 89: 450–472

Dinopoulos E, Sener M (2003) New directions in Schumpeterian growth theory. University of Florida. Mimeo

Dinopoulos E, Syropoulos C (2000) Innovation and rent protection in the theory of schumpeterian growth. University of Florida. Mimeo

Dinopoulos E, Thompson P (1996) A contribution to the empirics of endogenous growth Eastern Economic Journal 22: 389–400

Dinopoulos E, Thompson P (1998) Schumpeterian growth without scale effects. Journal of Economic Growth 3: 313–335

Dinopoulos E, Thompson P (1999) Scale effects in schumpeterian models of economic growth. Journal of Evolutionary Economics 9: 157–185

Eriksson C, Lindh T (2000) Growth cycles with technology shifts and externalities. Economic Modelling 17: 139–170

Griliches Z (1957) Hybrid corn: an exploration in the economics of technological change. Econometrica 25: 501–522

Grossman, G, Helpman E (1991) Quality ladders in the theory of growth. Review of Economic Studies 59: 43–61

Helpman E (1998) General purpose technologies and economic growth. MIT Press, Cambridge, MA

Helpman E, Rangel A (1998) Adjusting to a new technology: experience and training. NBER Working Paper 6551

Helpman E, Trajtenberg M (1998a) A time to sow and a time to reap: growth: based on general purpose technologies. In: Helpman E (eds) General purpose technologies and economic growth. MIT Press, Cambridge, MA

Helpman E, Trajtenberg M (1998b) Diffusion of general purpose technologies. In: Helpman E (eds) General purpose technologies and economic growth. MIT Press, Cambridge, MA

Hobijn B, Jovanovic B (2001) The information-technology revolution and the stock market: evidence. American Economic Review 91: 1203–1220

Jones C (1995a) Time-series tests of endogenous growth models. Quarterly Journal of Economics 110: 495–525

Jones C (1995b) R&D-based models of economic growth. Journal of Political Economy 103: 759–784

Jorgenson D (2001) Information technology and the U.S. economy. American Economic Review 91: 1–32

Jovanovic B, Rousseau P (2001) Vintage organization capital. NBER Working Paper 8166

Mulligan CB, Sala-I-Martin X (1991) A note on the time-elimination method for solving recursive economic models. NBER Working Paper 116

Mulligan CB, Sala-I-Martin X (1992) Transition dynamics in two-sector models of endogenous growth. Quarterly Journal of Economics 107: 739–773

Peretto P (1998) Technological change and population growth. Journal of Economic Growth 3: 283–311

Peretto P, Smulders S (1998) Specialization knowledge dilution, and scale effects in an io-based growth model. Center For Economic Research, Tilberg University, Paper No. 9082

Schumpeter JA (1934) The theory of economic development. Harvard University Press, Cambridge, MA

Segerstrom P (1998) Endogenous growth without scale effects. American Economic Review 88: 1290–1310

Segerstrom P, Zolnierek J (1999) The R&D incentives of industry leaders. International Economic Review 40: 745–766

Thirtly C, Ruttan V (1987) The role of demand and supply in the generation and diffusion of technical change. Harwood Academic Publishers, Switzerland

Young A (1998) Growth without scale effects. Journal of Political Economy 106: 41–63

Zachariadis M (2003) R&D innovation, and technological progress: a test of the Schumpeterian framework without scale effects. Canadian Journal of Economics (forthcoming)