India Studies in Business and Economics

Ambar Nath Ghosh
Asim K. Karmakar *Editors*

Analytical Issues in Trade, Development and Finance

Essays in Honour of Biswajit Chatterjee

Springer

India Studies in Business and Economics

For further volumes:
http://www.springer.com/series/11234

Ambar Nath Ghosh • Asim K. Karmakar
Editors

Analytical Issues in Trade, Development and Finance

Essays in Honour of Biswajit Chatterjee

Springer

Editors
Ambar Nath Ghosh
Department of Economics
Jadavpur University
Kolkata
West Bengal
India

Asim K. Karmakar
Department of Economics
Jadavpur University
Kolkata
West Bengal
India

ISSN 2198-0012 ISSN 2198-0020 (electronic)
ISBN 978-81-322-1649-0 ISBN 978-81-322-1650-6 (eBook)
DOI 10.1007/978-81-322-1650-6
Springer New Delhi Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014931400

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Analytical Issues in Trade, Development and Finance: Essays in Honour of Biswajit Chatterjee as a reference book for researchers, discusses some contemporary issues in Economics in the earlier mentioned areas. There are three major sections of the book: International Trade, Development Economics, and Macroeconomics and Finance. The papers in the first section address analytical issues relating to trade-environment linkage, capital accumulation for pollution abatement, possibility of technology diffusion by multinational corporations, nature of innovation inducing tariff protection, effects of import restriction and child labour, the linkage between exchange rate, direction of trade and financial crisis—the implications for India and on global economic crisis, financial institutions and global capital flows and Balance of Payment (BoP) imbalances. The second section consists of discussions on why widespread poverty persists in South Asia, development dividend associated with peace in South Asia, issues in well-being and human development, implications for endogenous growth through human capital accumulation on environmental quality and taxation, the rationale for the labour supply schedule for the poor, switching as an investment strategy, role of government and the strategic interaction in the presence of information asymmetry, government's role in controlling food inflation, the inter-state variations in levels and growth of industry in India, structural break in India's service sector development, and the phenomenon of wasted votes in the Parliamentary Elections of India—all refer to important frontier issues in the political economy of development in Less Developed Countries (LDCs) in general and India in particular. The third section on *Macroeconomics and Finance* deals with the effectiveness of monetary policy in tackling economic crisis, effective demand model of corporate leverages and recession, the empirical link between stock market development and economic growth in cross-country experience in Asia, an empirical verification of Mckinnon–Shaw hypothesis for financial development in India, the dynamics of the behaviour of the Indian stock market, efficiency of Non-Life Insurance Companies, econometric study of the causal linkage between FDI and current account balance in India and implications of contagious crises for the Indian economy. All together the 30 papers in the volume bring into focus the frontier issues in the domain of international trade, economic development, macroeconomics and finance.

Ambar Nath Ghosh
Asim K. Karmakar

Introduction

In the present volume, we have collected a number of papers from among this large set of colleagues, well-wishers and students, from home and abroad, who wish to record their intellectual debt to Professor Chatterjee, or their appreciation of his contributions in economics on the occasion of his attaining the age of 60 years in March 2014.

The papers presented in this edited volume deal broadly with three areas of economics: International trade, Development, and Macroeconomics and Finance. Within each of these areas, the individual papers have dealt with a wide variety of topics.

The first group of papers is devoted to international trade.

Section I: International trade

In his paper *Trade-Environment Linkage: A South-centric Model-specific Analysis,* Mallinath Mukherjee has shown that comparative advantage of a country in the production of polluting goods triggers trade-induced resource flows and raises doubts about the gains from free trade. The under-developed South may experience a seriously endangered environment because of their poorly defined property rights over environmental resources. Trade opportunities are generated due to different environmental standards. More trade and less pollution make the North unambiguously better-off; more trade but more pollution in the South cast doubts on globalization. However, more trade and the need to protect the environment have been at the root.

His paper is an attempt to capture these three possibilities in the context of a Small Open Economy. The North is assumed to have a perfectly elastic demand for exports from the South and a perfectly elastic supply of South's imports—both at the global prices. The contention of the model is that the countries of the South are seriously challenged by the developments of the last two decades in this respect. While each nation is unique in its own right, the model tries to isolate three fundamental developments, namely, growing food insecurity, increasing wage-gap and the possibility of exportable land services. Each issue is challenging and, therefore, the ongoing trade-environment debate would not become comprehensive unless such issues are identified and addressed by proper policy interventions.

The second paper in this section is *Accumulation of Foreign Capital for Pollution Abatement and Immiserizing Growth—A Theoretical Result for Developing*

Economies by Rupamanajari Sinha Ray. In this paper she articulates that various forms of externalities have opened up the cases for different forms of domestic distortions and the trade interventions and environmental interventions reflecting alternative forms of lobbying by the interested group, and the scope for devising an optimum framework for resolving such conflicts. In an open economy, when growth is driven by factor accumulation, the economy gains by way of expansion of domestic income at constant terms of trade, but the post-growth equilibrium international terms of trade may go against the growing nation if its growth is concentrated heavily in its exportable sector (ultra export-biased growth), and the level of national welfare of the growing economy in the post-growth scenario may turn out to be less than its pre-growth scenario, if the loss in terms of trade is stronger than the income gains due to growth for such an economy. This is the phenomenon of *immiserizing growth.* In this paper, she uses a two-sector general equilibrium framework for an open economy with environmental pollution acting as an externality to examine the conditions of immiserizing growth in developing countries under two situations: (a) domestic capital accumulation for pollution abatement purposes; and (b) foreign capital accumulation for pollution abatement purposes.

In their joint paper: *Optimal Entry Mode for Multinationals with Possibility of Technology Diffusion,* Nilanjana Mitra and Tanmoyee Banerjee (Chatterjee) evaluate the optimal entry mode of a multinational company (MNC) that is choosing among export, fragmented production structure with assembly line FDI in LDC or complete production in LDC with FDI. The results show that if the Intellectual Property Rights (IPR) restriction is strong and plant installation cost in LDC is below a certain critical level, then the MNC chooses complete LDC production with FDI over assembly line FDI, where the model assumes that a fake producer can copy the product if complete production takes place in LDC. In such a situation, government will choose to protect IPR only under some sufficient conditions; otherwise no monitoring is the optimal strategy of the government and MNC will choose the strategy of fragmented production structure and assembly line FDI will take place in LDC.

The fourth paper in this section *An Example of Innovative-Inducing Tariff Protection*, by Swapnendu Banerjee, illustrates how in a vertically differentiated domestic market with two types of consumers, tariff protection raises the incentive for quality innovation by a domestic monopolist. Similar in spirit of the Schumpeterian idea, this result is in sharp contrast to the oft-quoted argument that trade liberalization is conducive to quality innovation.

Runa Roy in her paper entitled *Import Restriction, Capital Accumulation and Use of Child Labour—A General Equilibrium Analysis* has developed a general equilibrium framework for a less developed economy using child labour in one of its sectors, and examines the effects of policies of import restriction and of domestic capital accumulation on the incidence or use of child labour in such an economy. The paper concludes that child labour problem in a less developed economy can be reduced by adopting policies favourable to economic growth through capital accumulation or by adopting restrictive trade policy in the sector where child labour is not used. One is directly altering the use of child labour; another is through changes in factor process. Thus, quantity effects and price effects are both important.

Rajendra Narayan Paramanik and Bandi Kamaiah in their paper entitled *Direction of Trade, Exchange Rate Regimes and Financial Crises: The Indian Case*, focus on the direction of trade of India with its selected major 25 trade partners during the years 1997–1998 to 2009–2010 with the help of gravity model of trade. To address the issues like heterogeneous impact of the trading partners on the trade volume of India, panel corrected standard error (PCSE) model has been used. Apart from the influence of traditional factors like GDP (gross domestic product), population, spatial distance (proxy for transaction cost), significant impacts of the financial crises and exchange rate regimes of the trading partners are tested. The expected positive roles of GDP and population of India as well as its trade partners have been vindicated. However, the negative impact of distance is not statistically discernible. Two major financial crises (Asian crisis in 1998 and the recent global meltdown in 2008) that occurred during this phase played havoc in terms of its impact on India's trade relation. Though in absolute terms advanced nations with freely floating exchange rate system trade more with India but proportional impact of the nations with fixed exchange peg or exchange rate arrangement with no legal tender has been found to be more prominent.

Dilip M. Nachane in his paper entitled *Global Crisis, Financial Institutions and Reforms: An Indian Perspective* has raised a lively controversy and sharp divisions among schools of economic thought regarding the relationship between finance and economic growth. He has pointed out that important areas of disagreement persist viz. the type of financial system most conducive to growth, private versus public ownership of financial institutions, the degree of regulation and supervision, the role of financial innovations and the pace and extent of financial liberalization. The Latin American crises of the 1980s and 1990s, the Asian financial crisis and the current global recession have once again brought the critical role of financial institutions under the scanner, and introduced some important caveats to the consensus. It is in this context that his paper aims to take stock of some of these issues in the Indian context. While it is certainly not being claimed that the Indian experience is representative of the entire South Asian region, it is nevertheless felt that some of the lessons drawn here, would have some relevance transcending their immediate context.

Rameshwar Tandon and Shariq Mohd. in their paper *Global Capital Flows & Payments Imbalances*, voice that the current financial and economic crisis is not a single phenomenon, with a single cause. Whether recession or depression, "the present crisis is a crisis of the capitalist system, like the East Asian crisis a decade ago or the Japanese crisis before that and it is the first crisis of the system since the 1930s." In the Anglo-American heartland, it began and remained until September 2008, primarily a financial crisis, though with "real economy" causes (a structural deficit in the production of tradable goods and services), and also 'real' economy effects. In Japan, Germany and much of the periphery, it began later, primarily as an export crisis, in response to slow-down in the Anglo-American heartland, reflecting a structural surplus capacity to produce tradable goods and services but the export crisis then fed through finance and wider growth problems.

The explosion of international financial intermediation after the 1980s and the rising incidence of financial crisis, with cross-border affects, were obviously related. For policy makers, the principal question was what it has long been—when 'real' economy growth rates were sought in excess of those capable of being generated by domestic savings, how were the benefits and costs of financial openness to be distributed? In principle, inward flows of privately owned capital make it possible for real economies to grow more rapidly than if they rely solely on domestic resources. In practice, the extra costs associated with crisis—induced capital outflows, bailouts, and the lost confidence of investors occasionally, threaten to undermine the real economies, and set back the process of industrialization and disrupt underlining political and social order. The current crisis shows that the basic premise of the traditional risk management theory is wrong and that financial markets indeed can be inherently unstable, especially due to their increasing complexity.

The second group of papers is devoted to Development Economics.

Section II: Development

In this section the first paper entitled *Widespread Poverty Amidst High Economic Growth: Some Lessons from South Asia* begins with G. K. Chadha's assessment on South Asia's formidable economic agenda to be pursued to conquer poverty. He opines that there is no scope, whatsoever, for complacency on the growth front; it has to grow faster in the coming times. Then, the pattern of growth has also to change: agriculture has to be accorded its due importance, most ostensibly, because the poverty alleviation effects of higher agricultural growth are proven to be far more pronounced, and enduring, than those emanating from non-farm growth that is typically selective in offering employment and is hugely disparate in wage rates and per person earning capabilities. There are many studies, for example in India, to show that "aggregate GDP growth matters in poverty alleviation but agricultural growth matters more despite a sharp reduction in its contribution to GDP"; more pointedly, "the elasticity of the headcount poverty ratio, measured at the cut-off point of $ 1.25 (2005, PPP), with respect to agricultural value added is nearly twice as high as that of GDP (both on a per capita basis)." Taking cognizance of South Asia's economic realities, the most effective medium-term strategy of making a marked dent into poverty, especially rural poverty, is to ensure a high, and sustained, growth of agriculture.

Sucha Singh Gill's paper *Development Dividend of Peace: Experience of South Asia* attempts to understand the long-term dynamism of development in relation to peace and security in South Asia. This is organized into Five Sections. The Sect. I introduces the theme and Sect. II relates the process of economic development with peace and security. Section III attempts to look introspectively at historical experience of this region over long period of colonial domination. The Sect. IV makes an attempt to analyze the post colonial experience of peace and development among SAARC countries. The last section of the paper provides summing up and attempts the framework for peace, security and development in this region.

Human development is defined as the search for the expansion of the real human freedoms of people to pursue lives that they value and have reason to value. The Human Development Index (HDI) was introduced as a way of measuring such progress.

To obtain a full picture of the evolution of human development, we must go beyond the dimensions in the HDI. Significant aggregate progress in health, education and income is qualified by high and persistent inequality, unsustainable production patterns and disempowerment of large groups of people around the world.

In this backdrop P. K. Chaubey's paper *Well-Being in Human Development Framework: Constituents and Aggregation* stresses that welfare economics ignored for long capability dimension of development and development economics did not pay sufficient attention to the softer side of human progress. The primary specification of a person's well-being is in terms of a functioning vector. It can be converted into a scalar measure of well-being only through a real-valued 'valuation function' $v(.)$, which maps functioning vectors into numerical representations. This valuation may not even be complete. But the most difficult part would be to judge well-being level of social states, which involve (1) comparisons over time or across space and (2) preferences of different people. Chaubey argues that although the whole capability approach concentrates on the individual, numerical presentation has exclusively been made only at an aggregate level. Aggregates may be small groups, communities, nations. All these elements have to be aggregated. There could be two ways. One could be that we attempt aggregation of each individual's vectors into their respective well-beings and then individuals' well-beings into society's well-being. The other could be that we attempt to aggregate a single being/doing of all individuals (like literacy or life expectancy) and then these aggregated social states of beings/doings into society's well-being. This purports to converting each row vector into a scalar and thus first creating a column vector of n-size and then converting this column vector into a scalar number. The Human Development Reports have attempted the second course. For example, health-beings of all individuals were first converted into life-expectancy of the society and literacy capabilities of all individuals were first converted into literacy level of the society. The same could be said about enrolment ratios or schooling years but not so clearly about per capita GDP. Then, these societal components of functionings (beings and doings) have been aggregated into Human Development index in a linear/geometric fashion.

The fourth paper in this section is by Bidisha Chakraborty and Manash Ranjan Gupta. In their paper *Human Capital Accumulation, Environmental Quality, Taxation and Endogenous Growth*, they consider a Lucas (1988) type model of endogenous growth in which the environmental quality positively affects the rate of human capital accumulation and the environmental quality itself is positively affected by human capital accumulation and is negatively affected by physical capital accumulation. Thereafter, they analyse the effects of taxation on the steady state equilibrium growth rate in this model. They also analyse the transitional dynamic properties of this model.

Pradipta Chaudhury's paper on *Labour Supply Schedule of the Poor: A Commonsense Approach* emphasizes the dissimilarities between the realities of poor societies and the economic conditions analysed in the text books while drawing a labour supply schedule. There are three fundamental differences: (a) In poor societies, the family is the unit of labour supply, not an individual. (b) The first objective of a poor family is to meet its subsistence needs by selling labour. This gives rise to the idea of a target income, which in turn leads to a rectangular hyperbola shape of the family's labour supply curve. (c) The downward slope of the labour supply schedule is for low wages, unlike the backward bending part of a labour supply curve in a standard textbook, which is for high wages.

He then pinpoints that there are some interesting implications of these three ideas: the first two lead to a typical sequence in the supply of labourers; usually, in a poor family first the male, then the female and then the child join the labour force. Consequently, we observe a corresponding wage differential, male wage rate is higher than female wage rate which in turn is higher than the child wage rate, even when there is no difference in the productivity. Under such conditions, a rise in the wage rate, or an increase in employment opportunities for adults, caused either by economic change or by public policy and funding leads to a decline in supply of labour, particularly child labour. This may be unwelcome for the employers who are enamoured by abundance of cheap labour. But more and regular adult employment and better wages will improve education and nutrition which will lead to increases in skills and efficiency.

He also points out to some conspicuous facts suggesting the shape of the labour supply schedule of the poor are discussed. In the third section the labour supply schedule is derived from these facts and commonsense; several implications of the curve are drawn. In the fourth section our discussion is extended to the issue of differential wage rates of male, female and child labour in a traditional or backward economy. The last section summarizes the central message of the paper.

The sixth paper in this section is *Switching as an Investment Strategy: Revisiting Parrondo's Paradox* by Avik Chakraborti highlights that randomly switching investments between assets with negative expected returns can, indeed, yield positive expected returns. The analytical apparatus, in line with those adapted to Parrondo's paradox, is based on stochastic properties of discrete-time Markov chains. An intuitive explanation of the result is provided in terms of Brownian ratchets.

Information asymmetry poses several problems in the world. Sometimes Government also plays a role in distorting the symmetry of the country. It is in this context that Somdeep Chatterjee and Asim K. Karmakar in their paper *Asymmetric Information, Non-cooperative Games and Impatient Agents: Modelling the Failure of Environmental Awareness Campaigns* outline the importance of how a market should be allowed to function free of government intervention and given the presence of adequately conscious agents, such a framework suffices to run an economy. The paper takes up the example of environment and awareness programme related to it. Firstly, the cause of failure of such a programme is identified and then a simple interactive market mechanism is designed without any central interference such that the same can be overcome, thereby highlighting the basic theoretical premise that is being addressed underlying the title of this paper.

The eighth paper in this section is *Government's Role in Controlling Food Inflation* by Hiranya Lahiri and Ambar Nath Ghosh. They opine that the major driver of recent food inflation in India has been vegetables, pulses and oilseeds for which there is no public procurement. Their paper aims at modelling the behaviour of big retailers or middlemen who hoard such perishable commodities and add to food inflation by creating artificial shortages due to speculative hoarding. They show the adverse impact of speculative buffering on average price. Lastly, they argue that import of food items and execution of open market sale by the government will help reduce inflation not only by bridging the supply gap, but also by reducing speculative buffering. The paper also shows how operation of PDS not only brings down food inflation in case of a supply shock, but also regulates behaviour of middlemen.

T. S. Papola in his paper *Interstate Variations in Levels & Growth of Industry: Trends During the Last Three Decades* articulates that inter-regional disparity in the levels of economic development and per capita income has been a major issue in development debate and policy in India. There are large variations in the different indicators of development among the states which finally get reflected in the differences in per capita incomes and levels of living. There have, of course, been changes in extent of disparities and in the relative positions of different states over the years.

He opines that it is primarily the level of industrialisation and growth of industry that determine the relative levels of economic development of different regions. For, development of agriculture is primarily dependent on the quantity and quality of land which is more or less given, and, growth of services mostly follows the growth of agriculture and industry. It is for this reason that most policy instruments for balanced regional development such as investment licences and fiscal and financial incentives that have been adopted in India have been directed towards industry, with the over objective of "industrial development of backward areas". It is, therefore, interesting to study the pattern of industrial growth in the post-reform period when most of the "interventionist" measures have been removed in comparison with the pre-reform period when they were in place. In this context, this paper looks at the changes in the levels of industrialisation, rates of industrial growth and shares of different states in all-India industrial output and employment. In the process it also examines whether rates of industrial growth and changes in the levels of industrialisation have gone together with GSDP growth rates of different states. His paper also makes an attempt to examine the factors that have led to differences in the rates of industrial growth, particularly, in the more recent period. It may be noted that 'industry' is confined to 'manufacturing', in this paper.

Purba Roy Choudhury in her paper *Unit Root and Structural Break: Experience from the Indian Service Sector* tries to analyze the trends in the service sector growth in India from 1950–2010. This paper is organized in the following way. Section I undertakes a selective survey of literature on the growth of services and employment and its role in the process of economic development with reference to India. Section II tries to analyze the relative share of agriculture, industry, services in the GDP of the Indian economy as a whole, along with a decomposition of the subsectors of service sector in India. Section III also takes into account the econometric methodology of the unit root properties of time series data trying to define the different tests of unit root with or without break. The breakpoints are estimated with the help of the Zivot Andrews test. Section IV also takes into account the empirical results and its implications. Section V concludes the study.

Prakash Singh and N. R. Bhanumurthy in their paper: *Infrastructure Development and Regional Growth in India* endeavour to understand the trends and determinants of economic growth in Indian states. For this, it considers two important determinants such as infrastructure and financial development. With the help of panel time series models, the study concludes that although both the variables are highly correlated with economic growth, it is the social sector development that is having higher impact on the economic growth. In terms of the role of financial sector, the results show that although it is necessary to have development in terms of increase in number of bank branches, it is the extent of bank business that is more important in the growth process.

The last paper in this section is Sanmitra Ghosh's paper on *The Phenomenon of Wasted Vote in the Parliamentary Election of India*. His paper takes up the phenomenon of strategic voting in the Indian context, and tries to find empirical support for predictions of *Duverger's Law*: the volume of *wasted votes* should be very low in large elections. The paper finds an interesting fact: while the volume of wasted votes is quite large, much of it is ineffective in terms of its being pivotal in changing the outcome of an election. This study also identifies the importance of ethnic heterogeneity as a determinant of the volume of wasted votes. An apparent puzzle is also identified: when elections between top two candidates are close, the volumes of wasted votes tend to be larger, which contradicts a prediction of calculus of voting theory.

Finally, the last group of papers is devoted to Macroeconomics and Finance.

Section III: Macroeconomics and Finance

The first paper in this section is written jointly by Chandana Ghosh and Ambar Nath Ghosh. In their paper *Monetary Policy and Crisis* they argue that in times of crisis central bank's conventional monetary policy of cutting interest rates in the face of declining output may be counterproductive. Instead of reversing the contraction in aggregate income such a policy may reinforce the recessionary forces. The reason lies in the likely behaviour of saving and investment in times of crisis.

The next joint paper entitled *An Effective Demand Model of Corporate Leverage and Recession* jointly by Rilina Basu and Ranjanendra Narayan Nag, analyses the role of balance sheet effect in triggering of financial meltdown. Specifically, their paper extends Paul Krugman's model (2008) to analyse how financial crisis can generate economy-wide recession. We construct an effective demand model to examine how channels of transmission operate through the balance sheet of highly leveraged firms.

Fr. John Felix Raj S. J. and Samrat Roy in their paper *Empirical Evidence on the Relationship between Stock Market Development & Economic Growth: A Cross-Country Exploration in Asia* examine the causal relationship between stock market development and economic growth for selected Asian economies covering the period from 1980–2010. The study has been done for China, India, Pakistan, Sri Lanka, Indonesia, Malaysia, Philippines and Singapore. The empirical part of this study applies Granger-causality technique to arrive at causal relationships between the variables in the study. The evidence obtained from a sample of eight countries suggests that a well-developed stock market can foster economic growth in the long run. It also provides support to theories according to which well-functioning stock markets can promote economic development by fuelling the engine of growth through faster capital accumulation, and by tuning it through better resource allocation.

Next paper in this section is *Financial Development in India: An Empirical Test of the McKinnon-Shaw Model* by Mahendra Pal. He articulates that India is a case of financial openness. His paper examines theoretically and empirically the McKinnon–Shaw model in India. According to McKinnon, a basic complementarity exists between money and physical capital. The model predicts that a high real interest rate policy will stimulate savings and investment and promote economic growth. The view stands in sharp contrast with the Neo-Classical and Keynesian view which contend that lowering the interest rate will stimulate investment and economic growth. Using time-series data for India for the period of 40 years (1971–2010), a simultaneous equations model for Money Demand Function and Saving Function is tested with the help of 2SLS and OLS technique. Positive and significant results are indicated for India. Results support continued financial liberalization with effective macroeconomic management.

Sadhan Kr. Chattopadhyay's paper on *Dynamics of Indian Stock Market* argues that the Indian stock market is considered to be one of the earliest in Asia, which is in operation since 1875. However, it remained largely outside the global integration process until 1991. A number of developing countries in association with the International Finance Corporation and the World Bank took steps in establishing and revitalizing their stock markets as an effective way of mobilizing and allocating of funds. In line with the global trend, reform of the Indian stock market also started with the establishment of Securities and Exchange Board of India (SEBI), although it became more effective after the stock market scam in 1991. With the establishment of SEBI and technological advancement, Indian stock market has now reached the global standard. The major indicators of stock market development show that significant development has taken in the Indian stock market during the post-reform period. His paper examines in this context whether reform in the Indian stock market has led to integration with the developed stock markets in the world. The study finds that contrary to general belief, Indian stock market is not co-integrated with the developed market as yet. Of course, some short-term impact does exist, although it is found to be unidirectional for obvious reasons. That is to say, the developed stock markets, viz., USA, UK and Hong Kong stock markets Granger cause the India stock market but not vice versa. However, the study does not find any causality between the Japanese stock market and Indian stock market. It is derived from the study that although some positive steps have been taken up, which are responsible for the substantial improvement of the Indian stock market, these are perhaps not sufficient enough to become a matured one and hence not integrated with the developed stock markets so far.

Anirban Dutta and Partha Pratim Sengupta in their joint paper *Analysis of Revenue Efficiency—Empirical Study of Indian Non-Life Insurance Companies* have analysed the revenue efficiencies of Indian non-life insurance companies in post-liberalised regime using the new cost based non-parametric Data Envelopment Analysis (DEA) model proposed by K. Tone (2002) for both scale assumptions of constant and variable returns to scale. We have formulated three inputs and two outputs model for analysis of twelve Indian non-life insurance companies over the period of 2005–2006 to 2009–2010. The overall revenue efficiency for life insurance sector is 86 and 96.4 %

under CRS and VRS assumption respectively. This implies that there is a scope of 14 and 3.6 % improvement of revenue under both the assumptions, respectively.

The seventh paper in this section is *Empirics on Fiscal Smoothing: Some Econometric Evidence for the Indian Economy* by Narain Sinha and R. C. Sharma. They argue that Fiscal Smoothing plays an important role in macroeconomic policies. It implies that the fiscal behaviour in an economy is sustainable under unchanged fiscal policies. Fiscal smoothing includes both tax smoothing and revenue smoothing in an economy. Optimal collection of taxes has been an area of interest. Assuming that the taxes are distorting; Barro (1979) propounded the tax smoothing hypothesis which suggests that tax rates should be smoothed overtime using a dynamic optimal control problem to be solved by the government. Employing similar framework, Mankiw (1987) derived the revenue smoothing hypothesis as a part of positive theory monetary and fiscal policy. The underlying basic principle is that an increase in government revenue requires the use of both fiscal and monetary policies. Several attempts have been made in the past to test the tax smoothing hypothesis at the national and subnational levels. Using the contemporaneous single equation OLS Mankiw (1987) and Poterba and Rotemberg (1990) find support for the revenue-smoothing hypothesis. Employing more advanced theory of econometric analysis of time series, Trehan and Walsh (1990), Froyen and Waud (1995), Ghosh (1995) and Evans and Amey (1996) generally reject the revenue smoothing. Tax smoothing hypothesis is generally supported for the federal governments and rejected for the state and local levels (Strazicicich 1997).

In their paper an attempt is made to test both tax smoothing hypothesis and revenue smoothing for central and state taxes for the Indian economy implying that the tax rates are a martingale. Their paper tests a version of Barro's tax-smoothing model, and the Mankiw revenue smoothing hypothesis using the Indian data for 1970–1971 and 2000–2001, respectively. They use the OLS method of estimation to test the revenue smoothing hypothesis and unit root tests based on the single equation approach in which the tax rate is treated as predetermined analysis of the time-series characteristics of tax-tilting behaviour indicating that fiscal behaviour in India is consistent with tax smoothing. The analysis is organized as follows. Section 2 reviews the tax-smoothing and revenue smoothing models and presents their testable theoretical propositions. Section 3 reports the empirical results followed by Sect. 4 concluding the paper.

In achieving inclusive growth, financial inclusion plays a pivotal role. The Committee on Financial Inclusion defines financial inclusion as the process of ensuring access to financial services, and timely and adequate credit when needed by vulnerable groups such as weaker sections and low income groups at an affordable cost. With the arrival of banking technology and realization that poor are bankable with good business prospects, financial inclusion initiatives will strengthen financial deepening further and provide resources to the banks to expand credit delivery. Thus, financial inclusion along with the government developmental programmes will lead to an overall financial and economic development in the country. In this context, Ram Pratap Sinha's paper entitled *Index of Financial Inclusions: Some Empirical Results* focussing on empirical evidences, suggests that there are considerable inter-state variations in the provision of financial services across the Indian states and Union Territories. However, cross-state comparison in this matter requires the construction of some index which could be gainfully employed to make such a comparison. Sarma (2008) proposed a framework for the computation of an index of financial inclusion in the lines of the Human Development Index. His paper however, shows how such indices could also be developed using the non-parametric approaches. For this, he computes the index of financial inclusion of 29 Indian States and 6 Union Territories for the year 2005–2006 using both Euclidean and Output Distance Function Approach. Also, he captures the regional variation in the scores which indicate the extent of regional asymmetry which is persistent in the Indian economy in the matter of financial inclusion.

Joydeep Mukherjee, Debashis Chakraborty and Tanaya Sinha in their paper *The Casual Linkage between FDI and Current Account Balance in India: An Econometric Study in the Presence of Endogenous Structural Breaks* begin with an overview of the Indian reform process: In 1991 as par the recommendations of the IMF, India followed a structural adjustment programme. The new economic philosophy shifted towards export-oriented growth model, where augmenting competition in the domestic market through reforms in licensing provisions and adoption of better technological capabilities through FDI collaborations have played an extremely important role. Over the last decade, the high economic growth in India resulting from the reforms has motivated massive FDI inflow in the country. The continuous inflow has caused India's share in global FDI inward stock to increase from 0.08 % in 1990 to 0.22 and 1.03 % in 2000 and 2010 respectively. However, the improved FDI scenario in India has simultaneously witnessed a decline in the current account balance (CAB) of the country. In this background, the current paper attempts to explore the underlying long term co-integrated relationship between FDI inflow in India and CAB by analyzing quarterly data over 1990–1991: Q1 to 2010–2011: Q4. Their result indicates that there exists a unique long-run relationship among FDI and CAB with two endogenous structural breaks. The analysis also reveals a unidirectional causality from India's FDI to CAB at 5 % level. The findings imply that although FDI may seem beneficial as a source of financing for the current account deficit, it may eventually lead to balance of payments problems due to adverse effects on current account. In this respect, even the role of FDI on economic growth can be questioned. Secondly, the huge outflow of foreign exchange from the country in recent years in the form of profit remittances raises the concerns over the optimality of allowing 100 % profit repatriation.

History is replete with financial ups and downs. One can focus on the depression years of 1929–1939, on the 'great inflation' of 1973–1979, on the third-world—especially the Latin American Debt Crisis beginning from Mexico of the 1980s, on the stock market crashes of 1987, on the Japan's lost decade of 1991–2003, the 1992 EMS crisis being the harbinger of the Tequila effect of Mexico of 1994–1995 aggravated by the continuous depreciation of paper currency, which undermined confidence and ultimately, capital flows and the Asian flu in 1997–1998, on the emerging market crisis of 1997–2002, on the Russian financial crisis in 1998, on the Argentine Financial Crisis (1999–2002) in 2001–2002, on the global credit crisis and recession of 2007–2009, on the European Debt crisis in 2010–2011 leading to stagflation in Europe and elsewhere or on a part of other dramatic events. In this context Asim K. Karmaker's paper on *Contagious Financial Crises in the Recent Past and Their Implications for India*, recalls early financial market crises in economic history, which helps us to understand more about the contagion processes observed in recent years. In the following section, his aim is to picture the important events that occurred in Chile in 1982, crisis of the ERM during the early 1990s, the Mexican Crisis of 1994–1995. Last of all, he provides a chronicle of the early crisis of 1997–1998 in Thailand and the subsequent turmoil in many Asian economies, as well as global economic meltdown and the Eurozone crisis of recent years and their implications for India.

The themes of the papers in this edited volume are of contemporary relevance in the advancing the literature on international trade, development economics, macroeconomics as well as finance, It is hoped that the contributions come up to the high academics and standards set by Professor Chatterjee as a teacher of economics and a scholar par excellence. We sincerely hope this collection will be useful to the serious scholars of the subject.

Kolkata

Ambar Nath Ghosh
Asim K. Karmakar

Biswajit Chatterjee

Biswajit Chatterjee
Born (03.03.1954)

Blessed with a brilliant academic career, Professor Biswajit Chatterjee has earned the reputation as an acclaimed economist, an excellent author and editor, learned scholar and researcher of good repute, a very effective, inspiring teacher and research guide, and outstanding administrator and organizer. He has also served as Dean, Faculty Council of Arts at Jadavpur University, Kolkata for 8 years and Head, Department of Economics, Jadavpur University. He is now the Vice-President of the Indian Society of Labour Economics and was President of the Indian Econometric Society during 2011–2012 and 2012–2013, besides being President of BEA since 2005. He has carried out excellent research work on different issues which are considered helpful for India. He has contributed many research papers in reputed journals, authored several books and edited volumes on themes of development economics, international trade and labour economics which are being used in many Indian and International Universities. His books and edited volumes include *Towards a District Development Report for West Bengal* (co-authored, 2003)*, Trade, Finance and Development* (co-edited, 2004)*, Economics of Asymmetric Information* (co-edited, 2006)*, Growth, Distribution and Public Policy: A Study of West Bengal* (2007)*, Development Perspectives: India—Contribution of P.R. Brahmananda* (co-edited, 2007)*, Globalisation and Health Sector in India* (2009)*, Capital Account Convertibility in India* (co-edited, 2011) and *Climate Change, Trade and Natural Disasters* (2011)*, Regional Disparities in Development* (co-edited, 2011)*, Food Security in India* (co-edited, 2012). For the last 36 years, he has taught at Presidency College, Kolkata, Kalyani University, Jadavpur University, University of Calcutta and Rabindra Bharati University and is at present a Visiting Professor at CRRID, Chandigarh. He has taught various subjects like Indian Economy, Macroeconomics, International Trade and Development Economics at the Ph.D., M. Phil., M.A. and UG Honours. levels. So far 17 of his research students placed now in different reputed organizations in India have obtained their Ph.D. degrees from Jadavpur University.

Contents

Part II Development

Part III Macroeconomics and Finance

Contributors

Tanmoyee Banerjee (Chatterjee) Department of Economics, Jadavpur University, Kolkata, West Bengal, India

Swapnendu Banerjee Department of Economics, Jadavpur University, Kolkata, India

Rilina Basu Department of Economics, Rabindra Bharati University, Kolkata, West Bengal, India

N. R. Bhanumurthy National Institute of Public Finance and Policy, New Delhi, India

G. K. Chadha South Asian University, New Delhi, India

Avik Chakrabarti Department of Economics, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Bidisha Chakraborty Department of Economics, Jadavpur University, Kolkata, West Bengal, India

Debashis Chakraborty Indian Institute of Foreign Trade (IIFT), New Delhi, India

Somdeep Chatterjee Department of Economics, University of Houston, Houston, USA

Sadhan Kumar Chattopadhyay Mumbai, India

P. K. Chaubey Indian Institute of Public Administration (IIPA), New Delhi, India

Pradipta Chaudhury Centre for Economic Studies and Planning, School of Social Sciences, Jawaharlal Nehru University, New Delhi, India

Purba Roy Choudhury Department of Economics, The Bhawanipur Education Society College, Kolkata, India

Anirban Dutta NSHM Business School, Durgapur, West Bengal, India

Ambar Nath Ghosh Department of Economics, Jadavpur University, Kolkata, India

Chandana Ghosh Economic Research Unit, Indian Statistical Institute, Kolkata, West Bengal, India

Sanmitra Ghosh Department of Economics, Jadavpur University, Kolkata, West Bengal, India

Sucha Singh Gill Centre for Research in Rural and Industrial Development (CRRID), Chandigarh, India

Manash Ranjan Gupta Economic Research Unit, Indian Statistical Institute, Kolkata, West Bengal, India

Bandi Kamaiah Department of Economics, University of Hyderabad, Hyderabad, Andhra Pradesh, India

Asim K. Karmakar Department of Economics, Jadavpur University, Kolkata, West Bengal, India

Hiranya Lahiri Department of Economics, Jadavpur University, Kolkata, India

Nilanjana Mitra Department of Economics, Susil Kar College, Ghoshpur, West Bengal, India

Shariq Mohd. International Institute of Special Education, Lucknow, India

Jaydeep Mukherjee Indian Institute of Foreign Trade (IIFT), New Delhi, India

Mallinath Mukherjee Department of Economics, St. Xavier's College (Autonomous), Kolkata, India

Dilip M. Nachane Indira Gandhi Institute of Development Research (IGIDR), Mumbai, India

Ranjanendra Narayan Nag St. Xavier's College (Autonomous), Kolkata, West Bengal, India

Mahendra Pal Department of Commerce, Delhi School of Economics, Delhi, India

T. S. Papola Institute of Studies in Industrial Development (ISID), New Delhi, India

Rajendra Narayan Paramanik Department of Economics, University of Hyderabad, Hyderabad, Andhra Pradesh, India

J. Felix Raj St. Xavier's College (Autonomous), Kolkata, West Bengal, India

Runa Ray Department of Economics, Vidyasagar College, Kolkata, India

Rupamanjari Sinha Ray Management Development Institute, Gurgaon, Haryana, India

Samrat Roy St. Xavier's College (Autonomous), Kolkata, West Bengal, India

Partha Pratim Sengupta National Institute of Technology (NITD), Durgapur, West Bengal, India

R. C. Sharma Department of Management, Central University of Rajasthan, Kishangarh, Rajasthan, India

Prakash Singh Research Scholar at Dept. of Business Economics, Delhi University & ICSSR, New Delhi, India

Narain Sinha Department of Economics, University of Botswana, Gaborone, Botswana

Ram Pratap Sinha Government College of Engineering and Leather Technology, Kolkata, India

Tanaya Sinha Amity University, New Delhi, India

Rameshwar Tandon International Institute of Special Education, Lucknow, India

About the Editors

Prof. Dr. Ambar Nath Ghosh is a professor of economics at Jadavpur University, Kolkata and a former professor of economics at Presidency College, Calcutta. He has contributed many research articles in respected national and international journals. He specializes in macroeconomic theory, public economics, international trade and finance. His published books include: *Economics of Public Sector* (2008) and *Macroeconomics* (2011), both with Prentice Hall Pvt. Ltd., New Delhi.

Dr. Asim K. Karmakar is assistant professor of economics at Jadavpur University, Kolkata and managing editor of the quarterly referred journal, Artha Beekshan. He is also the Executive Council Member of 'The Indian Econometric Society'. He has contributed many research articles in national and international journals of repute. He specializes in balance of payments, Indian economy and development economics. His published books include: *Balance of Payments Theory and Policy* (2010) and *Capital Account Convertibility in India* (Co-edited) (2011), Deep & Deep Publications, New Delhi and *Food Security in India* (co-edited) (2012), Regal Publications, New Delhi.

Part I
International Trade

Chapter 1
Trade–Environment Linkage: A South-centric Model-Specific Analysis

Mallinath Mukherjee

1 Introduction

Forces of globalization and concern for the environment seem to be the core issues of national and international debates during the last almost two decades with convergence of ideas increasingly appearing to be a distant possibility. Big trans-boundary environmental issues like climate change, global warming, etc. as well as local environmental externality problems have been drawing ever-increasing public attention Anderson (1992) Bovenberg and Smulders (1995), Copeland (1994) Copeland and Taylor (1995). Often, due to the forces of globalization, the sources of externality may be shifted from one nation to another. Brander and Taylor (1997) The comparative advantage of a country in the production of a polluting good triggers trade-induced resource flows and raises doubts about the gains from free trade. The underdeveloped South may experience a seriously endangered environment because of their poorly defined property rights over environmental resources, environment not entering the utility function as a normal commodity in the South, strategic reductions in the domestic environmental standards to gain pseudo-comparative advantage and the like. Trade opportunities are generated due to different environmental standards. More trade and less pollution make the North unambiguously better off; more trade but more pollution in the South cast doubts on globalization. This has been the traditional perception.

However, more trade and the need to protect the environment have been at the root of some new challenges the less-developed nations are facing. First, the food security of the poor nations is increasingly at stake. Second, in a skill-divided society, the medium-skilled labour in the South might have been gaining while both the high-skilled and the unskilled labour might have been losing. Finally, the use-right of land might have been silently coming up as a tradable product! Globalization implies some debilitation of the powers of the nation states. If use-right of land becomes

M. Mukherjee (✉)
Department of Economics, St. Xavier's College (Autonomous),
Kolkata 700016, India
e-mail: mukherjeemallinath@rediffmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_1,

tradable (in the sense that such use-right may be contractually leased out for long periods to 100 % equity-holding foreign firms), it may have serious implications for future consensus on both more trade and better environment.

This paper is an attempt to capture these three possibilities in the context of a small open economy. The North shall not be explicitly taken into account in this paper to concentrate only on the South. The North is assumed to have a perfectly elastic demand for exports from the South and a perfectly elastic supply of South's imports—both at the global prices given parametrically to the South.

An introduction to the proposed tradability of land services is in order. We do not have the competitive market framework for such services. In the absence of arbitrage, there is no reason for the land rents to get equalized across the globe. Domestic and foreign firms willing to purchase such land services operate through the intermediation of the State in the South. Thus, for example, the purchasing firms and the selling farmers do not negotiate the price directly. There is, therefore, a transaction cost. Because of differences in the transaction cost in the different nations of the South or, even in different parts of the same nation, land rents still vary.

The rest of the paper is organized as follows. Section 2 spells out the model. Section 3 gives the working of the model. Section 4 provides the basic results of the model. Section 5 works out some comparative statics to analyse the policy implications. Section 6 concludes.

2 Trade–Environment Linkage: A Small Open Economy Framework

The small open economy comprises the traded and the non-traded sectors. The traded sector, consisting of the import-competing and the export sectors, practises clean production; clean goods ensure international acceptability. The non-traded sector (i.e. the food sector in this paper), however, uses several environmental resources as free inputs; thus, for example, water and water resources, forests and forest resources, grazing land, etc. are free inputs to the non-traded sector. This is, thus, a three-sector model.

Labour is a heterogeneous input, differentiated by different levels of skill. Each sector uses a different level of skill. We assume that the import-competing sector is the most skill-intensive, as it has to compete with the imported goods. The non-traded sector is the least skill-intensive, and the export sector needs a medium level of skill. Exportables are assumed primarily to be outsourced services, products of the extractive industries, gems and jewellery, leather goods, etc. Gems and jewellery, leather goods, etc. do need artistic skills of high order. However, it is assumed that such skills are not achieved through expensive formal training. Rather, such skills are often inherited or gifted. Heterogeneous labour is, however, downwardly mobile. When trade-induced resource reallocations take place, labour can move from the high-skilled to the low-skilled sectors (but not the other way round). Thus, the import-competing sector uses only high-skilled labour. Export production is possible with both high-skilled and medium-skilled labour—both used at the medium level of skill. *For the convenience of arriving at a solution, We assume that the non-traded sector does not use any displaced labour from the traded sector*. Our small open economy practised import-substituting industrialization with very strong state intervention during 1950–1980 and even up to a part of the 1980s. The import-competing sector, therefore, became the hot breeding ground of equally strong trade unionism (often patronized by the State). In our model, such historical trade unionism is captured by a given real wage of the high-skilled labour in the import-competing sector indexed to the food price. Thus, $w_s{}^o$—the food price-indexed real wage in the import-competing sector—is a parameter given by $w_s{}^o = W_s/P_Z$, where W_s = money wage in the import-competing sector and P_Z = the food price. Given initial autarky equilibrium, P_z is in its initial equilibrium, and therefore, W_s is determined. As the South opens up to globalization, W_s changes provided the food price changes and change in the food price depends upon domestic demand for and the supply of food. Food price is endogenously determined.

Food demand is a given proportion of total wage income, i.e. $Z_d = \beta \cdot I$, $\beta > 0$, where I = total wage income = $W_s \cdot$ (total high-skilled labour used in the import-competing sector) + $W_M \cdot$ (total labour used in the export sector) + $W \cdot$ (total labour used in the non-traded sector) $\cdot$ W_M is the wage rate in the export sector and W = the wage rate in the non-traded sector; both these are fully flexible so that full employment is achieved.

Capital (K) is specific to the import-competing sector (X); the rental rate on capital (r) is fully flexible that ensures full use of the available capital stock in Sector-X.

Land services are used in both export production (Sector-Y) and the non-traded sector (Sector-Z). It is assumed that services of land are a homogeneous input usable in both these sectors. Given the stock of available land, the stock of available land services is also given. We assume that land services are tradable with a country-specific transaction cost. Under autarky, the initial rental rate on land (R_o) is endogenously determined. After the South opens up, the rental rate is given by the transaction cost adjusted world rent. Assuming a given transaction cost per unit of land service, this transaction cost-adjusted world rent is a parameter in our model.

Production functions in all the three sectors obey constant returns to scale. Production in the non-traded sector follows a Cobb–Douglas function. The non-traded sector uses land, low-skilled labour and free environment. This free environment affects productivities of land and labour in the sector. Thus, the production function in the non-traded sector is given by $Z = A(E) \cdot T_z{}^{1-\alpha} \cdot L^{\alpha}$, $0 < \alpha < 1$, where L = labour used in the non-traded sector, T_z = land used in Z and E = free environmental resource available to the firms in the non-traded sector such that $A'(E) > 0$. E is policy-determined by the State. Thus, if the State, in order to improve the quality of environment, makes the environment laws stricter, E declines, and therefore, "A" declines reducing, thereby, the factor productivities. We capture the trade–environment linkage in a small open economy through policy-induced changes in E generating a productivity effect in the non-traded sector.

Our model may be analytically presented as follows:

The three production functions for the three sectors are:

1. The import-competing sector: $X = X(K, S)$, where $K =$ capital stock, $S =$ skilled labour used in sector X. The unit output input coefficients are 'k' and 's', respectively. There is a given S_o available to the firms. In the initial no-trade situation, entire S_o is used in the import-competing sector. Once the economy opens up, S of S_o is used in Sector-X (depending upon the given world and the endogenously determined domestic prices) and the rest, if any, $(S_o - S)$ has to move to the sector needing less skill, i.e. Sector-Y in our model. Such labour gets less wages as well.
2. The export sector: $Y = Y(T_Y, M)$, where T_Y is land used in export production and M is labour used in this sector. The economy has a given amount of medium-skilled labour, M_o. In the initial autarky equilibrium, the entire M_o is used in Y. Once the small economy opens up, the given world prices and the endogenously determined domestic prices set the profit-maximizing demand for labour at M that uses the whole of $S_o - S$ in export production over and above M_o. W_M is flexible to achieve it. The unit output input coefficients are 'a_{ty}' and 'm', respectively.
3. The non-traded sector: $Z = Z(E, T_z, L) = A(E) \cdot T_z^{1-\alpha} \cdot L^{\alpha}$, $0 < \alpha < 1$, where E is free environmental resource affecting factor productivities in Z through A, T_z is land used in the non-traded sector and L is labour used in the non-traded sector. The economy has a given amount of unskilled labour, L_o. In the initial equilibrium, entire L_o is used in food production. As the economy opens up, the non-traded sector adjusts to the new prices and in the process, competes with the traded sectors for resources, and L amount of unskilled labour is demanded in Z. We assume that $L = L_o \cdot W$, the wage rate of unskilled labour, is flexible to achieve this. Thus, if due to globalization, production in the non-traded Sector-Z changes, the change is brought about by changes in policy-induced A(E) or, price-induced T_Z. The unit output input coefficients are 'a_{tz}' and 'l', respectively.

The nature of the non-traded sector is as follows:

1. Food demand: Nominal food demand is a given proportion of total wage income, i.e.
   $Z_d = \beta \cdot I$, where $I =$ total wage income $= W_s \cdot S + W_M \cdot M + W.L. = w_s^o \cdot P_z \cdot S + W_M \cdot M + W \cdot L$, $\beta > 0$.
2. Food supply: The domestic production of food is $Z = Z(E, T_z, L) = A(E) \cdot T_z^{1-\alpha} \cdot L^{\alpha}, 0 < \alpha < 1$. Food being the non-traded good, domestic market clearance implies $P_z \cdot Z = Z_d$, where Z is food production.

The following may be noted with respect to factor prices:

1. Wage rates in the three sectors: Wage rate in the high-skilled-labour-using import-competing sector is indexed to food price (food price is endogenously determined at the domestic market-clearing level) and such indexation is feasible due to trade unionism. This trade unionism in Sector-X has its legacy in the State-led import-substituting industrialization practised in our Southern nation earlier. Thus, real wage in Sector-X is given at w_s^o, i.e. W_s, the money wage in X, is $w_s^o \cdot P_z$. Any increase in the domestic food price raises W_s at the same rate.

While the wage rate in X is food-price indexed, wage rates in the export and the non-traded sectors are fully flexible. W_M is the wage rate in the medium-skill-using export sector and finally, W is the wage rate in the non-traded sector. Evidently, $W_s > W_M > W$.

2. Land rent: The nation has a given amount of homogeneous land services usable in both sectors Y and Z. Under the initial autarky equilibrium, land services are fully used and rent is endogenously determined at R_o. Once the small economy opens up transaction cost-adjusted land rent is given by R where R = world rental rate on land + nation-specific transaction cost (transaction cost per unit of land service is assumed to be parametrically given). Evidently, $R > R_o$. In the small open economy, services of land may become exportable if contraction of the non-traded sector due to openness releases more land than what is additionally demanded by the expanding export sector.
3. Capital rent: The rental rate on capital (r) is endogenously determined. Capital is a specific factor in the import-competing sector. Flexibility of r ensures full use of the available capital in the import-competing sector.

3 Working of the Model

The working of our model for the small open economy is as follows.

First, we may spell out the full-employment equilibrium for this small open economy. P_X and P_Y are given to this economy by the rest of the world. Given that, by assumption, X is the importable and Y is the exportable, P_X falls and P_Y rises as the small economy opens up. P_z—the food price—is endogenously determined at the domestic market-clearing level. Given competitive equilibrium, price equals unit cost. Thus, zero-profit conditions in the three sectors imply:

a. $$P_X = W_s \cdot s + r \cdot k = w_s^{o} \cdot P_z \cdot s + r \cdot k \tag{1.1}$$

b. $$P_Y = W_M \cdot m + R \cdot a_{ty} \tag{1.2}$$

and

c. $$P_z = W \cdot l + R \cdot a_{tz} \tag{1.3}$$

This price sub-system contains three equations in four unknowns—P_z, W_M, W and r. Given the endogenously determined P_z and the world P_x, the import-competing sector determines ‘r’. Given R and P_Y, the export sector determines W_M. Given R and P_z, the non-traded sector determines W. The factor prices determine the unit output input coefficients—k, s, m, l, a_{ty} and a_{tz}.

P_z is determined by domestic food market clearance. This gives us

d. $$P_z \cdot A(E) \cdot T_z^{1-\alpha} \cdot L^{\alpha} = \beta \cdot (w_s^{o} \cdot P_z \cdot S + W_M \cdot M + W \cdot L) \tag{1.4}$$

Finally, we have the full-employment conditions given by

e. $K = k{\cdot}X$ (1.5)

f. $T = T_Y + T_Z + T_E$, where T_E = exportable land services (1.6)

g. $S_o = S + SY$ (1.7)

where SY = high-skilled labour used as medium-skilled labour in export production,

h. $M_o = (M - SY)$ (1.8)

and

i. $L_o = L$ (1.9)

Our small open economy is described by Eqs. (1.1)–(1.9) which solve for the nine unknowns: X, Y, Z, P_z, W_M, W, r, S and M.

Intuitively, as the economy opens up from the initial autarky equilibrium, P_X falls and P_Y rises. Lower P_X reduces X and, given the specificity of capital in this import-competing sector, a part of S_o, i.e. high-skilled labour must be released. Therefore, openness implies $S < S_o$, i.e. $SY > 0$. This part of the labour force must accept lower wages in the export sector of the economy. This is the first negative effect of globalization on labour. Higher P_Y raises Y and this raises the derived demand for land and labour in the export sector. The export sector draws additional land from the non-traded sector and additional labour from the contracting import-competing sector. The movement from autarky to openness has two negative effects on the food sector. First, land drawn out by the expanding export sector has a negative supply effect on Z which tends to reduce Z and raise P_z. Second, transfer of every unit of labour from X to Y reduces wage income by $W_s - W_M = w_s{}^o \cdot P_z - W_M$. The higher the food prices, the greater the loss. Through the food demand equation, such loss of wage income has a negative demand effect on Z which tends to reduce Z and lower P_z. Expansion of trade implies contraction of the non-traded sector. Let there be simultaneous enforcement of stricter environmental standards. This reduces the availability of free environmental resources to the firms in Z and A(E) declines. This has adverse factor productivity effect in Z. The policy-induced adverse supply effect on the non-traded sector reduces Z and raises P_z. The first proposition that intuitively follows is an endangered food security in a small open economy trying to enforce stricter environmental standards. It calls for some policy prescription. Most such underdeveloped economies have significant amounts of common property resources which may be effectively exploited to raise E to the firms in Z. Globalization-induced negative supply and demand effects on Z may be countered by making more 'E' available to the producers in Z. As the common property resources are effectively exploited and the availability of E to the economy is raised, it is possible to make greater amounts of free resources available to Z and simultaneously improve the quality of the environment.

The interface between trade and environment has another major implication. As the small economy opens up, the expanding export sector can draw labour from the contracting import-competing sector, but not from the non-traded sector (due to only downward mobility of labour). Given the factor specificity of capital in X, the import-competing sector must release a part of S_o. The amount of this released labour would be a determinant of how conveniently the export sector can expand. Expansion of Y is also feasible through land drawn from Z. However, given tradability of land services, R—the post-trade land rent—is determined by global rent and parametrically given domestic transaction cost. Therefore, $R > R_o$ would imply parametrically given high rent cost and the export sector would tend to depend increasingly on labour released by X. It means, in a skill-divided society, globalization is limited by the availability of the right skill. This has a policy implication. If education makes a part of the unskilled labour of Z medium skilled, the expansion of the export sector would be less constrained. Given the twin dangers of endangered food security and skill-constrained pressure on expansion of the export sector, a package approach is called for. This comprises an effective exploitation of the common property resources combined with medium-level skill formation among the unskilled workers. In the absence of any such ameliorating policy package, increasing dependence on labour of the growing export sector would tend to raise W_M. For any given P_z, wage in the high-skilled import-competing sector, W_s, is unchanged and, thus, it reduces the wage gap between the two traded sectors. Private preference for medium level of skill formation may increase and private expenditure on high-skill formation may fall. This may endanger even the competitive existence of the import-competing sector.

The model gives us a reasonably transparent picture about what is welcome by whom. Since $W = P_z \cdot \text{MPL}$, stricter environment standards that reduce E, and therefore, MPL, and hence, W, are resisted by the unskilled working class. Demand contraction (through labour reallocation) tends to reduce P_z and supply contraction tends to raise it. Even if P_z rises, W cannot increase at a higher rate than P_z due to the fall in MPL. Stricter environmental standards in a small open economy make the weak weaker. The high-skilled workers, on the other hand, oppose globalization, but are indifferent to a better environment. Less E to the non-traded sector raises P_z and, through wage indexation, W_s rises at the same rate, hence the indifference to a better environment. However, such high-skilled labour of the import-competing sector resists globalization because the reallocation of a unit of labour from X to Y reduces wage income by $W_s - W_M$. Given fairly strong trade unions in X, collective bargaining would resist openness. The medium-skilled workers would generally dislike neither globalization nor stricter environmental standards. Openness raises job opportunities in the expanding export sector. Stricter environmental standards cause contraction of the non-traded sector. Given parametric rent, fall in land productivity (due to lower E) would drive land out of Z. More land in the export sector tends to raise marginal productivity of labour in Y. Medium-skilled labour gains from both the higher price of Y and higher marginal productivity of labour in Y (since $W_M = P_Y \cdot \text{MPM}$, where MPM = marginal productivity of labour in Y). However, globalization-induced labour reallocation raises labour in Y that tends to reduce MPM. Thus, MPM rises provided the effect of land transfer dominates the effect of labour transfer. The capitalists are clear losers and oppose both better environment and globalization. They oppose globalization by the standard result of a specific factor trade model. Their resistance to better environment originates from the contraction of the non-traded sector that raises P_z and, through wage indexation, raises $W_{s.}$ Given the world price of X and a zero-profit condition, the rental rate on capital (r) falls steeply, and hence, the resistance. The attitude of the landed class is ambivalent.

Finally, the model opens up a possibility of exportable surplus of land. High rent implies substitution of land by labour in the export sector restricting the additional demand for land in Y. If, therefore, the contracting non-traded sector releases more land than what is additionally demanded by Y, an exportable surplus of land services is generated. This has serious implications for changing complexion of international relations. It has policy significance as well. If the surplus land is procured by the native State and used for developing common property environmental resources (e.g. for forestry, rainwater harvesting, etc.), even stricter environmental standards need not reduce E and the probable food insecurity of the small open economy may be avoided.

4 Basic Results of the Model

Given cost minimization, we get the following for the three sectors:

$$\theta_s \cdot \hat{s} + \theta_K \cdot \hat{k} = 0, \tag{1.10}$$

where θ_s and θ_K are the respective factor shares in X,

$$\theta_M \cdot \hat{m} + \theta_{Ty} \cdot \hat{a}_{ty} = 0, \tag{1.11}$$

where θ_{Ty} and θ_M are the respective factor shares in Y

and

$$\theta_L \cdot \hat{l} + \theta_{Tz} \cdot \hat{a}_{tz} = 0 \tag{1.12}$$

where θ_{Tz} and θ_L are the respective factor shares in Z.

In each equation, the sum of the two relevant factor shares adds up to unity.

Given cost minimization, zero-profit conditions of the three sectors give the following:

$$\hat{P}_x = \theta_s \cdot \hat{W}_s + \theta_k \cdot \hat{r} = \theta_s \cdot \hat{P}_z + \theta_k \cdot \hat{r}, \tag{1.13}$$

$$\hat{P}_Y = \theta_M \cdot \hat{W}_M + \theta_{Ty} \cdot \hat{R}, \tag{1.14}$$

and

$$\hat{P}_z = \theta_L \cdot \hat{W} + \theta_{Tz} \cdot \hat{R}. \tag{1.15}$$

We have the three elasticities of substitution given by

$$\sigma_x = -(\hat{s} - \hat{k})/(\hat{W}_s - \hat{r}) = -(\hat{s} - \hat{k})/(\hat{P}_z - \hat{r}), \tag{1.16}$$

$$\sigma_y = -(\hat{m} - \hat{a}_{ty})/(\hat{W}_{\mathrm{M}} - \hat{R}), \tag{1.17}$$

and

$$\sigma_z = -(\hat{l} - \hat{a}_{\mathrm{tz}})/(\hat{W} - \hat{R}) = 1\,(\text{by the Cobb} - \text{Douglas production function}) \tag{1.18}$$

We turn now to the solution of the model:

From Eq. (1.13),

$$\theta_{\mathrm{k}} \cdot \hat{r} \quad = \quad \hat{P}_x - \theta_{\mathrm{s}} \cdot \hat{P}_z,$$

or

$$\hat{r} \quad = \quad \hat{P}_z + \left(\frac{1}{\theta_k}\right) \cdot (\hat{P}_x - \hat{P}_z). \tag{1.19}$$

Similarly, from Eq. (1.14),

$$\hat{W}_{\mathrm{M}} \quad = \quad \hat{P}_Y + (\theta_{Ty}/\theta_{\mathrm{M}}) \cdot (\hat{P}_Y - \hat{R}), \tag{1.20}$$

and from Eq. (1.15)

$$\hat{W} \quad = \quad \hat{P}_z + \frac{1-a}{a}(\hat{P}_z - \hat{R}). \tag{1.21}$$

Thus, given the changes in R and the prices of X, Y and Z, the changes in factor prices are fully determined. Given the changes in the factor prices, the optimal or cost-minimizing changes in the factor coefficients are determined. Given the changes in the cost-minimizing factor coefficients and the full-employment conditions, the changes in the output levels are determined. It is evident from Eqs. (1.19)–(1.21) that, while the changes in R, P_X and P_Y are parametrically given to our economy, the change in P_z is endogenously determined. Therefore, our system is solved once the change in P_z is determined. We have to turn to the non-traded sector now. Unskilled labour of the non-traded sector cannot move to any other sector, i.e. $L = L_o$. Flexibility of W ensures that the simplifying assumption is viable. The full-employment conditions of labour imply $S_o = S + SY$ and $M_o = M - SY$. Armed with this simplification we solve for P_z.

P_z is solved by the domestic market clearance of the non-traded sector, i.e.

$$P_z \cdot A(E) \cdot {T_z}^{1-\alpha} \cdot L^{\alpha} = \beta\,(W_s \cdot S + W_M \cdot M + W \cdot L),$$

or, $\hat{P}_z + \widehat{A(E)} + (1-\alpha)(\hat{a}_{\mathrm{tz}} - \hat{l}) = a \cdot \hat{s} + \mathrm{a} \cdot \widehat{W}_{\mathrm{s}} + \mathrm{b} \cdot \widehat{M} + \mathrm{b} \cdot \widehat{W}_M + \mathrm{c} \cdot \widehat{L} + \mathrm{c} \cdot \widehat{W} - \widehat{L},$ (1.22)

where ‘a’, ‘b’ and ‘c’ are sector-specific shares of labour income in total wage income.

Now, dS (i.e. displaced high-skilled labour from X) $= S - S_o = -SY$

Thus, $dS/S = \hat{S} = -SY/S = e$ (displaced labour as a proportion of post-globalization employment in X). Here, 'e' is negative.

Similarly, dM (i.e. additional labour demand in Y) $= M - M_o = SY$.

Thus, $dM/M = \hat{M} = SY/M = \frac{S}{M} \cdot \frac{SY}{S} = -f \cdot e$ (where f = post-globalization employment ratio in X to Y). We have, by assumption, $\hat{L} = 0$.

Therefore, from Eq. (1.22),

$$\widehat{P_z} + \widehat{A(E)} + \delta_z \cdot (1-\alpha) \cdot (\widehat{W} - \widehat{R}) = a \cdot e + a \cdot \widehat{W_s} - b \cdot f \cdot e + b \cdot \widehat{W_M} + c \cdot \widehat{W},$$

$$\text{given that } \widehat{L} = 0 \text{ and } \widehat{a_{tz}} - \widehat{l} = \delta_z(\widehat{W} - \widehat{R}), \tag{1.23}$$

or given $\delta_z = 1$ and $\hat{W}_s = \hat{P}_z$,

$$\widehat{P_z} = g + h \cdot \widehat{W_M} + i \cdot \widehat{W} + J \cdot \widehat{R} - n \cdot \widehat{A(E)}, \tag{1.24}$$

where $g = (a - b \cdot f) \cdot e/1 - a$, $h = b/1 - a$, $i = (c + \alpha - 1)/1 - a$, $j = (1 - \alpha)/1 - a$, and $n = 1/1 - a$.

Using Eqs. (1.20), (1.21) and (1.24),

$$\hat{W} = \left[\frac{g - n \cdot \hat{A}}{a - i}\right] + \left[\frac{h}{\{\theta_M(a-i)\}}\right] \cdot \hat{P}_y + \left[\frac{\left\{(a + j - 1) - h \cdot \left(\frac{\theta_{TY}}{\theta_M}\right)\right\}}{a - i}\right] \cdot \hat{R}. \tag{1.25}$$

Since $\hat{A} = \widehat{A(E)}$ is a parameter induced by the environmental policy of the government and $\hat{P}_y$ and $\hat{R}$ are globalization-generated parameters, $\hat{W}$ is determined by Eq. (1.25).

Similarly, Eqs. (1.19), (1.20) and (1.24) together solve for $\hat{r}$. Equation (1.20) gives the solution to $\hat{W}_M$.

As evident from Eqs. (1.19)–(1.24), the model is open to alternative outcomes. We shall highlight only one possibility. P_x, P_Y and R are given from outside. Given our model specifications, $\hat{P}_x < 0$, $\hat{P}_Y > 0$ and $\hat{R} > 0$. Let us analyse the implications of $\hat{P}_y > \hat{R}$. Enforcement of stricter environment standards on Z in our small open economy has two adverse supply effects as well as an adverse demand effect. Assuming the policy-induced supply effect to be stronger, the price of the non-traded good rises. Let $0 < \hat{P}_z < \hat{R}$. By Eq. (1.20) and the assumption that $\hat{P}_y > \hat{R}$, we have $\hat{W}_M > \hat{P}_y > \hat{R}$ and by Eq. (1.21), and the assumption that $\hat{P}_z < \hat{R}$, we have $\hat{R} > \hat{P}_z > \hat{W}$. By wage indexation of labour in the import-competing sector, $\hat{P}_z = \hat{W}_s$. Combining, $\hat{W}_M > \hat{P}_y > \hat{R} > \hat{P}_z = \hat{W}_s > \hat{W}$. Two points come out clearly from this possibility. First, the wage gap between the high-skilled and the medium-skilled workers declines; W_M increases at a faster rate than W_s. Second, the wage gap between the high-skilled and the medium-skilled workers on the one hand and the unskilled workers on the other hand increases. Trade and environment linkage analysis must address such skill-specific effects on wage rates; otherwise, policy prescriptions may have little significance. This gives us our hypothesis on intra-country wage disparities. Our hypothesis, however, does not rule out other possibilities.

To establish the possible threat to food security in this small open economy enforcing simultaneously stricter environmental standards, we return to the production function of the non-traded sector.

This is

$$\mathrm{Z} = \mathrm{A(E)} \cdot \mathrm{T_z}^{1-\alpha} \cdot L^{a}, 0 < \alpha < 1.$$

Thus, $\widehat{Z} = \widehat{A(E)} + (1-\alpha)\cdot \widehat{T}_{\mathrm{z}} + \alpha \cdot \widehat{L}$.

Since $\hat{L} = 0$, by assumption, we have

$$\widehat{Z} = \widehat{A(E)} + (1-\alpha)\cdot \widehat{T}_{\mathrm{z}}. \tag{1.26}$$

Now, from Eq. (1.15),

$$\hat{P}_{\mathrm{z}} = \theta_{\mathrm{L}} \cdot \hat{W} + .\theta_{Tz} \cdot \hat{R}.$$

Thus, $\hat{R} = \hat{P}_{\mathrm{z}} + \frac{\theta_L}{\theta_{Tz}}(\hat{P}_{\mathrm{z}} - \hat{W})$.

Hence,

$$\hat{R} - \hat{W} = \frac{1}{\theta_{Tz}}(\hat{P}_{\mathrm{z}} - \hat{W}). \tag{1.27}$$

Now, $\hat{L} = \hat{l} + \hat{Z}$ and, $\hat{T}_{\mathrm{z}} = \hat{a}_{\mathrm{tz}} + \hat{Z}$, since $\mathrm{L} = l \cdot z$ and $\mathrm{T_z} = a_{tz} \cdot Z$

or $\hat{L} - \hat{T}_{\mathrm{z}} = \hat{l} - \hat{a}_{\mathrm{tz}}$.

However, $\hat{l} - \hat{a}_{tz} = \sigma_z \cdot (\hat{R} - \hat{W}) = (\hat{R} - \hat{W})$, since $\sigma_z = 1$.

Therefore, $\hat{L} - \hat{T}_{\mathrm{z}} = \hat{R} - \hat{W} = \frac{1}{\theta_{Tz}}(\hat{P}_{\mathrm{z}} - \hat{W})$.

$$\text{Given, } \hat{L} = 0, \hat{T}_{\mathrm{z}} = -\frac{1}{\theta_{Tz}} \cdot (\hat{P}_{\mathrm{z}} - \hat{W}). \tag{1.28}$$

Since $\hat{P}_{\mathrm{z}} - \hat{W} > 0, \hat{T}_{\mathrm{z}} < 0$. The contracting non-traded sector releases land in our small open economy enforcing stricter environmental standards. Now, from Eqs. (1.26) and (1.28),

$$\widehat{Z} = \widehat{A(E)} - \frac{1-\alpha}{\theta_{Tz}} \cdot (\widehat{P}_{\mathrm{z}} - \widehat{W}). \tag{1.29}$$

In our model, $\widehat{A(E)}$ is negative. Therefore, from Eq. (1.29), $\hat{Z} < 0$. Stricter environmental standards enforced in a small open economy reduce Z (i.e. food production) and endanger the food security of the nation. This establishes our hypothesis on food insecurity.

We turn now to our last hypothesis on the tradability of land services, or the emergence of exportable surplus of land services in this small open economy. From Eq. (1.14),

$$\hat{R} - \hat{W}_{\mathrm{M}} = \frac{1}{\theta_{Ty}} \cdot (\hat{P}_{\mathrm{y}} - \hat{W}_{\mathrm{M}}). \tag{1.30}$$

Now, $\hat{M} - \hat{T}_{\mathrm{y}} = \hat{m} - \hat{a}_{\mathrm{ty}}$ and $\hat{m} - \hat{a}_{\mathrm{ty}} = \sigma_{\mathrm{y}} \cdot (\hat{R} - \hat{W}_{\mathrm{M}}) = \sigma_{\mathrm{y}}/\theta_{Ty} \cdot (\hat{P}_{\mathrm{y}} - \hat{W}_{\mathrm{M}})$.

Therefore, $\hat{T}_y = \hat{M} - \sigma \mathrm{y}/\theta_{Ty} \cdot (\hat{P}_{\mathrm{y}} - \hat{W}_{\mathrm{M}})$.

We know $\hat{M} = -\mathrm{f} \cdot \mathrm{e}$.

$$\text{Therefore, } \hat{T}_{\mathrm{y}} = -\mathrm{f} \cdot \mathrm{e} - \sigma_{\mathrm{y}}/\theta_{Ty} \cdot (\hat{P}_{\mathrm{y}} - \hat{W}_{\mathrm{M}}). \tag{1.31}$$

Since $\mathrm{e} < 0$ and $\hat{P}_{\mathrm{y}} - \hat{W}_{\mathrm{M}} < 0, \hat{T}_{\mathrm{y}} > 0$. The expanding export sector draws land from the contracting non-traded sector. Now, land released by the contracting food sector is given by Eq. (1.28) and land additionally demanded is given by Eq. (1.31). Adding the two,

$$\hat{T}_{\mathrm{y}} + \hat{T}_{\mathrm{z}} = \frac{[-f \cdot e - \frac{\sigma y}{\theta_{Ty}} \cdot (\hat{P}y - \hat{W}_M)]}{> 0} + \frac{[-\frac{1}{\theta_{Tz}} \cdot (\hat{P}z - \hat{W})]}{< 0}. \tag{1.32}$$

Surplus land services emerge when the expanding export sector demands less additional land than what is released by the contracting non-traded sector, i.e. when

$$\hat{T}_{\mathrm{y}} + \hat{T}_{\mathrm{z}} < 0, \text{or} - \mathrm{f} \cdot \mathrm{e} - \sigma_{\mathrm{y}}/\theta_{Ty} \cdot (\hat{P}_{\mathrm{Y}} - \hat{W}_{\mathrm{M}}) < 1/\theta_{Tz} \cdot (\hat{P}_{\mathrm{z}} - \hat{W}). \tag{1.33}$$

5 Policy Analysis

The model outlined earlier hypothesizes on growing food insecurity, gain for the medium-skilled labour, loss for high-skilled and unskilled workers and the possibility of the emergence of land as an exportable service in a small open economy enforcing stricter environment standards. The outcomes call for appropriate policy interventions. We shall focus on *the role of education* as a policy parameter. Education and skill formation may be addressed in our model by introducing the concept of 'efficiency labour' in each category and analysing the comparative static results. We shall deviate from this approach. Our primary concern is the state of the common property resources which are available to the firms in the non-traded sector as a 'free environment resource'. Expenditure on education and appropriate skill formation in the non-traded sector on rainwater harvesting, use of water bodies, afforestation and use of forest resources, use of canals for both irrigating water in and out etc. raises E (and therefore, A) and even at a given E, raises A. We shall just focus on the latter as a policy to counter the effect of stricter environmental standards in our economy. Let the public expenditure on such skill formation be G. Thus, A = A(E, G) such that $\mathrm{A_E} > 0$ and $\mathrm{A_G} > 0$. We return to Eq. (1.29) and adjust it to accommodate G. We have

$$\hat{Z} = (\mathrm{A_E}/\mathrm{A}) \cdot \mathrm{dE} + (\mathrm{A_G}/\mathrm{A}) \cdot \mathrm{dG} - \frac{1-\alpha}{\theta_{Tz}}(\hat{P}_z - \hat{W}). \tag{1.34}$$

Since dE < 0 by the environmental standards and dG > 0 by public expenditure on appropriate skill formation, it follows from Eq. (1.34) that $\hat{Z} > 0$ if $(A_G/A)\cdot dG$ exceeds the absolute value of $(A_E/A).dE - \frac{1-\alpha}{\theta_{Tz}}.(\hat{P}Z - \hat{W})$. The food security of the nation can be guaranteed despite stricter environmental standards in this small open economy, provided the State spends an appropriate amount of G on appropriate skill formation. Similarly if E is defined as $E = \lambda.E_0$, where E_0 is the available stock of environmental resource and λ is the policy determined rate of its use, lower λ (due to stricter environmental standards) need not reduce E if E_0 is raised by appropriate management of the common property resources. The surplus land services generated in our model may be exploited for raising E_0.

6 Conclusion

Our model is an attempt to capture the complex interactions between expanding trade and growing concern for the environment. The contention of the model is that the countries of the South are seriously challenged by the developments of the last two decades in this respect. While each nation is unique in its own right, our model tries to isolate three fundamental developments, namely, growing food insecurity, increasing wage gap and the possibility of exportable land services. Each issue is challenging, and therefore, the ongoing trade–environment debate would not become comprehensive unless such issues are identified and addressed by proper policy interventions. The model has highlighted some such policies. The assumptions of the analysis may be a point of departure for future scrutiny of the results.

7 Mathematical Appendix

a. Equation (1.10) of the text is derived as follows:

Cost minimization implies that the marginal rate of factor substitution equals the factor–price ratio, i.e.

$$\frac{dk}{ds} = -\frac{W_s}{r},$$

$$\text{or } W_s \cdot ds + r \cdot dk = 0,$$

$$\text{or } \frac{w_s \cdot s}{p_x} \cdot \frac{ds}{s} + \frac{r \cdot k}{p_x} \cdot \frac{dk}{k} = 0,$$

$$\text{or } \theta_s \cdot \hat{s} + \theta_k \cdot \hat{k} = 0.$$

Similarly, the Eqs. (1.11) and (1.12) are derived.

b. Equation (1.13) is derived as follows:

The zero-profit condition in Sector-X implies that the price of X equals the unit cost of X.

$$\therefore \quad P_x = W_s \cdot s + r \cdot k.$$

$$\therefore \quad dP_x = (dW_s) \cdot s + W_s(ds) + (dr) \cdot k + r \cdot (dk).$$

$$\therefore \quad \frac{dp_x}{p_x} = \frac{w_s \cdot s}{p_x} \cdot \frac{dw_s}{w_s} + \frac{w_s \cdot s}{p_x} \cdot \frac{ds}{s} + \frac{r \cdot k}{p_x} \cdot \frac{dk}{r} + \frac{r \cdot k}{p_x} \cdot \frac{dr}{r}.$$

$$\text{Or, } \hat{P}_x = \theta_s \cdot \hat{W}_s + \theta_s \cdot \hat{s} + \theta_k \cdot \hat{r} + \theta_k \cdot \hat{k}.$$

$$\text{By cost minimization, } \theta_s \cdot \hat{s} + \theta_k \cdot \hat{k}. = 0.$$

$$\therefore \hat{P}_x = \theta_s \cdot \hat{W}_s + \theta_k \cdot \hat{r}.$$

Now, $W_s = W_s{}^o \cdot P_z$.

$$\therefore \hat{W}_s = \hat{P}_z.$$

$$\therefore \hat{P}_x = \theta_s \cdot \hat{P}_z + \theta_k \cdot \hat{r}.$$

Equations (1.14) and (1.15) are derived similarly.

c. Equation (1.22) is derived as follows:

$$\text{Total wage income} = W_s \cdot S + W_M \cdot M + W \cdot L = \text{I (as in the text)}.$$

$$\text{Now, } P_z \cdot A(E) \cdot \left(\frac{a_{tz}}{l}\right)^{1-\alpha} = \frac{\beta(W_s \cdot S + W_M \cdot M + W \cdot L)}{L}.$$

$$\therefore \hat{P}_z + \widehat{A(E)} + (1-\alpha) \cdot (\widehat{a_{tz}} - \hat{l}\,)$$

$$= \frac{d[\beta(W_s \cdot S + W_M \cdot M + W \cdot L)]}{\beta(W_s \cdot S + W_M \cdot M + W \cdot L)} - \hat{L}$$

$$= \frac{d(W_s \cdot S + W_M \cdot M + W \cdot L)}{I} - \hat{L}$$

$$= \frac{W_s}{I} \cdot dS + \frac{S}{I} \cdot dW_s + \frac{W_M}{I} \cdot dM + \frac{M}{I} \cdot dW_M + \frac{W}{I} \cdot dL + \frac{L}{I} \cdot dW - \hat{L}$$

$$= \tfrac{W_s \cdot s}{I} \cdot \tfrac{ds}{s} + \tfrac{W_s \cdot s}{I} \cdot \tfrac{dWs}{W_s} + \tfrac{W_M \cdot M}{I} \cdot \tfrac{dM}{M} + \tfrac{W_M \cdot M}{I} \cdot \tfrac{dW_M}{W_M} + \tfrac{W.L}{I} \cdot \tfrac{dL}{L} + \tfrac{W.L}{I} \cdot \tfrac{dW}{W} - \hat{L}$$

$$= a\hat{s} + a \cdot \hat{W}_s + b\hat{M} + b\hat{W}_M + c \cdot \hat{L} + c \cdot \hat{W} - \hat{L}, \text{as in the text.}$$

d. Equation (1.25) is derived as follows:

We have $\hat{P}_z = g + h \cdot \widehat{W_M} + i \cdot \widehat{W} + j \cdot \hat{R} - n \cdot \widehat{A(E)}$, by Eq. (1.24). We know,
By Eq. (1.20), $\hat{W}_M = \hat{P}_y + \frac{\theta_{Ty}}{\theta_M} \cdot (\hat{P}_y - \hat{R}) = \frac{1}{\theta_M} \cdot \hat{P}_y - \frac{\theta_{Ty}}{\theta_M} \cdot \hat{R}$,
and by Eq. (1.21)., $\hat{W} = \hat{P}_z + \frac{1-\propto}{\propto} \cdot (\hat{P}_z - \hat{R}) = \frac{1}{\propto} \hat{P}_z - \frac{1-\propto}{\propto} \hat{R}$,

$$\text{From } \hat{W} = \frac{1}{\propto} \cdot \hat{P}_z - \frac{1-\propto}{\propto} \cdot \hat{R},$$

$$\frac{1}{\propto} \hat{P}_z = \hat{W} + \frac{1-\propto}{\propto} \cdot \hat{R}.$$

$$\text{Or, } \hat{P}_z = \propto \cdot \hat{W} + (1-\propto) \cdot \hat{R}.$$

Using this expression of $\hat{P}_Z$ in equation Eq. (1.24),

$$\propto \widehat{W} + (1-\propto).\hat{R} = g + h \cdot \widehat{W}_M + i \cdot \widehat{W} + j \cdot \hat{R} - n \cdot \widehat{A(E)}.$$

$$\therefore\ (\alpha - i) \cdot \widehat{W} = g + h \cdot \widehat{W}_M + (j + \alpha - 1) \cdot \hat{R} - n \cdot \widehat{A(E)}.$$

Now, replacing $\hat{W}_M$ by $\frac{1}{\theta_M} \hat{P}_Y - \frac{\theta_{Ty}}{\theta_M} \cdot \hat{R}$,

$$(\propto - i)\widehat{W} = g + h \left[\frac{1}{\theta_M} \widehat{P}_y - \frac{\theta_{Ty}}{\theta_M} \hat{R} \right] + (\propto + j - 1)\hat{R} - n \cdot \widehat{A(E)},$$

$$\text{or, } (\propto - i)\widehat{W} = (g - n \cdot \hat{A}) + \left[(\propto + j - 1) - h \cdot \frac{\theta_{Ty}}{\theta_M} \right] \hat{R} + \cdot \frac{h}{\theta_M} \hat{P}_y, \textit{using } \hat{A} \text{ for } \widehat{A(E)}.$$

$$\therefore\ \hat{W} = \frac{g - n\hat{A}}{a - i} + \frac{(\propto + j - 1) - h \cdot \frac{\theta_{Ty}}{\theta_M}}{a - i} \hat{R} + \frac{h}{\theta_M (a - i)} \cdot \hat{P}_y, \text{as in the text.}$$

References

Anderson K (1992) The standard welfare economics of policies affecting trade and the environment. In: Anderson K, Blackhurst R (eds) The greening of world trade issues. University of Michigan Press, Ann Arbor

Bovenberg LA, Smulders S (1995) Environmental quality and pollution augmenting technical change in two-sector endogenous growth model. Journal of Public Economics 57(3):369–391

Brander JA, Taylor MS (1997) International trade and open access renewable resources: the small open economy case. Canadian Journal of Economics 30(3):526–552

Chichilnisky G (1994) North–South trade and the global environment. Am Econ Rev 84

Copeland BR (1994) International trade and the environment policy reform in a polluted small economy. Journal of Environmental Economics and Management 26(1):44–65

Copeland BR, Taylor MS (1994) North–South trade and the environment. Q J Econ 109

Copeland BR, Taylor MS (1995) Trade and transboundary pollution. American Economic Review 85(4):716–737

Marjit S, Acharyya R (2001) International trade, wage inequality and the developing economy—A general equilibrium approach. Research Monograph

Markkusen J (1975) Cooperative control of international pollution and common property resources. Quarterly Journal of Economics 89(4):618–632

Chapter 2
Accumulation of Capital for Pollution Abatement and Immiserizing Growth—A Theoretical Result for Developing Economies

Rupamanjari Sinha Ray

1 Introduction

In an open economy, when growth is driven by factor accumulation, the economy gains by way of expansion of domestic income at constant terms of trade (TOT), but the post-growth equilibrium international TOT may go against the growing nation if its growth is concentrated heavily in its exportable sector (ultra-export-biased growth), and the level of national welfare of the growing economy in the post-growth scenario may turn out to be less than its pre-growth scenario, if the loss in TOT is stronger than the income gains due to growth for such an economy. This is the phenomenon of *immiserizing growth* as coined by Bhagwati (1958). In the standard Heckscher–Ohlin framework for an open economy, the results of changes in welfare gains and losses due to factor accumulation hold true when the international equilibrium remains stable both before and after growth and there is free trade between nations and the markets are perfectly competitive.

Concerns for the environment in an economy have received articulated attention by economic analysts in recent years, who point to the various forms of externalities that pollution or environmental degradation of any form emits either through consumption or through production. These externalities have opened up the cases for different forms of domestic distortions and the trade interventions and environmental interventions reflecting alternative forms of lobbying by the interested group, and the scope for devising an optimum framework for resolving such conflicts. Under the new era of climate change, it is interesting to ask whether global environmental problems such as global warming and climate changes could be tackled by a shift of capital from developed to developing nations. The developed nations have already entered into a process of adopting eco-friendly technologies, particularly the economies of the European Community. The USA, being the highest contributor of global warming, had not ratified the Kyoto Protocol which requires that the countries must reduce its carbon emission below the 1990 level. Thus, there is room for

R. S. Ray (✉)
Management Development Institute, Gurgaon, Haryana, India
e-mail: rupamanjari@mdi.ac.in

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_2,

improvement within the developed economies from an environmental perspective, but that has to be accompanied with loss in production. But the damages that have been made to the environment due to the process of rapid industrialization in developed countries are irrevocable as far as climate changes are concerned. However, any further damage can be prevented through international cooperation, and one such area of cooperation is to allow for movement of foreign capital into developing countries for abatement of pollution. We shall use a two-sector general equilibrium framework for an open economy with environmental pollution acting as an externality to examine the conditions of *immiserizing* growth in developing countries under two situations: (1) domestic capital accumulation for pollution abatement purposes and (2) foreign capital accumulation for pollution abatement purposes.

From the developing country's perspective, the driver for the companies may be earning carbon credits without diverting or less diverting the resources away from production. This would lead to a steady growth process of the emerging economies. However, the question that remains unanswered is the welfare aspect in terms of using foreign capital for carbon emission purposes and using domestic capital for the same purpose. This chapter is limited to this comparative study of conditions for immiserizing growth patterns through trade when domestic and foreign capital is used for pollution abatement purposes. The environmental distortion is expressed as the cost borne by the society to combat the problems of pollution. We keep the analysis simple in terms of general equilibrium model of international trade with full employment of resources as developed among others by Kemp (1969) and Hazari and Sgro (1983).

In case of domestic labour and capital accumulation, we observe what would be the conditions that are needed for welfare improvement in the economy. When the Rybczynski result is valid, then capital accumulation in the economy would lead to welfare gain only when expansion of capital-intensive sectors and improvement in TOT are larger than the contraction of labour-intensive sectors and the loss due to degradation of environmental quality. When the Rybczynski result does not hold, then, under the Marshall–Lerner condition, the economy faces a net welfare loss as the contraction of capital-intensive sectors and deterioration in TOT adds to the loss due to contraction of labour-intensive sectors and degradation of environmental quality. In this case, it is to be noted that environmental degradation, to a large extent, could not be overcompensated by any production gains or gains in TOT. This leads to immiserizing growth. The best policy to avoid immiserizing growth, in this case, is to impose environmental tax and to give production subsidy to the capital-intensive sectors such that TOT also moves in favour of the domestic country. In the case of labour accumulation, we find a different story when the Rybczynski result does not hold good; there would be undoubtedly a welfare gain. Thus, an economy may improve when the accumulation of factor, which generates less environmental degradation, would lead to welfare gains. Even if the factor is more pollution intensive, then also the economy gains from its accumulation provided the impact on environmental quality is not so considerable. However, immiserizing growth would be inevitable when the second sector contracts more unless environmental tax is imposed on sector I and a subsidy is given to sector II.

When domestic and foreign capital accumulation for pollution abatement purposes are adopted by the country separately, then by comparing the two results it is inferred that an economy's welfare improves more when the economy invites foreign capital for pollution abatement purposes rather than siphoning away part of domestic capital for such purposes. However, a prerequisite to the foreign capital accumulation for pollution abatement is that the Marshall–Lerner condition should be satisfied. However, this is a necessary condition but not a sufficient condition as welfare improvement also depends on the utilization capacity of foreign capital and the rate of interest at which foreign capital is being provided. In this context, it is to be noted that international cooperation among countries for common interests of improving the environment requires that developing countries be provided financial assistance by the developed countries to meet the international environmental standards. Financial assistance in the form of foreign capital for pollution abatement at low interest rates could improve the welfare of least developed countries (LDCs), both in terms of environmental quality and in terms of net production gain. Foreign capital technologies are assumed to be environment saving. However, if the interest rates were too high to be repatriated by the gains in TOT and production, then immiserizing growth would take place. The best policy is a tax policy. Not only should an environmental tax be imposed but also a production tax on both the sectors should be imposed. When TOT moves against the country, then immiserizing growth could be avoided by imposing environmental tax on the price of sector I only. Domestic capital accumulation for pollution abatement in developing countries would siphon off the domestic capital resources, which could otherwise be used in the production process. The Rybczynski theorem's validity would lead to immiserizing growth when contraction of a labour-intensive sector becomes more than the improvement in environmental quality and expansion in a capital-intensive sector taken together. Then the best policy to overcome such distortion is to impose environmental tax and to provide production subsidies to both the sectors with the proceedings of the tax.

The plan of the present chapter is as follows. Section 2 describes the model. Section 3 describes the effects of accumulation of domestic factors. Section 4 analyses the effects of introducing foreign capital in the model, and the conditions for immiserizing growth. Section 5 explains how the country is benefited if it utilizes foreign capital instead of its own capital for pollution abatement. Section 6 contains the concluding remarks.

2 The Model

The impact of capital accumulation on the welfare of an economy may lead to immiserizing growth in the presence of some externality. In the present chapter, the possibility of immiserizing growth in the presence of environmental resource as a factor of production is checked, firstly, when capital accumulation occurs domestically with a part being used for the purpose of pollution abatement, and secondly, when foreign capital is introduced in the model only for the purpose of pollution abatement.

The model developed here is a simple general equilibrium Heckscher–Ohlin model of trade with environment as the third factor of production. Including Meadetype production externality, the changes in the model have been observed in the existing literature of international trade. Unlike Judith Dean (1999) and Ramon Lopez (1994), weak separability between environmental resource and conventional factors of production is not assumed.

The production function is defined as follows:

$$X_i = F(K_i, L_i, E_i), \text{where}$$

K_i Capital input of the i[th] sector,
L_i Labour input of i[th] sector,
E_i Environmental resource of the i[th] sector and
X_i Output level of the i[th] sector.

The following assumptions are made about the above production function:

1. It follows constant returns to scale. Thus, the production function can be written as follows:

$$X_i = L_i f(k_i, e_i), \tag{2.1}$$

where k_i is the capital–labour ratio and e_i is the environmental resource and labour ratio.

2. The assumptions on the marginal productivity functions may be stated as follows: The marginal productivity of labour is

$$\frac{\partial X_i}{\partial L_1} = f_i - k_i f_{ik_i} > 0, \tag{2.2}$$

where $f_{ik_1} = \frac{\partial X_i}{\partial k_i}$.

The marginal productivity of capital is

$$\frac{\partial X_i}{\partial K_i} = f_{ik_1} > 0. \tag{2.3}$$

The marginal productivity of environmental resource is

$$\frac{\partial X_i}{\partial E_i} = f_{ie_1} > 0. \tag{2.4}$$

The analysis has been carried out in an open economic framework. The home country is assumed to be labour-abundant compared to the foreign country, which is capital-abundant. Commodity 1 is assumed to be more labour-intensive, and commodity 2 is assumed to be more capital-intensive. Thus, (1) $L_1 > L_2$ and (2) $K_1 < K_2$. Thus, country 1 will export commodity 1 and import commodity 2. We also assume that there is a difference in the pre-trade price ratios of the two countries, such that $\left(\frac{p_2}{p_1}\right)_1 > \left(\frac{p_2}{p_1}\right)_2$, where p_1 and p_2 are prices of commodities 1 and 2. Commodity 2 is assumed to use more environmental resources compared to commodity 1. Production of commodity 2 creates more pollution than commodity 1 as commodity 2 is capital-intensive and commodity 1 is labour-intensive. Thus, we assume that the capital technologies in the countries are such that more pollution is created by the use of capital in both the countries. Thus, $E_1 < E_2$.

The balance of trade (BoT) equilibrium is

$$B_1 = pM_2, \tag{2.5}$$

where

B_1 is the export of commodity 1, $B_1 = X_1 - C_1$,
M_2 is the import level of commodity 2, $M_2 = C_2 - X_2$,
p is the relative price level of importable, $p = \frac{p_2}{p_1}$ and
p_1 and p_2 are prices of commodities 1 and 2.

The social welfare function is defined as follows:

$$W = W(C_1, C_2, E), \tag{2.6}$$

where

C_1 is the domestic consumption level of commodity 1,
C_2 is the domestic consumption level commodity 2 and
E is the level of environmental quality.

Competition prevails in all markets. Thus, wage rate w and rental rate r appear exogenous to the model and are expressed in terms of commodity 1. Thus, we can conclude, w equals the marginal productivity of labour and r equals the marginal productivity of capital. Two equations are, thus, obtained:

$$w = f_1 - k_1 f_{1k_1} - e_1 f_{1e_1} = p(f_2 - k_2 f_{2k_2} - e_2 f_{2e_2}), \tag{2.7}$$

$$r = f_{1k_1} = p f_{2k_2}. \tag{2.8}$$

From the above specification of the model, it could be observed what the general equilibrium condition holds, i.e. whether the autarky TOT is tangent to the production possibility curve (PPC) at autarky equilibrium.

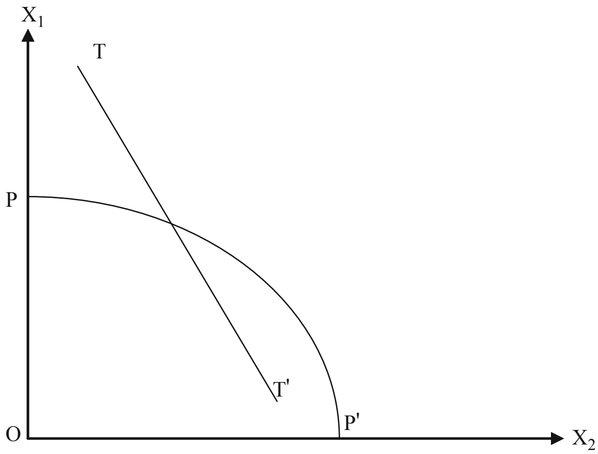
Slope of the PPC is defined by $\frac{dX_1}{dX_2}$, and the slope of the TOT line is $(-p)$, which is nothing but $\left(-\frac{p_2}{p_1}\right)$.

From the marginal productivity of labour, capital and environmental resource conditions, we obtain

$$\frac{dX_1}{dX_2} = -\frac{p(1+\alpha_1)}{(1+\alpha_2)} = -p\gamma, \tag{2.9}$$

where $\alpha_i = \frac{f_{ie_i} dE_i}{F_{iK_i} dK_i + F_{iL_i} dL_i}$ and $\gamma = \frac{1+\alpha_1}{1+\alpha_2}$. Since capital intensive sector is assumed to be more environmental resource-intensive, we assume $\alpha_1 < \alpha_2$. Thus, $\left|\frac{dX_1}{dX_2}\right| < p$. Therefore, we see from Fig. 2.1 that at equilibrium the TOT TT$'$ is intersecting the PPC PP$'$.

Fig. 2.1 Initial equilibrium



3 Conditions for Immiserizing Growth in Case of Domestic Factor Accumulation

To analyse the conditions of immiserizing growth under domestic factor accumulation, under the model presented in Sect. 1, only the effect of capital accumulation in the economy is noticed in this chapter, as our main objective of the study has been capital accumulation for the pollution abatement purpose. Therefore, in this chapter we ignore effects of the second factor, i.e. labour accumulation and concentrate only on effects of capital accumulation, K_D.

We define welfare function, $W = W(C_1, C_2, E)$ from Eq. (2.6). Taking total differentials,

$$dW = W_1 dC_1 + W_2 dC_2 + W_3 dE.$$

Dividing both sides by W_1 we get

$$\frac{dW}{W_1} = \frac{W_1 dC_1 + W_2 dC_2}{W_1} + \frac{W_3 dE}{W_1} = \frac{dC_1 + pdC_2 + W_3 dE}{W_1}, \tag{2.10}$$

where $p = p_2/p_1$.

From budget constraint, we know that $(X_1 + pX_2) = (C_1 + pC_2)$.

Taking total differential, we get $dX_1 + pdX_2 + X_2 dp = dC_1 + pdC_2 + C_2 dp$

or
$$dC_1 + pdC_2 = dX_1 + pdX_2 - M_2 dp, \tag{2.11}$$

where $M_2 = C_2 - X_2$.

Putting Eq. (2.11) in Eq. (2.10), we get

$$dW/W_1 = dX_1 + pdX_2 - M_2 dp + (W_3/W_1)dE.$$

Differentiating with respect to K, capital accumulation,

$$(1/W_1)(dW/dK) = dX_1/dK + p(dX_2/dK) - M_2(dp/dK) + (W_3/W_1)(dE/dK). \tag{2.12}$$

Now from the Rybczynski result, in the presence of environmental resource as a factor of production, it is observed that X_1 decreases with the increase in capital endowment:

$$dX_1/dK = (f_1 - e_1)(dL_1/dK) + f_1 e_1(dE_1/dK) < 0.$$

Effects on X_2 are ambiguous as $dX_2/dK = (f_2 - e_2)(dL_2/dK) + f_2 e_2(dE_2/dK)$, and

$$dX_2/dK \quad \text{will be } > 0 \qquad \text{if} \qquad |(f_2 - e_2)(dL_2/dK)| > |f_2 e_2(dE_2/dK)|.$$

The changes in the TOT could be obtained as follows:

Differentiating the balance of payments equilibrium equation with respect to capital endowment, K, we get

$$\frac{dp}{dK} = \frac{p\dfrac{\partial M_2}{\partial K}}{(\eta_x + \eta_m - 1)M_2}, \tag{2.13}$$

where η_x and η_m are the import elasticities of the two countries.

M_2 is defined[1] as $M_2 = D_2(p, Y) - X_2(p, Y)$,

where D_2 is the domestic demand of the second commodity and X_2 is total output level of the second commodity.

Thus, initially p remaining constant, the partial derivative of M_2 with respect to K is $\frac{\partial M_2}{\partial K} = \frac{\partial D_2}{\partial Y}\frac{dY}{dK} - \frac{\partial X_2}{\partial K}$, and using Eq. (2.9),

$$\frac{dY}{dK} = \frac{dX_1}{dK} + p\frac{dX_2}{dK} = p(1-\gamma)\frac{dX_2}{dK}.$$

Then $\frac{\partial M_2}{\partial K} = [pm_h(1-\gamma) - 1]\frac{dX_2}{dK}$,

where m_h is the marginal propensity to consume, $\frac{\partial D_2}{\partial Y}$.

Thus, $\frac{dp}{dK} = \frac{p[pm_h(1-\gamma)-1]}{(\eta_x+\eta_m-1)M_2}\frac{dX_2}{dK}$. For simplicity, initially we put $p = 1$. Then

$$\frac{dp}{dK} = \frac{m_h(1-\gamma) - 1}{(\eta_x + \eta_m - 1)M_2}\frac{dX_2}{dK}. \tag{2.14}$$

According to Marshall–Lerner condition, the sum of import elasticities is greater than unity and both γ and m_h are less than unity.[2] Thus, from the above equation, we get that if the production of X_2 increases (decreases), then the TOT would move in favour of (against) country 1.

---

[1] This definition is obtained from B. Hazari (1983) where the domestic demand is a function of price and income and the production of importables is also a function of price and income level.

[2] $\gamma < 1$, by assumption as $\alpha_1 < \alpha_2$. m_h is the mpc which is less than unity.

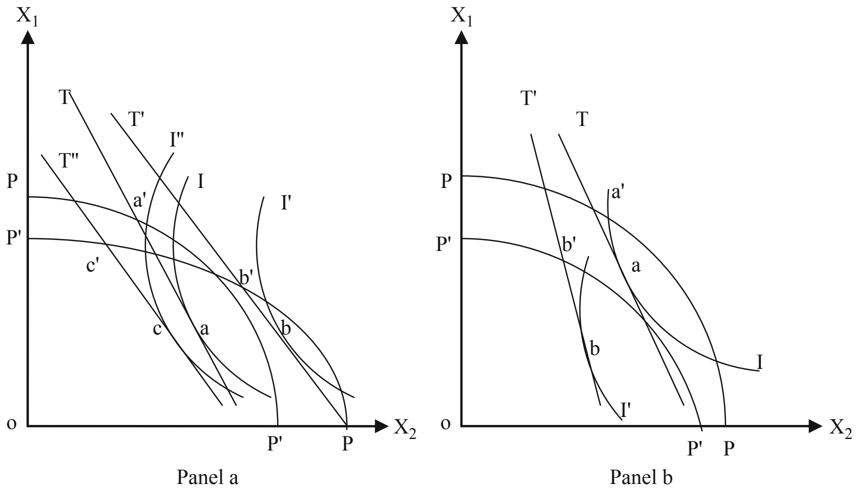


Fig. 2.2 Depicting conditions of immiserizing growth under domestic capital accumulation

As production of import sector rises (X_2), then $p = p_2/p_1$ will fall. Therefore, TOT will move in favour of the home country. Now E will fall as more capital is accumulated.

Therefore, $(dE/dK) < 0$.

Thus,

$$(1/W_1)(dW/dK) = dX_1/dK + p(dX_2/dK) - M_2(dp/dK) + (W_2/W_1)(dE/dK) > 0,$$
only when $\quad dX_1/dK + (W_1/W_2)(dE/dK)\,|<|\,p(dX_2/dK) - M_2(dp/dK)|$.

So if this condition is satisfied, the country will improve its welfare with capital accumulation. Thus, condition of immiserizing growth is

$|dX_1/dK + (W_1/W_2)(dE/dK)| > |p(dX_2/dK) - M_2(dp/dK)|$. This is depicted in Fig. 2.2a.

From Fig. 2.2a, we observe that initial TOT, T, intersects initial PPC, PP, at a′ and is tangent to the initial indifference curve at point a. Welfare will improve with capital accumulation when falling TOT, T′, intersects PPC, P′P′, at point b′. However, if the falling TOT shows a large deterioration along with environmental degradation, then immiserizing growth condition is captured at point c where welfare reduces to I″ from I and the TOT, T″ intersects PPC at point c′.

If $dX_2/dK < 0$, i.e. if $|(f_2 - e_2)(dL_2/dK)| < |f_2 e_2(dE_2/dK)|$, then when both X_1 and X_2 will fall, the country's export supply falls and import demand rises. TOT will move against the country under the Marshall–Lerner condition. In that case, $(1/W)(dW/dK) < 0$, as each term in Eq. (2.12) becomes negative. Thus, if the Rybczynski theorem (Reference Appendix) is not valid for the capital-intensive sector, then capital accumulation leads to immiserizing growth. This is shown in Fig. 2.2b, where immiserizing growth is captured by the shift of welfare indifference curve from I to I′. Here PPC moves from PP to P′P′. The best policy to avoid immiserizing growth would be to impose environmental tax and give production subsidies to the sectors to implement environment-friendly technology in order to expand production of both the sectors.

If the Rybczynski theorem is valid, then in the presence of the environmental resource factor, welfare improves by domestic capital accumulation only when the expansion in the capital-intensive sector and improvement in the TOT is more than the absolute contraction of the labour-intensive sector and deterioration in environmental quality. Under the Marshall–Lerner condition, there would be immiserizing growth if the Rybczynski theorem were not valid.

4 Effects of Foreign Capital Introduced in the Economy for Pollution Abatement Purposes

In the earlier sections, capital available in the economy is the domestic capital, K. When foreign capital is introduced in the economy, then total capital K is divided into two parts—domestic capital K_{D} and foreign capital K_{F}. Foreign capital can only be used for pollution abatement purposes. Domestic capital is used for production purposes. Then the question arises whether the economy's welfare gains due to pollution reduction leading to more production would be wiped out by the welfare losses due to deterioration in TOT and repatriation of foreign capital. Thus, the economy would have either a net gain or a net welfare loss.

The production function is rewritten as $X_i = F_i(K_{Di}, Li, E_i) = L_i f_i(k_{Di}, e_i)$. Assumption: Foreign capital invested for pollution abatement is partly or fully absorbed by the society, $dE = \theta dK_F (0 < \theta \leq 1)$. Environmental quality is partially or fully improved according to the economy's utilization capacity of foreign capital for pollution abatement purposes. Thus, θ is defined as the parameter for utilization capacity.

4.1 Effect on Output Levels

Differentiating the production function with respect to K_F,

$dX_1/dK_F = f_1 e_1 (dE_1/dK_F)$.

Similarly, $dX_2/dK_F = f_2 e_2 (dE_2/dK_F)$.

The effects of improvement in environmental quality on the employment of environmental resources by the sectors are positive, as productivity of environmental resources is enhanced as these resources become less pollution generating. Thus, $\frac{dE_1}{dK_F} > 0, \frac{dE_2}{dK_F} > 0$,

$$\therefore \frac{dX_1}{dK_F} = L_1 f_1 e_1 \left(\frac{dE_1}{dK_F} \right) > 0 \tag{2.15}$$

$$\frac{dX_2}{dK_F} = L_2 f_2 e_2 \left(\frac{dE_2}{dK_F} \right) > 0 \tag{2.16}$$

Therefore, gains from trade due to production should increase as the export sector expands and the import sector's expansion reduces imports.

4.2 Effects of Foreign Capital Growth on Balance of Trade (BoT) Equilibrium

BoT equilibrium Eq. (2.5) is rewritten as
$B_1 = pM_2 + iK_F$.

Thus, from BoT equilibrium, we can infer that trade surplus in the economy should cover the interest payments on capital account, i.e. capital deficit of iK_F. Another way to interpret the changes in BoT equilibrium is that total production must exceed total consumption to cover the interest payments to the foreign country. Whether the country's trade balance improves or deteriorates with the import of foreign capital can be checked in the following way. Trade balance will improve only when

$$\left(\frac{dB_1}{dK_F}\right) \geq \left[\frac{d(pM_2 + iK_F)}{dK_F}\right]$$

or

$$\frac{dB_1}{dp}\frac{dp}{dK_F} \geq p\frac{\partial M_2}{\partial K_F} + p\frac{\partial M_2}{\partial p}\frac{dp}{dK_F} + M_2\frac{dp}{dK_F} + i \tag{2.17}$$

Dividing throughout by $M_2(dp/dK_F)$ we get,

$$(\eta_m + \eta_x - 1) \geq \eta_m + i,$$

or $\eta_x \geq i + 1$, where η_m is the import elasticity of demand for the second commodity and η_x is the elasticity of exports.

The balance of trade would only improve when the difference between the export elasticity and rate of interest, at which the home country is borrowing foreign capital, is greater than unity.

4.3 Effects on National Income

National income in terms of exportable is defined as $Y = X_1 + pX_2$. Out of this national income, foreign capital, which is taken only for the purpose of pollution abatement, has to be repaid to the foreigners in terms of interest payments, iK_f. Thus, national income should be defined as, $Y = X_1 + pX_2 - iK_F$. So, TOT remaining constant, the partial effect on national income is, using Eq. (2.9),
$\frac{\partial Y}{\partial K_F} = \frac{\partial X_1}{\partial K_F} + p\frac{\partial X_2}{\partial K_F} - i = p(1-\gamma)\frac{\partial X_2}{\partial K_F} - i$. Thus, national income would rise only if the gains from production are greater than the rate of interest at which the foreign capital is repatriated.

4.4 Effects on TOT

From Eq. (2.17), $\frac{dp}{dK_F} = \frac{p(\partial M_2/\partial K_F)+i}{(\eta_x+\eta_m-1)M_2}$,

M_2 is defined as $M_2 = D_2(p, Y) - X_2(p, Y)$.

Thus, initially p remaining constant, the partial derivative of M_2 with respect to K_F is $\frac{\partial M_2}{\partial K_F} = \frac{\partial D_2}{\partial Y}\frac{dY}{dK_F} - \frac{\partial X_2}{\partial K_F}$,

or $\frac{\partial M_2}{\partial K_F} = m_h[p(1-\gamma)-1]\frac{dX_2}{dK_F} - m_h i$. Thus, initially putting $p = 1$, for simplicity,

$$\frac{dp}{dK_F} = \frac{(m_h(1-\gamma)-1)\frac{dX_2}{dK_F} + (1-m_h)i}{(\eta_x+\eta_m-1)M_2}. \tag{2.18}$$

Under the Marshall–Lerner condition, the change in TOT remains ambiguous as $m_h < 1$, $\gamma < 1$ and $\frac{dX_2}{dK_F} > 0$. This implies that the second term in the numerator is positive while the first term is negative. Thus, TOT will move in favour of the home country (country 1) with foreign capital introduced in the economy for pollution abatement, only if $\left|(m_h(1-\gamma)-1)\frac{dX_2}{dK_F}\right| > |(1-m_h)i|$. If this condition is not satisfied, then the TOT moves against the country.

4.5 Conditions for Immiserizing Growth Under Foreign Capital (K_F) Accumulation

In the basic model, we have defined social welfare function, $W = W\ (C_1, C_2, E)$. The main objective of this exercise is to observe the effects of foreign capital invested for the purpose of pollution abatement only on the level of social welfare:

$W = W\ (C_1, C_2, E)$.

Taking total differential and dividing both sides by W_1, we get,

$$\begin{aligned} dW/W_1 &= dC_1 + (W_2/W_1)dC_2 + (W_3/W_1)dE \\ &= dC_1 + pdC_2 + (W_3/W_1)dE. \end{aligned} \tag{2.19}$$

From Balance of Payments (B.O.P.) equilibrium condition, we know that,

$$(X_1 + pX_2) - (C_1 + pC_2) = iK_F,$$

or
$$(C_1 + pC_2) = X_1 + pX_2 - iK_F,$$

or
$$dC_1 + pdC_2 + C_2dp = dX_1 + pdX_2 + X_2dp - idK_F,$$

or
$$dC_1 + pdC_2 = dX_1 + pdX_2 - M_2dp - idK_F. \tag{2.20}$$

Combining Eqs. (2.19) and (2.20),

$$(dW/W_1) = dX_1 + pdX_2 - M_2dp - idK_F + (W_3/W_1)dE.$$

Therefore,

$$\begin{aligned}(1/W_1)\,(dW/dK_F) &= dX_1/dK_F + p\,(dX_2/dK_F) - M_2\,(dp/dK_F) - i \\ &\quad + (W_3/W_1)\,(dE/dK_F) \\ &= dX_1/dK_F + p\,(dX_2/dK_F) - M_2\,(dp/dK_F) - i \\ &\quad + (W_3/W_1)\,\theta,\end{aligned}$$

as $dE = \theta dK_F$.

Now, $dX_1/dK_F + pdX_2/dK_F > 0$, as from Eqs. (2.19) and (2.20), dX_1/dK_F and $dX_2/dK_F > 0$.

From Eq. (2.18), dp/dK_F is ambiguous. Now let us first assume $dp/dK_F < 0$, i.e. it moves in favour of the country.

Under this condition, $dX_1/dK_F + p(dX_2/dK_F) - M_2(dp/dK_F) > 0$.

But welfare will increase only when $(1/W_1)(dW/dK_F) > 0$, i.e. $dX_1/dK_F + p(dX_2/dK_F) - M_2(dp/dK_F) + (W_3/W_1)\,\theta > i$. So gains from production increase and improvement in TOT under the Marshall–Lerner condition should be greater than the rate of interest, which the home country pays the foreign country by using the foreign capital for the purpose of pollution abatement. If this condition is not satisfied, then that leads to immiserizing growth. This is described with the help of Fig. 2.3a.

From Fig. 2.3a, the initial PPC, PP, intersects the initial TOT line, T, at a′. The initial consumption takes place at point a where the T is tangent to the welfare indifference curve I. With the initiation of pollution abatement with foreign capital, the PPC shifts to P′P′ from PP. The TOT, T, shifts to T′ becoming flatter. There would be an overall improvement if

$$dX_1/dK_F + p(dX_2/dK_F) - M_2(dp/dK_F) + (W_3/W_1)\theta > i.$$

This improvement is depicted in the figure by the shift of I to a higher indifference curve I′. But if the above condition is not satisfied, i.e.

$$dX_1/dK_F + p(dX_2/dK_F) - M_2(dp/dK_F) + (W_3/W_1)\theta < i,$$

then, the country would suffer an immiserizing growth, in spite of the expansion in production sectors. The interest payments to the foreigners become so high that it undermines the improvement of production. The TOT also shifts to the left to T″ as interest is high, but is flatter due to the expansion of sectors. The welfare is reduced, shown by the shift of the indifference curve to I″. An environmental tax coupled with production tax should be imposed, such that the PPC shifts to P″P″ and TOT line shifts to T‴, leading to a rise in welfare level to I‴. So the Marshall–Lerner condition alone cannot ensure the country's welfare improvement. The rate of interest at which the foreign capital is repatriated plays an important role in this context. The rate of absorption of foreign capital, θ, by the economy for pollution abatement also plays a major role. Higher the rate of absorption, higher the improvement in environmental quality will be. Therefore, welfare of the economy increases.

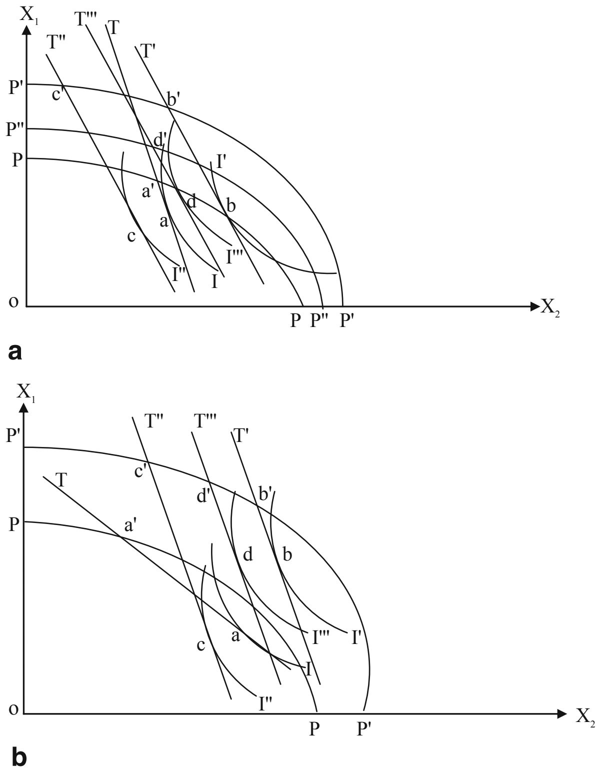


Fig. 2.3 Conditions of immiserizing growth with foreign capital for pollution abatement

If $dp/dK_F > 0$, then the condition for immiserizing growth would be
$dX_1/dK_F + p(dX_2/dK_F)M_2(dp/dK_F) + (W_3/W_1)\theta < i + M_2(dp/dK_F)$.
This is depicted in Fig. 2.3b.

Welfare improvement is shown by the shift of welfare indifference curve from I to I′. However, immiserizing growth is shown by a shift to I″. This could be avoided by reducing TOT. This could be brought about by imposing an environmental tax on the price of sector I, such that the TOT line shifts to the right. On the same PPC, P′P′, the new TOT line, T‴, is flatter than the TOT line, T″, indicating a fall in TOT, thereby shifting the production point to d′ from c′. Thus, effects on TOT becomes one of the major criteria for welfare improvement when foreign capital is introduced in the economy for pollution abatement purposes, along with the rate of interest at which foreign capital is being repatriated.

5 Effects of Domestic Capital for Pollution Abatement Purposes

In Sect. 3 of this chapter, foreign capital for pollution abatement purposes was introduced in the model and how the economy may become better off both in TOT and environmental improvement was analysed, and it was observed that an economy's welfare improves if it is able to repatriate the foreign capital from production and trade improvement due to pollution abatement. However, it would be interesting to observe whether welfare improves more in the case of domestic capital being utilized for pollution abatement or foreign capital being used for the same.

For the analysis, we have introduced some simple changes in the model. Domestic capital is now used for two purposes: (1) production and (2) pollution abatement. Thus, total capital endowment, K, in the economy is divided into two parts—K_p, part of total capital used for production purposes, and K_a, for pollution abatement purposes. Thus, $K = K_a + K_p$.

Assumptions:

1. α part of total capital is used in pollution abatement purpose.
2. μ is the rate of utilization of capital for pollution abatement purposes.

The production function is rewritten as $X_i = F_i(K_{pi}, Li, E_i) = L_i f_i(k_{pi}, e_i)$.

If there is domestic capital accumulation, i.e. if there is an increase in the endowment of domestic capital, then in order to observe the welfare effects, we have to observe the following effects.

5.1 Effects on Output Levels

At first, we try to observe the effects on labour and capital inputs of both the industries. We rewrite the capital endowment equation as follows:

$$\begin{aligned} K &= K_a + K_p \\ &= K_a + K_{p1} + K_{p2}, \qquad \text{where } K_{pi} = \text{capital input in the i}^{\text{th}} \text{ industry} \\ &= k_1 L_1 + k_2 L_2 + K_a \end{aligned}$$

Change in total capital endowment on labour inputs and environmental resources of the industries—

$$dL_1/dK = -(1-\alpha)/(k_2 - k_1),$$

and

$$dL_2/dK = -(1-\alpha)/(k_2 - k_1), \tag{2.21}$$

$$dE/dK = (dE/dK_a)(dK_a/dK) = \alpha\mu. \tag{2.22}$$

dE_1/dK and dE_2/dK are assumed to be positive as improvement in environmental quality has positive effects on the employment of environmental resources, which now have greater productivity.

Using the Rybczynski result (Reference Appendix), we obtain the effects on output levels as follows:

$$(dX_i/dK) = (f_i - e_i)(1-\alpha)(dL_i - dk) + f_i e_i (dE_i/dK).$$

Thus,

$$(dX_1/dK) = -(f_1 - e_1)(1-\alpha)/(k_2 - k_1) + f_1 e_1 (dE_1/dK), \tag{2.23}$$

$$(dX_2/dK) = (f_2 - e_2)(1-\alpha)/(k_2 - k_1) + f_2 e_2 (dE_2/dK). \tag{2.24}$$

The capital-intensive sector expands unambiguously as there is an increase in the productive capital and environmental resource due to domestic capital accumulation. The contraction of the labour-intensive sector due to domestic capital accumulation cannot be unambiguously determined. However, contraction in the labour-intensive sector due to the increase in productive capital endowment in the economy should be more than its expansion due to increase in the environmental resource, which is a result of the improvement in environmental quality as part of the domestic capital being used for the purpose of pollution abatement. If the cost–benefit approach is taken into consideration, then pollution abatement cost should be lower than the production cost. Hence, ($\alpha < 1 - \alpha$), the part of the domestic capital used for the pollution purpose should be less than the part used for production purpose, such that the cost incurred for pollution abatement does not exceed the benefits derived. In such a situation, the Rybczynski result would also hold for the labour-intensive industry.

5.2 *Effects on TOT*

From Eq. (2.14),

$$\frac{dp}{dK} = \frac{m_h(1-\gamma) - 1}{(\eta_x + \eta_m - 1)M_2} \frac{dX_2}{dK}.$$

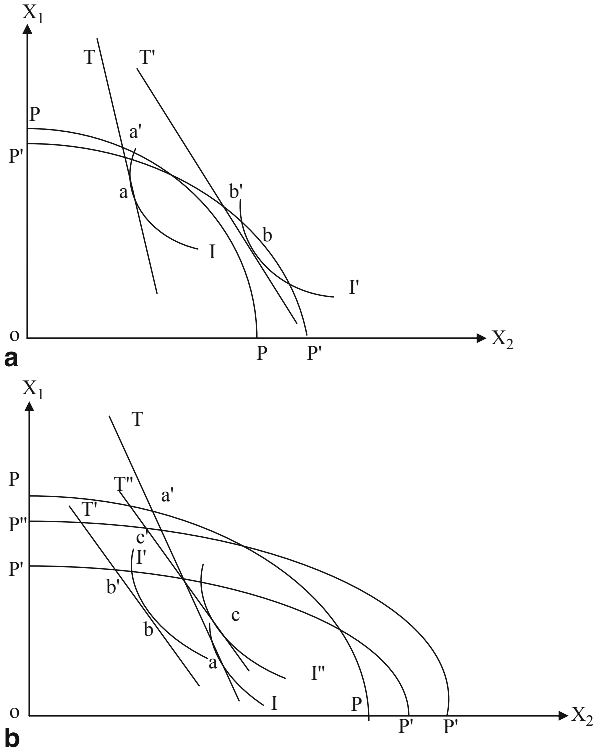
Following the Rybczynski result, the capital-intensive sector will expand and the labour-intensive sector will contract due to domestic capital accumulation. Then the import demand will fall as the importable capital-intensive sector expands. As a result, the price of imported commodity in the world market will fall. However, due to the contraction of labour-intensive sector, export supply will fall, raising the price of exports in the world market. As a result, the TOT (p $= \frac{p_2}{p_1}$) will move in favour of the country.

5.3 *Conditions for Immiserizing Growth Under Domestic Capital (Kp + Ka) Accumulation*

We obtain the condition where the welfare changes are positive. This is done as follows:

$$(1/W_1)(dW/dK) = dX_1/dK + p(dX_2/dK) - M_2(dp/dK) + (W_3/W_1)(dE/dK).$$

Fig. 2.4 Conditions of immiserizing growth under domestic capital for pollution abatement purpose



The first term on the right-hand side is only negative. Other terms are positive. Thus, the condition for immiserizing growth is derived as follows:

$(1/W_1)(dW/dK) > 0$ is the condition for immiserizing growth. This implies,

$p(dX_2/dK) - M_2(dp/dK) + (W_3/W_1)(dE/dK)| > dX_1/dK|$. This implies that absolute decrease in output level of labour-intensive export sectors should be less than the increase in output level of the capital-intensive sectors and the positive TOT effect. This improvement in welfare is shown in Fig. 2.4a.

Improvement in TOT and environmental quality and expansion in capital-intensive sectors should overcompensate for the loss of welfare due to contraction of labour-intensive sectors. Welfare rises from I to I′ and TOT moves from line T to T′. The PPC would shift from PP to P′P′.

However, if the shift of PPC is such that there is a little expansion in TOT and capital-intensive sector, while contraction of the labour-intensive sector is more, then PPC shifts to P′P′ as shown in Fig. 2.4b.

The TOT line becoming flatter shifts to T′ from T and there is a welfare loss shown by the shift of welfare curve from I to I′. This is a case for immiserizing growth. If an environmental tax is imposed and by these proceedings if a production subsidy is given to both the sectors, then contraction of labour-intensive sectors would be less and expansion of capital-intensive sectors would be more. The PPC would shift to P″P″. TOT would shift up to T″T″ as a result of the expansion of capital intensive sector.

5.4 *A Comparative Study of Foreign Capital and Domestic Capital for Pollution Abatement Purposes and its Implications on Carbon Credits*

The studies mentioned earlier have observed the conditions of immiserizing growth if foreign capital and domestic capital are used for pollution abatement purposes. Comparing the two studies, it is inferred that if the world rate of interest is sufficiently low, then a country will be better off if the economy opts for foreign capital for pollution abatement and does not waste its domestic resources. Even if the domestic capital is used, then production expands through the use of eco-friendly technologies that lead to less wastages and improvement in environmental quality. However, welfare improvement is realized when contraction in labour-intensive export sector is less than the expansion in the capital-intensive import sector. The implication here is that a production subsidy is to be given to the labour-intensive sector, and a tax is to be collected from the capital-intensive sector when a part of the gains of adopting eco-friendly technology is siphoned off from the capital-intensive sector and distributed to the labour-intensive sector. There is no doubt that such a redistribution will increase national welfare, but that itself provides a disincentive to the capital-intensive sector of the economy. Therefore, resource diversion and redistribution of gains is neither sustainable nor justified.

If the rate at which foreign capital is repatriated back is kept low, then, however, a higher rate of foreign capital repatriation may be detrimental to the growth and welfare process of the economy. In the example of carbon emissions and carbon credit, use of foreign capital in developing nations may be an ideal strategy wherein the rate of foreign capital repatriation is very low, and, on the other hand, the foreign capital is compensated by the earning of carbon credits, the trading that involves full return on such capital investment. Under such a framework, it seems that foreign capital investment in environmental projects in developing economies may benefit both developed and developing economies, rather than addressing the problem of climate change individually with their respective domestic resources.

6 Conclusion

The chapter had theoretically analysed two different conditions of immiserizing growth under foreign capital and domestic capital used for pollution abatement purposes. The entire theoretical analysis is based on a two-country model. However, it may be extended to a multi-country case. It is better for a developing nation to opt for foreign capital for pollution abatement purposes. In such a case, the benefits gained both economically and environmentally would be higher than the costs. The resources may be used for production purposes rather than being diverted from production sector to undertake expensive eco-friendly projects. However, environmental quality is also improved in the process. However, such initiatives from foreign counterparts need international cooperation. The entire explanation of growth has been based on welfare analysis in terms of environmental quality improvement and BoT situation of a country, which in turn explains gains from trade in terms of TOT and output expansion. It has been shown that a redistribution of gains from trade from polluting capital-intensive sector, that adopts eco-friendly technology, to labour-intensive sector as production subsidy may prevent immiserizing growth in case domestic capital is used for production purpose. However, using foreign capital for improving the environment is a better choice provided the rate of absorption of such foreign capital by the domestic economy is high and rate of repatriation is low.

This model can also explain the theoretical basis of carbon credit earning by developed economies by undertaking green business projects in developing economies by which both economies may be beneficial in terms of economy and environment. The benefits acquired through Clean Development Mechanism of the Kyoto Protocol can also be analysed by this model. Through international cooperation, further damages to the environment and climate change threats can be prevented in the world economy as a whole. This chapter is limited in its theoretical assumptions of a neoclassical general equilibrium set-up. Further extension of the model may be feasible to explore the intergenerational gains from cooperation in a dynamic set-up.

References

Bhagwati J (1958) International trade and economic expansion. American Economic Review 48(5):941–953

Chatterjee B (2000) Environment, trade and globalization—emerging issues and policy options. N.K. memorial lecture delivered at the 4th annual conference of Economic Association of Bihar, Patna. May 28:2000

Chatterjee B, Karmakar AK (1995) Environment and development in an open economy. IEA Conference Volume. Gwalior, December 1995

Chatterjee B, Sen R (2001) Environment standards, WTO and the Indian economy. In: Chadda GK (ed) WTO and the Indian Economy. Deep and Deep, New Delhi

Chichilnisky G (1994) North–south trade and the global environment. Am Econ Rev 84(4):851–873

Copeland BR, Taylor MS (1994) North-south trade and the environment. Q J Econ 109(3):755–787 (August 1994)

Copeland BR, Taylor MS (1997) A simple model of trade, capital mobility, and the environment. NBER Working Paper Series, vol 5898. National Bureau of Economic Research, Cambridge

Dean J (1992) Trade and environment: a survey of the literature. In: Low P (ed) International trade and the environment. World Bank Discussion Papers, vol 159. World Bank, Washington D.C.

Dean J (1999) Testing the impact of trade liberalization on the environment. In: Fredriksson P (ed) Trade, global policy and the environment. World Bank Discussion Papers, vol 402. World Bank, Washington, D.C., pp 55–63

Hazari BR, Sgro PM (1975) Theorems on immiserising (normal growth) in the non-traded goods and wage differentials framework. Southern Econ J 41:515–519

Hazari BR, Sgro PM (1976) Some notes on technical progress in the framework of factor market imperfections and non-traded goods. Aust Econ Pap 25(March):1–43

Hazari BR, Sgro PM (1983) International trade theory in a two country general equilibrium model with traded intermediate goods, Vol 2, published by School of Economics, La Trobe University

Kemp MC (1969) The pure theory of international trade and investment. Prentice Hall, Englewood Cliffs

Kolstad CD (2010) Environmental economics. Oxford University Press, New York

Lopez R (1994) The environment as a factor of production: the effects of economic growth and trade liberalization. J Environ Econ Manag 27:163–184

World B (1992) World development report. World Bank, Washington D.C

Chapter 3
Optimal Entry Mode for Multinationals with Possibility of Technology Diffusion

Nilanjana Mitra and Tanmoyee Banerjee (Chatterjee)

1 Introduction

The relationship between strong intellectual property rights (IPR) protection, foreign direct investment (FDI), and technology transfer is an issue of great interest. Protection of IPR is becoming a challenge to less-developed and developing countries like India. An interesting report in *The Economic Times* shows how counterfeit and fake products have infested the Indian market. The report says that India tops the world counterfeit pharma products. Over 35 % of the automotive parts sold in India are fake, while the value of counterfeit and pirated software is over US$ 1.5 billion. Such a glaring figure shows the importance of IPR restrictions in a developing economy like India.[1]

Empirical evidences showed that a poor IPR protection rate would lead to a low level of technology transfer. The study by Bascavusoglu and Zuniga (2001) examines how international differences in foreign patent protection affect decisions to transfer technology. The paper empirically evaluates the role of intellectual property protection, technology endowments, and market size on technology receipts of French firms from abroad. Results show that high-technology sectors such as chemicals, pharmaceuticals, manufacturing of machines and instruments, electronics, etc. are indeed more sensitive to IPR protection overseas. On the other hand, IPR in low-technology sectors have a negative effect, but are not significant.

An empirical study by Wakasugi and Ito (2005) showed that for Japanese multinational firms technology transfer measured by royalty payments of affiliate to parent

[1] Page 3 of *The Economic Times* dated 30 July, 2007.

N. Mitra (✉)
Department of Economics, Susil Kar College, 24 Pgs (South),
Ghoshpur, West Bengal 743 330, India
e-mail: bnilanjana2011@hotmail.com

T. Banerjee (Chatterjee)
Department of Economics, Jadavpur University, Kolkata, West Bengal 700 032, India
e-mail: tanmoyee@hotmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_3, © Springer India 2014

firms is substantial in the countries where the enforcement of IPRs is strict, and that it increases in the countries where IPRs are strong. Another empirical study by Smarzynska Javorcik (2004) sheds light on the relationship between IPR protections and structure of FDI using a unique firm-level data set describing investment projects in eastern Europe and the former Soviet Union. First, the study indicates that investors in sectors relying heavily on protection of intellectual property are deterred by a weak IPR regime in a potential host country. There is also some evidence that weak IPR protection may discourage all investors, not just those in the sensitive sectors. Second, the lack of IPR protection deters investors from undertaking local production and encourages them to focus on distribution of imported products. Interestingly, this effect is present in all sectors, not only those relying heavily on IPR protection.

Given the empirical findings, governments all over the world try to create an investment-friendly environment hoping that multinational corporations (MNCs) will bring new technologies, management skills, and marketing know-how.[2] In the light of the earlier mentioned findings, the present study theoretically explains the relationship between IPR restriction and structure of foreign investment. The study tries to find out different social welfare maximizing rates of IPR protection on the basis of which a foreign multinational firm decides whether to enter the less-developed country (LDC) market via FDI or just export the product, and second, if it enters via FDI it also considers the choice between assembly-line FDI, which implies a fragmented production structure where the high-technology intensive production procedure is undertaken in developed, IPR-protected economies and assembly-line activities are transferred to less developed, weak, IPR-protected economies,[3] and, finally, FDI with transfer of disembodied technology and complete production process.[4]

---

[2] A ready reference of the change in attitude is evident in the Indian pharmaceutical industry. After independence, the Indian Patent Act, 1970, was enforced to ensure rapid industrialization in a newly independent country as well as serve public interest in a balanced manner. The main feature of the Patent Act of 1970 was complete absence of product patents for pharmaceutical, food, and chemical-based products. The industrial sector was also covered by process patents. In order to integrate the country with the global pharmaceutical industry, the Patent Act, 1970, was amended in March, 2005. The new Patent Act introduces a product patent regime, covering drugs, foods, and chemicals. This is in compliance with the TRIPS Agreement of WTO designed to bring an end to the copy of drugs patented abroad by Indian pharma companies. This was allowed under the previous Act as long as the Indian companies used different manufacturing processes. Thus, the Parliament of India approved the third Patents (Amendment Bill), 2005, expecting to encourage foreign investment in research and development projects consequently benefiting the Indian economy. As expected, the FDI in the pharma industry is estimated to be US$ 172 million during 2005–2006 recording a CAGR of 62.6 % during the period 2002–2006. In case of R&D, in 2005–2006 the R&D expenditure of 50 major companies totaled US$ 495.19 million growing at a rate of 26 % over the previous year. This shift to a higher growth path is largely attributable to the new product patent act in 2005. Pharmaceutical Industry Analysis News by Bio Spectrum Asia: http://www.biospectrumasia.com.

[3] A common example of this type of FDI is in the case of Coca Cola—one of world's leading beverage suppliers. The MNC prepares the concentrate in the USA which is then exported to different countries where the bottling units of the MNC are located either as complete subsidiary units or as joint ventures.

[4] In this respect, we must mention Public Notice No. 60 issued by the Ministry of Commerce, Government of India, in December 1997, to increase the technology-intensive FDI in the automobile

A large number of theoretical papers have dealt with the matter of technology transfer and entry of foreign firms in the LDC market.[5] However, none of the papers considered the case of multinational firms' decision over assembly-line FDI vis-à-vis transfer of complete technology and production in the LDC via FDI under different IPR regimes.

The paper on IPR and the mode of technology transfer by Viswasrao (1993) formulates a model where the lack of IPR protection in the Southern countries affect the nature of licensing contract offered by the North as well as the mode of technology transfer. The choices available are examined in a partial equilibrium game-theoretic setting where asymmetric information adversely affects licensing of low-cost technologies to the South. The paper concludes that a Northern firm may opt for subsidiary production or monopoly licensing which lowers Southern welfare. Nicholson (2000), in a theoretical paper, considered the manner in which multinational enterprises facilitate technology transfer from the North to the South and the role played by the protection of intellectual property. Different industries respond to changes in intellectual property protection (IPP) regimes differently and alter their mode of entry accordingly. Firms with complex but easily imitable products will tend to internalize production through FDI, but firms that face a lower risk of imitation will tend to license production to nonaffiliated Southern firms.

---

sector of India. The policy placed import of capital goods and automotive components under open general license, but restricted import of cars and automotive vehicles in completely built unit (CBU) form or in completely knocked down (CKD) or in semi-knocked down (SKD) condition. Car manufacturing units were issued licenses to import components in CKD or SKD form only on executing a Memorandum of Understanding (MoU) with the Director General Foreign Trade (DGFT). Eleven companies signed MoUs with the DGFT under which they agreed to: (1) Establish actual production of cars and not merely assemble vehicles. (2) Bring in a minimum foreign equity of US$ 50 million if a joint venture involved majority foreign equity ownership. (3) Indigenize components up to a minimum of 50 % in the third and 70 % in the fifth year or earlier from the date of clearance of the first lot of imports; thereafter, the MoU and import licensing will abate. (4) Neutralize foreign exchange outgo on imports (CIF) by export of cars, auto components, etc. (FOB). This obligation was to commence from the third year of start of production and was to be fulfilled during the currency of the MoU. From the fourth year, imports were to be regulated in relation to the exports made in the previous year. However, this notice was abolished with effect from April 1, 2001. On December 21, 2001, the World Trade Organization's Ddispute Settlement Body (DSB) arrived at a decision that the "indigenisation" condition, as contained in Public Notice No. 60 and in the MoUs entered into thereunder, is in violation of Article III:4 of GATT 1994 as at the date of its establishment. With the Panel having announced its decision, India would not be able to impose, in any manufacturing area, conditions of the kind specified in its December 1997 notification, so long as it remains a member of the WTO. Sources: (a) Auto Policy, Government of India, Ministry of Heavy Industries and Public Enterprises, Department of Heavy Industry, New Delhi, March 2002. (b) Frontline, Volume 19, Issue 1, Jan 05–18, 2002, Published by *The Hindu*, http://www.hinduonnet.com/fline/fl1901/19011030.htm.

[5] According to Maskus (1998), the increase in international investments in the 1990s and the problem of protection of IPR in the same period have led to the inquisitiveness about the link between technology transfer and IPR protection. A number of papers (Helpman (1993); Lai (1998); Yang and Muskus (2001)) used endogenous growth models to show that protecting IPR could benefit the South by increasing the flow of technology to the South. The papers also considered the role of IPR protection on rate of FDI and rate of innovation.

A paper by Zigic (1998) rejects the common belief that the South generally benefits from relaxing IPR protection, while the North is worse off in a North–South duopoly framework with technological spillover. In this respect, the congruence of interest with respect to a Southern IPR protection regime should not be an exceptional or impossible state of affairs. Another paper by Zigic (2000) analyzed the issue of optimal tariffs when the Northern and Southern firms compete in quantities in an imperfectly competitive Northern market and there are potentially varying degrees of IPR violation by the South. IPR violation is reflected through the leakage of technological knowledge ("spillovers") from the Northern to the Southern firm creating unit cost reduction. It is shown that optimal tariffs in this framework are always higher than in the simple duopoly model. However, this paper did not discuss the matter of FDI.

Similarly a paper by Mattoo et al. (2004) explores the preferences of a foreign firm and a welfare-maximizing host-country government over two modes of FDI—direct entry or acquisition of existing domestic firms in the presence of costly technology transfer. The paper shows that a purely welfare-maximizing government might use FDI restrictions in order to influence the foreign firm's choice between different modes of entry. However, this paper does not give insights about the IPR protection and entry of foreign firm in the LDC market.

The model developed by Eicher and Woo (2005) tried to integrate optimal entry modes as a function of market size, FDI, fixed cost tariffs, and transport costs. The results highlight that, even in the presence of high tariffs, large countries are more likely to attract acquisition investment, while intermediate-sized countries may be predominantly served by trade. In this case also, the issue of IPR protection has not been analyzed.

The matter of IPR protection and FDI decision is analyzed in the paper by Naghavi (2005) in a North–South framework. The model endogenizes Southern IPR policy and the Northern firm's decision on whether to serve the Southern market through exports to obstruct exposure of its technology or by engaging in FDI to avoid trade costs. The Southern firm is assumed to be incapable of acquiring the production technology unless the Northern firm moves production to the South. In other words, the Northern firm acquires a monopoly position by producing at home. If the Northern firm chooses to move production to the South, the Southern firm can enter the market and the two firms compete in a Cournot duopoly setting. Furthermore, the Northern firm is capable of engaging in research and development (R&D) aimed at innovating more cost-effective production technologies. Knowledge gained through R&D is, however, assumed to have a public good character and can be imitated at zero cost. The model results show that a strict IPR regime is optimal for Southern firm as it triggers technology transfer by inducing FDI in less R&D-intensive industries and stimulates innovation by pushing multinationals to deter entry in high-technology sectors.

In the present chapter, we try to analyze the optimal entry mode for multinationals where the MNC firm can chose from any of the following options:

- Conduct the entire production in the developed, IPR-protected country and then export the finished product to the LDC
- Fragment the production between the developed country (DC) and the LDC and shift the assembly-line units to LDC
- Implementing the entire production in the LDC

Like Nagavi (2005), the model assumes that the imitator firm in the LDC market is incapable of acquiring the production technology unless the Northern firm moves complete production to the LDC market. If the MNC chooses to move production to the LDC, the imitator firm can enter the market and the two firms compete in a duopoly setting where the MNC acts as a Stackelberg leader. In other words, the product imitation is not possible if the MNC adopts the export strategy or the fragmented production strategy and, thereby, invests in assembly-line units only. The paper assumed that when the imitator enters the market, he/she sells a "look alike" of the original product and faces the same demand curve as the foreign firm.[6] In this case, we have endogenized the choice of IPR protection rate by the LDC government. Finally, the welfare implications of the different modes of entry are examined.

The rest of the chapter is structured as follows: Sect. 2 describes the basic model and the assumptions. Section 3 gives the optimal strategy choice of the MNC. Section 4 describes the welfare-maximizing choice of IPR protection rate by government. Finally, Sect. 5 gives the conclusion.

2 Model

The model considers an MNC located in the DC with the following options for production:

1. It can produce entirely in the DC market and export the finished product to the LDC with a per-unit positive shipment cost.
2. It can fragment the production process in two stages between the DC and the LDC. In the first stage, production of the core material takes place in the DC.

[6] Banerjee, Banerjee(Chatterjee), Raychaudhuri (2008) empirically pointed out the existence of the form of piracy where the imitator sells a look-alike copy of the original. The paper showed that the Recording Industry Association of America (RIAA) in its annual record of "commercial piracy" (non-Internet) statistics and enforcement efforts released on July 13, 2005 mentions, "Because of the high quality and seeming authenticity of counterfeit CDs, this genre of illicit product is increasingly finding its way to legal music retail outlets, often at prices that approach or equal the retail price of legitimate product." See http://www.hispanicprwire.com/news. A report published in Sify on November 11, 2005 mentions that, in India, fake producers have copied the latest hologram on the HP cartridge pack, making the fake cartridge almost identical to the original. See http://sify.com/printer_friendly.php?id=13940748ctid=2lid=1. The same report also highlights the copying of DaimlerChrysler spare parts where the copied product had similar number coding as found in the original. A report published on http://english.people.com.cn/200204/16/eng20020416_94113.shtml, on April 16, 2002 describes the destruction of pirated international brand name products like Rolex watches, Nike and Adidas sportswear, and Toyota and Honda automotive parts by Chinese customs officers in east China. These evidences suggest the existence of commercial piracy where the fake products are identical to the originals.

In the second stage, assembling of the core material takes place in the LDC. A common example of this type of production is in the case of Coca Cola—one of world's leading beverage suppliers. The MNC prepares the concentrate in the USA, which is then exported to different countries where the bottling units are located.

3. It can undertake the entire production in the LDC by opening up the entire manufacturing and assembling unit with FDI. Here, the government is introduced as a monitoring authority to restrict technology leakage to other competing LDC firms. The model assumes that, in the third case, where the production of core materials is taking place in the LDC market, leakage of technology can take place.

This will in turn lead to entry of competitive domestic firms if the IPR protection regime is weak.

The optimal entry mode of foreign firms depends on the rate of IPR protection already chosen by the government of the LDC firm and the cost of setting up plants in the LDC and the per-unit transport cost of the product.

The model considers the following functional forms:

The DC firm is facing a linear demand function which is given as

$$q = a - p \tag{3.1}$$

where q = quantity demanded, p = price of the final product, and a = market size parameter introduced as a positive constant.[7]

Given this demand function, we proceed with the production option for the MNC under the three different production options.

2.1 Case 1: Production Conducted Entirely in DC and Product Exported to LDC

First, we consider the situation where the foreign firm undertakes the entire production process in its own country.

The total cost function of the MNC is defined as follows:

$$c_{DC} = cq + A + tq. \tag{3.2}$$

The model assumes that the production process is divided into two stages. In the first stage of production, core materials are produced by undertaking the sunk cost A and, in the second stage of production, the assembling or finishing tasks are undertaken by incurring a per-unit variable cost c and t the per-unit positive shipment cost for transferring the finished product from the DC to the LDC.

[7] It can be shown that profitability conditions of the producer will determine the feasible values of a.

The profit function of DC firm is defined as follows:

$$\begin{aligned}\pi_{Export} &= pq - cq - tq - A \\ &= (a-q)q - cq - tq - A.\end{aligned} \quad (3.3)$$

From the first order profit-maximizing conditions, we get the monopoly output, price, and profit of the MNC as follows[8]:

$$\begin{aligned}q_{export} &= (a-c-t)/2 \\ p_{export} &= (a+c+t)/2 \\ \pi_{export} &= (a-c-t)^2/4 - A.\end{aligned} \quad (3.4)$$

2.2 *Case 2: Fragmented Production Structure Between DC and LDC*

Second, the MNC may choose the strategy of fragmenting the production between the DC and the LDC. Thus, it can conduct the manufacturing in the DC (thus bringing in embodied technology to the LDC) and complete the assembly in the LDC. The total cost function of the DC firm is given by

$$c_{Frag} = wq + A + tq + F, \quad (3.5)$$

where w is the per-unit cost of assembling the semifinished product in the LDC. It is assumed that $w < c$ due to cheap labor in the LDC. The per-unit shipment cost to transfer the intermediate product to the LDC is t. For simplicity, it is assumed to be same as the shipment cost of the finished product and F is the initial plant installation cost to transfer the production partly to the LDC. Thus, the total sunk cost of production becomes (A + F), where A is the sunk cost of production undertaken in the DC to manufacture the core material.

The profit function of the DC firm under fragmentation is given as

$$\begin{aligned}\pi_{Frag} &= pq - wq - tq - A - F \\ &= (\text{a-q})\text{q-}\ wq - tq - A - F.\end{aligned} \quad (3.6)$$

From the first-order profit-maximizing conditions, equilibrium quantity, price level, and profit are given as follows[9]:

$$\begin{aligned}q_{Frag} &= (a-w-t)/2 \\ p_{Frag} &= (a+w+t)/2 \\ \pi_{Frag} &= (a-w-t)^2/4 - A - F.\end{aligned} \quad (3.7)$$

The price output combinations are the monopoly combinations of the DC firm since the firm is the sole producer of the good in the LDC market.

[8] A positive profit implies that $a > c + t + 2\sqrt{A}$.

[9] A positive profit implies that $a > w + t + 2\sqrt{A + F}$.

2.3 Case 3. Complete Production in the LDC

The third alternative to the DC firm is to produce entirely in the LDC through FDI and undergo complete technology transfer. That is, in this case, the DC firm is bringing in disembodied technology to the LDC. The DC firm will act as a monopolist in the LDC until and unless a fake producer (producing with diffused technology from the DC firm) enters the market and sells an exact replica of the original product. Thus, with the entry of the fake producer both the firms will operate as duopolists reaching a Subgame Perfect Nash Equilibrium (SPNE) where the incumbent DC firm acts as a leader and fake firm operates as a follower. The impact of IPR restrictions is introduced in the form of a government sector acting as a monitoring authority trying to resist technology diffusion and entry of the fake producer by choosing a suitable level of the IPR protection rate.

Thus, the game plan is as follows:

1. The government first chooses a rate of IPR protection which is defined as the probability of detection of the fake firm. The probability of detection of the fake producer is $(1 - \alpha)$where $0 < \alpha \leq 1$.[10]
2. Second, the DC firm initiates production entirely in the LDC.
3. Technology diffuses to another LDC firm that enters the market by producing a replica of the original product.

If the piracy is detected, the MNC or the foreign investor continues to act as the monopolist; otherwise, the firm can at best be a Stackelberg leader with the fake producer acting as a follower.[11]

2.4 Assumptions

Let α be the probability of the entry of the fake producer, where $(1 - \alpha)$ is the endogenously determined rate of IPR protection.

Let A be the sunk cost of production incurred by the foreign investor and $(F + R)$ be the plant installation cost of the foreign firm when the firm shifts the entire production unit to LDC.[12] Thus, total sunk cost of production becomes $(A + R + F)$. As assumed in the previous model, w is the per-unit assembling cost of the semifinished product in the LDC. Let G be the government punishment/penalty cost to be paid by the fake producer, if detected. Let C be the cost of acquiring technology to be incurred by the fake producer. The fake producer does not have to incur the fixed cost for plant installation.

[10] $\alpha > 0$ implies that perfect monitoring is impossible as monitoring by the government is costly.

[11] In this case, we have assumed that incumbent firm and the fake entrant firm are involved in quantity competition. The basic reason behind this idea is that once the technology or the know-how has been copied the unit variable cost of production or marginal cost of production is negligible or may be declining for the fake firm.

[12] It is assumed that $R > 0$.

The expected profit of the incumbent MNC is given by

$$\pi_{\text{Incumbent}} = \{(1-\alpha)((\text{a}-\text{q}_1)\text{q}_1 - \text{wq}_1) + \alpha(\text{q}_1(\text{a}-\text{q}_1-\text{q}_2) - \text{wq}_1)\} - \text{A} - \text{F} - \text{R}$$
$$= \{\text{q}_1(\text{a}-\text{q}_1) - \text{wq}_1 - \alpha \text{q}_1\text{q}_2 - \text{A} - \text{F} - \text{R}\} \qquad (3.8)$$

where q_1 is the output of the incumbent MNC and q_2 is the output of the fake producer.

The expected profit of the fake producer may be given as

$$\pi_{\text{fake}} = \alpha[(\text{a}-\text{q}_1-\text{q}_2)\text{q}_2 - \text{wq}_2 - \text{A}] - (1-\alpha)\text{G} - \text{C} \qquad (3.9)$$

where $(1-\alpha)G$ is the expected penalty paid by the fake producer if detected. Solving for

$$\frac{\delta\pi_{fake}}{\delta q} = 0 \quad \text{gives} \quad q_2 = (a - q_1 - w)/2. \qquad (3.10)$$

This is the reaction function of the fake producer. Given the reaction function of the fake producer, the reduced form profit of the incumbent MNC is defined in the following way:

$$\pi_{\text{Incumbent}} = (\text{q}_1(\text{a}-\text{q}_1) - \text{wq}_1 - \alpha\text{q}_1(\text{a}-\text{q}_1-\text{w})/2 - \text{A} - \text{F} - \text{R}).$$

Maximizing this with respect to q_1 gives

$$q_{1LDC} = (a - w)/2. \qquad (3.11)$$

From Eq. (3.10), the equilibrium output level of the fake producer is given as

$$q_{2LDC} = (a - w)/4. \qquad (3.12)$$

Let p_{LDC} be the expected price under this situation.[13] It is given as

$$p_{LDC} = \alpha(a + 3w)/4 + (1-\alpha)(a + w)/2 = (\text{a}(2\text{-}\alpha) + \text{w}(2+\alpha))/4.$$

The equilibrium profits of the incumbent firm and the fake firm are given as[14]

$$\pi_{Incumbent} = (a - w)^2(2-\alpha)/8 - A - F - R \qquad (3.13)$$

$$\pi_{fake} = \alpha(a - w)^2/16 - \alpha A - (1-\alpha)G - C \qquad (3.14)$$

$$\pi_{fake} \geq 0 \Rightarrow \alpha \geq 16(G + C)/((a - w)^2 - 16A + 16G) = \hat{\alpha}. \qquad (3.15)$$

[13] In this case, when both the incumbent and the fake producer operate in the market, the price is given as *(a + 3w)/4,* and if the fake producer does not enter the market, the price is given as *(a + w)/2.*

[14] The incumbent firm will receive a positive profit even with $\alpha = 1$ if $a > w + 2\sqrt{2(A + F + R)}$.

Thus, the fake producer enters the market for $\alpha \in [\hat{\alpha},1]$.[15]

Thus, a high punishment level (G) or a high cost of copying technology C, or a high level of sunk cost of production deters the entry of a fake producer.

For $\alpha < \hat{\alpha}$, the incumbent firm acts as a monopolist and earns its monopoly profit given as[16]

$$\pi_{Incumbent}^{mono} = (a - w)^2/4 - A - F - R. \tag{3.16}$$

In the next section, we try to find the optimal strategy choice by the DC firm under different IPR regimes and different levels of plant set-up cost in the LDC under FDI.

3 Optimal Strategy Choice of the MNC

As already mentioned, the MNC has three possible strategies:

1. It can choose the export strategy where the production is taking place in the DC.
2. It can fragment the production process. In this case, core materials are produced in the DC; only the finishing or the assembling part takes place in the LDC.
3. It can produce the commodity entirely in the LDC.

First, we compare the profit of the MNC under the export and the fragmented production process.

3.1 Proposition 1

For certain values of plant installation cost (below the critical level given by F^*), the incumbent firm prefers the fragmented production structure over the export strategy. Again, the larger market size prefers this move while the higher transport cost discourages such effort, where $F^* = (2a - w - c - 2t)(c - w)/4$.

[15] From Eq. (9a), it is clear that given a monitoring rate an increase in the punishment level will reduce profitability of the fake producer. Thus, a high value of G will increase $\hat{\alpha}$, that is, as G increases the fake producer will enter the market only because of a lenient monitoring rate. It has been observed that $\partial\hat{\alpha}/\delta G = 16\{(a - w)^2 - 16A - 16C\}/\{(a - w)^2 - 16A + 16G\}^2 > 0 \Rightarrow \{(a - w)^2 - 16A - 16C\} > 0$. Also, we have $\delta^2\hat{\alpha}/\delta G^2 = -256G/[(a - w)^2 - 16A + 16G]^3 < 0$.Thus, as G increases $\hat{\alpha}$ rises at a decreasing ate. This, along with Eq. (13) implies that there does not exist a maximum value of G which will completely control the entry of the fake producer by making $\hat{\alpha} = 1$. From Eq. (13), it is clear that $\hat{\alpha} < 1$, as long as $\partial\hat{\alpha}/\delta G > 0$i.e.$\{(a - w)^2 - 16A - 16C\} > 0$. Thus, the LDC government cannot control the entry of a fake producer by simply charging a high penalty (to the fake producer). It has to undertake some amount of monitoring activity; otherwise, the fake producer will always have a positive probability of entry in the market. Here lies the importance of endogenous choice of IPR protection by the government of the LDC. Second, the condition $\{(a - w)^2 - 16A - 16C\} > 0$ imposes a restriction on the value of a given as $a > w + 4\sqrt{A + C}$. Thus, combining footnotes 8, 9, and 14 along with this condition we get the feasible level of a as $a > \max\{\mathrm{w} + 4\sqrt{A + C}, c + t + 2\sqrt{A}, w + t + 2\sqrt{A + F}, w + 2\sqrt{2(A + F + R)}\}$.

[16] If the condition given in footnote 14 is satisfied, the monopoly firm will always get the positive monopoly profit.

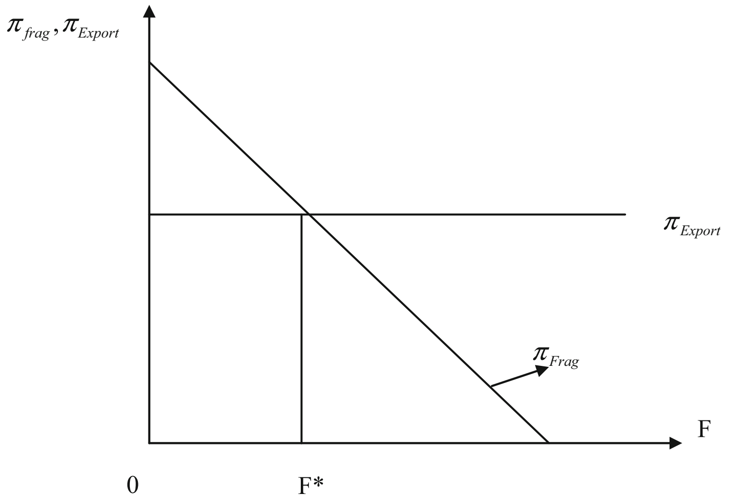


Fig. 3.1 For values of F below F^* fragmented production takes place; otherwise, the export strategy is chosen

Proof: $\pi_{\text{Frag}} - \pi_{\text{export}} = [(\text{a} - \text{w} - \text{t})^2/4 \ - \text{A} - \text{F}] - [(\text{a} - \text{c} - \text{t})^2/4 \ - \text{A}].$

Now $\pi_{\text{Frag}} \geq \pi_{\text{export}}$ implies that

$$F \leq (a - w - t)^2/4 \ - (a - c - t)^2/4 \ = (2a - w - c - 2t)(c - w)/4 \ . \quad (3.17)$$

It is assumed that

$$F^* = (2a - w - c - 2t)(c - w)/4. \quad (3.18)$$

Let F^* be the critical level of cost of opening a production or assembly-line unit in the LDC. If the actual value of F is above this level, from Eq. (3.17), it is clear that the foreign firm will choose the export strategy over the fragmentation strategy; only a low level of F will induce the fragmented production structure.

From Fig. 3.1, it is clear a reduction in the actual value of F by infrastructure development will shift the π_{Frag} schedule upward and induce the foreign firm to set up an assembly-line plant in the LDC instead of adopting the export strategy.

Now, let us analyze the impact of different parameters on the critical value of F:

1. Impact of market size parameter on F^*:

$$\delta F^*/\delta a \ = 1/2 > 0.$$

If the market size increases, the critical value of F increases. From Fig. 3.1, it can be observed that for an increase in "a" the market size parameter will shift the π_{Frag} schedule rightward and the π_{Export} schedule upward. An increase in F^* implies that the π_{Frag} schedule shifts more than proportionately the π_{Export} schedule. This implies that the profitability of production under the fragmented strategy increases more than under the export strategy. This increases the range of values of cost of the foreign investor to set up a plant in the LDC for which the fragmented production structure is profitable compared to the export strategy. Thus, a large market size encourages the fragmented production structure.

2. Impact of transport cost on critical value of F^*:

$$\delta F^*/\delta t \ = -1/2 < 0.$$

If the transport cost (cost of exporting the finished product to the LDC or transferring the intermediate product to the LDC) increases, the critical value of F^* decreases. An increase in transport cost reduces the profitability of the foreign firm under both strategies and the profit schedules will shift downward in Fig. 3.1 for an increase in "t." A fall in F^* implies that the downward shift in the π_{Frag} schedule is more than that of the π_{Export} schedule. Thus, an increase in the transport cost reduces the profitability of fragmented production more than proportionately than that of export strategy. Thus, a higher transport cost discourages the fragmented production structure compared to the export strategy.

Next, we compare the profit under fragmented and complete LDC strategy.

3.2 *Proposition 2*

The DC firm chooses to produce completely in the LDC for the values of α in the interval $\alpha \in [0, \max(\alpha^*, \hat{\alpha})]$ when $R < (4t(a-w) - 2t^2)/8$ otherwise it chooses the strategy of fragmented production, where

$$16(G+C)/(a-w)^2 - 16A + 16G \ = \hat{\alpha} \text{ and } 4t(a-w) - 2t^2 - 8R/(a-w)^2 \ = \alpha^*.$$

Proof: It has been shown that if the actual value of $\alpha < \hat{\alpha}$, then the incumbent firm operates as a monopolist in the market and gains a profit given by Eq. (3.16). This profit dominates the profit under fragmentation given by Eq. (3.7) as long as thetransportation cost t is positive and $R < (4t(a-w) - 2t^2)/8$.[17] Hence, for $\alpha < \hat{\alpha}$, the firm chooses the complete LDC strategy over fragmentation if and only if $R < (4t(a-w) - 2t^2)/8$ with a positive t or, in other words, t lies in an interval defined as $(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w) - 2\sqrt{A+F}$.

[17] If $R < (4t(a-w) - 2t^2)/8$, then t lies in the interval $(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w) + \sqrt{(a-w)^2 - 4R}$. From footnote 9, it is clear that $t < a - w\text{-}2\sqrt{A+F}$. Hence, the feasible range of t to satisfy the earlier mentioned condition for a given value of R is $(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w) - 2\sqrt{A+F}$. In this case, when the condition given in footnote 14 is satisfied, we always have $(a-w) - \sqrt{(a-w)^2 - 4R} < (a-w) - 2\sqrt{A+F}$. Alternatively, if R is high, or $R > (4t(a-w) - 2t^2)/8$ when fragmented profit dominates the monopoly complete LDC profit, then t must assume a value given as t $< (a-w) - \sqrt{(a-w)^2 - 4R}$.

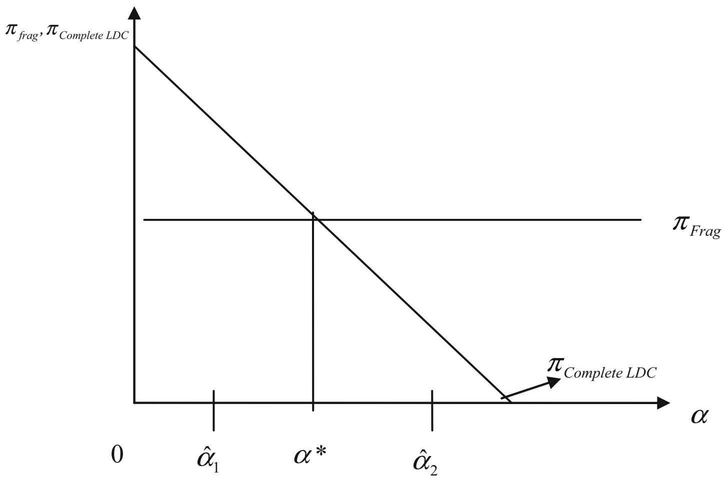


Fig. 3.2 If the fake producer actually operates in the LDC market, then the incumbent DC firm chooses the complete LDC strategy if $\alpha \leq \alpha^*$ where $R < (4t(a-w) - 2t^2)/8$. If $\hat{\alpha} < \alpha^*$ ($i.e\ \hat{\alpha}_1$) the complete LDC strategy is chosen

Alternatively if $\alpha \geq \hat{\alpha}$, the fake producer operates at the market along with the incumbent firm. In that case, comparing profits of the incumbent firm to that of fragmentation strategy we get the following results:

$$\pi_{\text{Incumbent}} - \pi_{\text{Frag}} \geq 0 \Rightarrow [(a-w)^2(2-\alpha)/8 - A - F - R] - [(a-w-t)^2/4 - A - F] \geq 0$$

implies that

$$\alpha \leq 4t(a-w) - 2t^2 - 8R/(a-w)^2 = \alpha^*. \tag{3.19}$$

Thus, for $R > (4t(a-w) - 2t^2)/8$, the complete LDC strategy is never chosen by the incumbent DC firm even if $\alpha = 0$. Alternatively, for $R < (4t(a-w) - 2t^2)/8$, the optimal decision of the DC firm depends on the rate of IPR protection.[18] The equilibrium configuration is given in Fig. 3.2.

From Fig. 3.2 when $\hat{\alpha} < \alpha^*$, the firm chooses the complete LDC strategy up to $\alpha = \alpha^*$, beyond which the fragmented strategy is adopted. Alternatively if $\hat{\alpha} > \alpha^*$, then the firm chooses complete LDC up to $\alpha = \hat{\alpha}$ because in the interval $0 < \alpha \leq \hat{\alpha}$, it operates as a monopolist. Beyond $\hat{\alpha}$ again fragmentation will take place.

The comparative static analysis with respect to market size "a" and the transport cost and R gives the following result:

$$\delta\alpha^*/\delta a = -4t(a-w-t)/(a-w)^3 < 0.$$

[18] The critical value of α given by $\alpha*$ will be positive if $R < (4t(a-w) - 2t^2)/8$. As from the last footnote, this condition can in turn be interpreted as a restriction on the value of transport cost t given as $(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w) - 2\sqrt{A+F}$. Alternatively, for $R > (4t(a-w) - 2t^2)/8$, we must have t $< (a-w) - \sqrt{(a-w)^2 - 4R}$. Thus, a small t implies that $\alpha*$ will assume a negative value such that the fragmented profit will dominate over the complete LDC profit.

Also, $\delta\hat{\alpha}/\delta a \quad =< 0$ which implies that a higher market size encourages the fake producer to enter the market.

Hence, a higher market size discourages complete LDC production. The logic is intuitive. The higher the market size, greater will be the profitability of production for the incumbent firm as well as the fake producer. So, more stringent IPR restriction is required for transfer of disembodied technology; otherwise, FDI is channeled to assembly-line sectors only.

Again, $\delta\alpha^*/\delta t = 4(a - w - t)/(a - w)^3 > 0$.

Hence, if t increases $\alpha*$ decreases implying a higher transport cost favoring complete production in LDC with FDI.

Finally, $\delta\alpha^*/\delta R = -8/(a - w)^2 < 0$. Here, R is the extra setup cost of transferring the entire production process to the LDC over the assembly-line unit. An increase in R favors the fragmented production structure compared to the complete LDC strategy. Thus, as R increases a strict IPR protection is required to induce the profitable transfer of disembodied technology over assembly-line production.

Next, we compare the profit under the export strategy and the complete LDC strategy and draw the following results.

3.3 Proposition 3

1. The firm will choose the export strategy over the complete LDC strategy if $F + R \geq F_1^*$.
2. For $F + R < F_1^*$, the incumbent firm chooses complete LDC production where $\alpha \in [0, \max(\alpha, \alpha_1^*)]$; otherwise, the export strategy is chosen.

Thus, the firm chooses complete LDC if the monitoring is strong, where

$$\frac{16(G + C)}{(a - w)^2 - 16A + 16G} = \hat{\alpha}, \quad \alpha_1^* = \frac{8(F^* - F)}{(a - w)^2} + \alpha^*,$$

$$F^* + (2(a - w) - t)t/4 \ = F_1^*, \quad \text{and} \quad F^* = (2(a - w - c - 2t)(c - w)/4.$$

Proof: First, we compare the export strategy profit of the monopolist to the monopoly profit of the incumbent firm under the complete LDC strategy when the fake producer is not entering the market. This is possible for $0 < \alpha < \hat{\alpha}$.

$$\begin{aligned} &\pi_{Incumbent}^{Monopoly} - \pi_{Export} \\ &\quad = [(\mathrm{a} - \mathrm{w})^2/4 \ - \mathrm{A} - \mathrm{F} - \mathrm{R}] - [(\mathrm{a} - \mathrm{c} - \mathrm{t})^2/4 \ - \mathrm{A}] \geq 0 \\ &\quad F + R \leq (2a - w - t - c)(c + t - w)/4 \ = F_1^* \qquad (3.20) \end{aligned}$$

where

$$\mathrm{F}_1^* = \mathrm{F}^* + [(2(\mathrm{a} - \mathrm{w}) - \mathrm{t})\mathrm{t}/4\ ]. \qquad (3.21)$$

Thus, Eq. (3.20) can be written as

$$\mathrm{F} < \mathrm{F}^{*} + \{((2(\mathrm{a}-\mathrm{w})-\mathrm{t})\mathrm{t}-4\mathrm{R})/4\ \}. \qquad (3.22)$$

For values of plant installation cost (F + R) above F_1^*, the export strategy is always adopted, as in this case even the monopoly profit under the complete LDC strategy is less than that of the export strategy.

One thing must be noted. In this case, if $R > (4t(a-w)-2t^2)/8$[19] along with $\mathrm{F} > \mathrm{F}^*$, then we have $F+R > F_1^*$; in this case, the export strategy is always chosen.

Next, we compare the profits for these two strategies when the fake producer operates in the market along with the incumbent firm[20]

$$\pi_{\text{Incumbent}} - \pi_{\text{Export}} \geq 0$$

$$\Rightarrow [(\mathrm{a}-\mathrm{w})^2(2-\alpha)/8 - \mathrm{A} - \mathrm{F} - \mathrm{R}] - [(\mathrm{a}-\mathrm{c}-\mathrm{t})^2/4\ - \mathrm{A}] \geq 0$$

$$(a-w)^2 - (a-c-t)^2/4\ - (F+R) \geq \alpha(a-w)^2/8$$

$$(2a-w-c-t)(c+t-w)/4\ - F - R \geq \alpha(a-w)^2/8$$

$$\Rightarrow \alpha(a-w)^2/8 + F \leq F^* + (4t(a-w) - 2t^2 - 8R)/8 = F_1^* - R. \qquad (3.23)$$

Thus, for the combination of (α, F), below the straight line depicted by Eq. (3.23), the complete LDC strategy is chosen over the export strategy.

This can be further simplified as[21, 22]

$$\alpha \leq 8\{F_1^* - (F+R)\}/(a-w)^2\ = \alpha_1^*. \qquad (3.24)$$

Given the values of F and R, if $F+R > F_1^*$ then $\alpha_1^* < 0$. This implies that condition (19) is impossible to hold as $\alpha > 0$. Hence, the export strategy is adopted. In this case, if $R > (4t(a-w)-2t^2)/8$[23] along with $\mathrm{F} > F^*$ then $F+R > F_1^*$ and by the same logic the export strategy is chosen. Next, we consider the cases for $R < (4t(a-w)-2t^2)/8$.[24] In this case also, if $F+R > F_1^*$, then$\alpha_1^* < 0$ and the export strategy is always adopted. Alternatively if $F+R \leq F_1^*$ for values of α less

---

19 or $\mathrm{t} < (a-w) - \sqrt{(a-w)^2 - 4R}$.

20 $(2a-c-w-t)(c+t-w)$ can be written as
$((2a-c-w-2t)+t)((c-w)+t) = (2a-c-w-2t)(c-w) + t(2a-2w-2t) + t^2$
$= 4F^* + t(2a-2w-2t) + t^2 = 4F^* + 2t(a-w) - t^2 = 4F_1^*$
where from Eq. (14), we have $F^* = (2a-w-c-2t)(c-w)/4$.

21 For $\alpha^* > 0(i.e.R < (4t(\text{a-}w)\text{-}2t^2)/8) \Rightarrow \alpha_1^* = 8(F^*-F)/(a-w)^2\ + \alpha^*$.

22 The optimal value of α_1^* is negatively related to F. Given a value of R, if F increases, then the optimal value of α_1^* falls. It implies that with higher plant installation cost in the LDC, the incumbent firm requires a stringent IPR protection regime for profitable production of complete production in the LDC.

23 Which in turn implies that $\mathrm{t} < (a-w) - \sqrt{(a-w)^2 - 4R}$ for a given value of R.

24 If this condition is satisfied, then t falls in the interval given as $(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w) - 2\sqrt{A+F}$.

than α_1^*, the incumbent firm chooses the complete LDC strategy as compared to the export strategy when the fake producer operates in the market (i.e., $\alpha \geq \hat{\alpha}$). In that case if $\alpha_1^* > \hat{\alpha}$, then the complete LDC is chosen for $\alpha \in (0, \alpha_1{}^*)$. In this interval, the firm will act as a monopolist for $0 < \alpha \leq \hat{\alpha}$, as it is not profitable for the fake firm to enter the market for the corresponding values of α. The incumbent firm receives the duopoly profit in the interval $\hat{\alpha} < \alpha \leq \alpha_1^*$. Instead, if $\alpha_1^* < \hat{\alpha}$ then the MNC will choose the complete LDC strategy for $\alpha \in (0, \hat{\alpha}]$. The incumbent firm receives the monopoly profit in this entire range of α values as the fake firm does not enter the market for $\alpha < \hat{\alpha}$. Combining the results, it is obtained that if $F + R < F_1^*$ the complete LDC strategy is chosen in the interval $\alpha \in [0, \max(\hat{\alpha}, \alpha_1^*)]$; otherwise, the export strategy is chosen. (Hence proved.)

Combining the three propositions, we get Proposition 4.

3.4 Proposition 4

For $R > (4t(a - w) - 2t^2)/8$, the foreign firm adopts the fragmented strategy for F< F^*; otherwise, the export strategy is adopted.

For $R < (4t(a - w) - 2t^2)/8$,

1. If $F + R \geq F_1^*$ the foreign firm always chooses the export strategy.
2. For $F^* < F < F_1^* - R$, the foreign firm chooses the complete LDC strategy for $\alpha \in [0, \max(\alpha, \alpha_1^*)]$, if $\widehat{\alpha} < \alpha^*$. Alternatively, if $\widehat{\alpha} > \alpha^*$, it chooses the complete LDC strategy for $\alpha \in [0, \alpha_1^*]$; otherwise, the export strategy is chosen.
3. For $0 \leq F \leq F^*$, the foreign firm chooses the complete LDC strategy if $\alpha \in [0, \max(\alpha,\ \alpha^*)]$; otherwise, the fragmented strategy is adopted.

The proof follows from the three other propositions.

The results obtained in Proposition 4 can be presented by Fig. 3.3. Figure 3.3 is drawn for $R < (4t(a - w) - 2t^2)/8$. Figure 3.3 depicts Eqs. (3.17) (the horizontal line through F^*), (3.19) (the vertical line through α^*), and (3.23) (the downward sloping intercept), respectively.

From proposition 4, it is clear that if the cost of plant installation in LDC is very high (i.e., $F + R > F_1^*$), then the firm prefers to choose the export strategy rather than involving in any type of FDI in the LDC. Alternatively, the foreign firm will be involved in transferring disembodied technology via FDI if and only if the rate of IPR protection is strong and $R < (4t(a - w) - 2t^2)/8$, where R is the extra cost of setting up the complete production structure in the LDC. In this framework, disembodied technology transfer will take place and the firm will undertake the complete production in the LDC if and only if the rate of monitoring is strong and falling in the range given by $\alpha \in [0, \max(\alpha, \min(\alpha_1^*, \alpha^*)]$. Given the assumption that transfer of the complete production process to the LDC may lead to entry of the fake producer supplying an imitation product with diffused technology from the DC firm, the incumbent firm transfers complete production if and only if the IPR protection rate is high. When the IPR protection rate is low and the cost of plant installation in

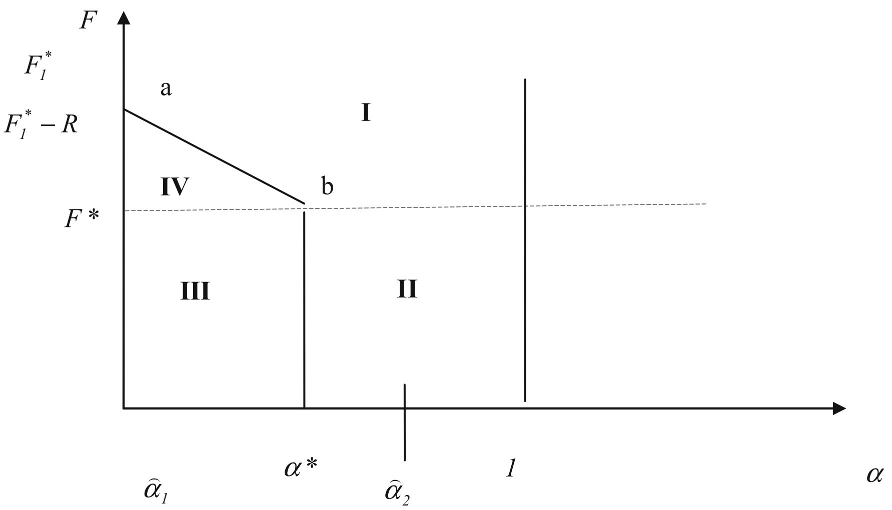


Fig. 3.3 In this figure, given the value of F, in region I the export strategy is always adopted if $\widehat{\alpha} < \alpha_1^*$; otherwise, the export strategy is chosen for $\alpha > \widehat{\alpha}$. In region II, the fragmented strategy is always adopted if $\alpha^* > \widehat{\alpha}$. Otherwise, the fragmented strategy is chosen for $\alpha > \widehat{\alpha}$. If F and a combinations fall in region III, the complete LDC strategy is chosen for values of α upto α^* over the fragmented strategy. If F and a combinations fall in region IV, the complete LDC strategy is chosen over the export strategy

the LDC is also low, the incumbent firm prefers to open the assembly-line units in the LDC and imports embodied technology to the LDC.[25] This fragmentation strategy is chosen if $\alpha > \max(\alpha^*, \alpha)$ and the plant installation cost $F \leq F^*$.[26]

[25] The model assumed that import of embodied technology does not have any technology spillover effect.

[26] In this case, we have assumed that under the fragmentation strategy the fake producer cannot copy the original product and hence does not enter the market. However, a paper by Van Long (2005) considers the case of "incomplete outsourcing" (that is, when the parent firm is fragmenting the production structure and transferring the production of upstream components to low-wage economies) in the presence of spillover through training of workers that benefit the rival firm in the low-wage economy. In the present structure, if we allow the technology spillover and entry of the fake firm under the fragmentation strategy, then the following cases will be observed. First, if the transport cost is negligible or almost nil, then the fragmentation strategy will dominate the complete LDC strategy and the incumbent firm will only choose between the export strategy and the fragmentation strategy. Thus, complete transfer of technology with FDI will not take place. Second, with positive transport cost, the fragmentation strategy will be adopted if and only if the rate of IPR protection is strong or, in other words, α is below a critical level. Otherwise, the firm chooses between the export strategy and the complete LDC strategy.

4 Social Welfare Analysis

The model assumes that the government of LDC maximizes the social welfare to choose the optimal value of the IPR protection rate.[27] The social welfare comparison of the different modes of production also gives important insights into the different modes of production chosen by the MNC. Social welfare can be defined as the sum total of profit retained by the foreign firm in the LDC, consumer surplus and government surplus less the cost of production, monitoring, etc.

The assumption made here is that under FDI, the DC firm fully repatriates profit, thus leaving social welfare as the sum total of

$$\text{SW} = \text{Consumer Surplus} + \text{Government Surplus} - \text{Costs.}$$

In case when production is conducted entirely in the DC, the government surplus and costs are both zero. The model assumes that if the product is exported to the LDC, then the probability of product imitation does not exist. Hence, the government does not incur any monitoring cost to protect the IPR of the exported product. So under the export strategy the social welfare is defined as

$$SW_{Export} = (a - c - t)^2/8. \tag{3.25}$$

In case of the fragmentation strategy, the social welfare is defined as follows:

$$SW_{Frag} = (a - w - t)^2/8. \tag{3.26}$$

Finally, we consider the situation where the fake producer enters the market. The fake producer will profitably operate in the market if and only if $\alpha > \hat{\alpha}$.

Then, the ex ante level of social welfare is defined as follows:

$\text{SW}_{\text{LDC}} =$ Expected profit of fake producer + Expected net level of consumer surplus + Government Earnings.

The expected profit of the fake producer is given by Eq. (3.14) as

$$\pi_{fake} = \alpha(a - w)^2/16 \ - \alpha A - (1 - \alpha)G - C.$$

The expected consumer surplus is as follows:[28]

$$CS_{LDC} = (a - w)^2(4 + 5\alpha)/32 \ . \tag{3.27}$$

---

[27] In this case, it is important to note that an LDC government can optimally choose the level of infrastructure, which determines the cost of setting up plants in the LDC by a foreign firm. That is, F and R can also be instrumental variables of the government of the LDC. Here, to find out the optimal effect of IPR protection rate, it is assumed that F and R are exogenously given.

[28] CS_{LDC} = (Probability that fake producer gets detected) *(Consumer surplus when incumbent firm operates as monopolist) + (Probability that fake producer cannot be detected)*(Consumer surplus when fake producer operates along with the incumbent firm.) $= 9\alpha(a - w)^2/32 + (1 - \alpha)(a - w)^2/8 = (a - w)(4 + 5\alpha)/32$.

The net government surplus is defined as

$$GS_{LDC} = (1-\alpha)G - d(\alpha) \tag{3.28}$$

where $d(\alpha)$ is assumed to be the government monitoring cost such that $d'(\alpha) < 0$, that is, as the monitoring cost increases, the probability of entry of the fake producer decreases and vice versa. Again $d''(\alpha) > 0$, implying that the monitoring cost increases at an increasing rate. Finally, it is assumed that the complete monitoring is impossible implying that

$$d(\alpha) \to \infty \quad \text{when} \quad \alpha \to 0.$$

Thus, social welfare under the complete LDC strategy when the fake producer is operating in the market (i.e., $\alpha > \hat{\alpha}$) is defined as follows:

$$SW_{LDC} = (a-w)^2(7\alpha+4)/32 \ -\alpha A - C - d(\alpha). \tag{3.29}$$

From Eq. (3.29), we have the following results:

$$\frac{dSW_{LDC}}{d\alpha} = 7(a-w)^2/32 \ - A - d'(\alpha) > 0. \tag{3.30}$$

As $d'(\alpha) < 0$ and $7(a-w)^2/32 \ - A > 0$ (this follows from the profitability condition of the fake producer),

$$\frac{d^2SW}{d\alpha^2} = -d''(\alpha) < 0. \tag{3.31}$$

Otherwise in the absence of the fake producer (i.e., $\alpha \le \hat{\alpha}$), the social welfare is defined as follows:

$$SW_{LDC} = (a-w)^2/8 - d(\alpha). \tag{3.32}$$

Finally, it can be shown that Eq. (3.29) dominates Eq. (3.32) around $\alpha = \hat{\alpha}$.

Thus, the social welfare function under the complete LDC strategy is increasing in α and discontinuous at $\alpha = \hat{\alpha}$.

Next, we compare the level of social welfare for three possible strategies and get the following proposition:

4.1 Proposition 5

1. *For* $R > (4t(a-w) - 2t^2)/8$, the government always chooses no monitoring strategy.
2. *For* $R < (4t(a-w) - 2t^2)/8$
   a. If $F + R \ge F_1^*$, the government chooses the no-monitoring strategy
   b. If $0 \le F \le F^*$

1. when $\hat{\alpha} < \alpha^*$, sufficient conditions that the government chooses $\alpha_{opt} = \alpha^*$ are $R/2 \geq d(\alpha^*)$ and $\alpha^* \geq 2/7$
2. Alternatively, when $\hat{\alpha} \geq \alpha^*$, a sufficient condition that the government chooses $\alpha_{opt} = \hat{\alpha}$ and $SW_{LDC}|_{\alpha \leq \hat{\alpha}} > SW_{Export}$ at $\alpha = \hat{\alpha}$ is , $R/2 \geq d(\hat{\alpha})$ otherwise, no monitoring is the optimal government strategy

c. $F^* < F < F_1^* - R$

1. when $\hat{\alpha} < \alpha_1^*$, sufficient conditions that $SW_{LDC}|_{\alpha > \hat{\alpha}}$ is higher than SW_{LDC} at $\alpha = \alpha_1^*$ are $(F^* + R)/2 > d(\alpha_1^*)$ and $\alpha_1^* \geq 2/7. So\, \alpha_{opt} = \alpha_1^*$
2. Alternatively when $\hat{\alpha} \geq \alpha_1^*$, a sufficient condition that government chooses $\alpha_{opt} = \hat{\alpha}$ and $SW_{LDC}|_{\alpha \leq \hat{\alpha}} > SW_{Frag}$ at $\alpha = \hat{\alpha}$ is $(F^* + R)/2 > d(\widehat{\alpha})$; otherwise, no monitoring is the optimal government strategy

Proof: Comparison of social welfare for the export strategy and the fragmented strategy gives the following results (from Eq. (3.18)):

$$
\begin{aligned}
&SW_{Frag} - SW_{Export} \\
&= (2a - w - c - 2t)(c - w)/8 = F^*/2 > 0.
\end{aligned}
$$

The consumer surplus under the fragmentation strategy is higher than that of the export strategy, which is obvious as the price of the product under the fragmentation strategy is lower than that of the export strategy.

From Proposition 4, it is clear that for $R > (4t(a - w) - 2t^2)/8$ the foreign firm adopts the fragmented strategy for F < F^*; otherwise, the export strategy is adopted. The earlier mentioned condition also implies that t is below the critical level $(a - w) - \sqrt{(a - w)^2 - 4R}$. Thus, a lower value of transport cost or alternatively a high setup cost for setting up a complete production unit in the LDC will induce an equilibrium where the export strategy is adopted or assembly-line FDI is taking place irrespective of the level of IPR protection. Thus, for this range of R values, the government does not incur any cost to protect IPR and the optimal value of α is $\alpha_{opt} = 1$.Given$\alpha_{opt} = 1$, the firm will choose its optimum strategy. If F < F^*, the firm chooses the fragmented production structure; otherwise, it chooses the export strategy.

Second, if $R < (4t(a - w) - 2t^2)/8$ then for $F + R \geq F_1^*$, the foreign firm always chooses the export strategy. Thus, in this range of plant installation cost the domestic government does not incur any cost to protect IPR and optimal value of α is $\alpha_{opt} = 1$ and the export strategy will be the Nash equilibrium of the game.

Next, we compare the social welfare under the complete LDC strategy and the fragmented strategy. However, from Proposition 4 it is clear that the complete LDC strategy is chosen only if $\alpha \in [0, \max(\hat{\alpha},\alpha^*)]$; otherwise, the fragmented strategy is chosen when $0 < F < F^*$.

For $\alpha \leq \hat{\alpha}$, the fake firm does not enter the market; hence, the difference between the social welfare under fragmented strategy and the complete LDC strategy is obtained by comparing Eqs. (3.26) and (3.32) as follows:

$$
\begin{aligned}
&SW_{LDC}|_{\alpha \leq \hat{\alpha}} - SW_{Frag} = [(\mathrm{a} - \mathrm{w})^2/8 - (\mathrm{a} - \mathrm{w} - \mathrm{t})^2/8] - \mathrm{d}(\alpha) \\
&SW_{LDC}|_{\alpha \leq \hat{\alpha}} - SW_{Frag} = (2(a - w)t - t^2)/8 - d(\alpha). \qquad (3.33)
\end{aligned}
$$

The first part is always positive, but given the assumption that $d(\alpha) \to \infty$ when $\alpha \to 0$, Eq. (3.33) will assume negative values for lower values of α and as α increases and the cost of monitoring declines then Eq. (3.33) may assume a positive value.

Given Proposition 4, for $0 < F < F^*$, if $\alpha^* < \hat{\alpha}$, the foreign firm chooses the complete LDC strategy in the interval $0 < \alpha \leq \hat{\alpha}$, and beyond this level the fragmented production is chosen. Given that the social welfare under the complete LDC strategy increases with α, the optimal value of α is $\hat{\alpha}$, if

$$SW_{LDC}|_{\alpha \leq \hat{\alpha}} \geq SW_{Frag} \text{ at } \alpha = \hat{\alpha}.$$

If the government chooses $\alpha = \hat{\alpha}$, the incumbent firm will choose the complete LDC strategy as the equilibrium strategy, and the fake firm does not enter the market as it receives a zero profit for this value of α.

At $\alpha = \hat{\alpha}$, we have

$$\begin{aligned}&(SW_{LDC}|_{\alpha \leq \hat{\alpha}} - SW_{Frag})_{\alpha = \hat{\alpha}} \\ &= 2(a - w)t - t^2/8 - d(\hat{\alpha}).\end{aligned} \tag{3.34}$$

As $R < (4t(a - w) - 2t^2)/8$ implies that

$$(2t(a - w) - t^2)/8 - d(\hat{\alpha}) > R/2 - d(\hat{\alpha}).$$

Thus, if $(R/2 > d(\hat{\alpha}))$, the complete LDC strategy is always adopted and the optimal value of α is $\hat{\alpha}$. If this condition is not satisfied, Eq. (3.34) is ambiguous in sign. The higher the value of transport cost, the higher will be the gain in consumer surplus from complete LDC production. However, for lower values of t, gain in consumer surplus may not be enough to cover the cost of IPR protection. In that situation, the social welfare under fragmentation may dominate over complete LDC production and the government chooses $\alpha = 1$. Then, the incumbent firm will choose the fragmentation strategy as $0 < F < F^*$ and the fake firm does not enter the market. Hence, assembly-line FDI takes place in the LDC.

Alternatively, if for $0 < F < F^*$, and $\alpha^* \geq \hat{\alpha}$, the foreign firm chooses the complete LDC strategy in the interval $0 < \alpha \leq \alpha^*$, and beyond this level fragmented production is chosen. Given that $\alpha^* \geq \hat{\alpha}$, the incumbent firm chooses the complete LDC production strategy and acts as a monopolist in the interval $0 < \alpha \leq \hat{\alpha}$ and as a duopolist in the interval $\hat{\alpha} < \alpha \leq \alpha^*$. For $\alpha > \hat{\alpha}$,

$$\begin{aligned}&SW_{LDC}|_{\alpha > \hat{\alpha}} - SW_{Frag} \\ &\quad = [(\mathrm{a} - \mathrm{w})^2/8 - (\mathrm{a} - \mathrm{w} - \mathrm{t})^2/8] - \alpha \mathrm{A} - \mathrm{C} - \mathrm{d}(\alpha) - [(\mathrm{a} - \mathrm{w} - \mathrm{t})^2/8] \\ &\quad = [\mathrm{t}(2(\mathrm{a} - \mathrm{w}) - \mathrm{t})/8] + \alpha[\mathrm{t}(2(\mathrm{a} - \mathrm{w}) - \mathrm{t})/8] - \mathrm{C} - \mathrm{d}(\alpha).\end{aligned} \tag{3.35}$$

As we have mentioned, SW_{LDC} is increasing in α and discontinues at $\alpha = \hat{\alpha}$. Given the strategy of the foreign firm is that it chooses complete LDC strategy only in the interval $0 < \alpha \leq \alpha^*$, otherwise it chooses the fragmented strategy, the optimal strategy for the government is to choose $\alpha = \alpha^*$ if $SW_{LDC}|_{\alpha > \hat{\alpha}} \geq SW_{Frag}$ at $\alpha = \alpha^*$.

Otherwise, the government chooses $\alpha = 1$.

A sufficient condition is that the $SW_{LDC}|_{\alpha > \hat{\alpha}} \geq SW_{Frag}$ at $\alpha = \alpha^*$ if $R/2 \geq d(\alpha^*)$ and $\alpha^* \geq 2/7$.[29]

When $0 < F < F^*$, if $\alpha^* \geq \hat{\alpha}$, the optimal strategy of the government is to choose $\alpha = \alpha^*$ when $R/2 \geq d(\alpha^*)$ and $\alpha^* \geq 2/7$. For $\alpha_{opt} = \alpha^*$, the incumbent firm chooses the complete LDC strategy and the fake firm will enter the market.

Otherwise if the cost of monitoring is very high such that $SW_{LDC}|_{\alpha > \hat{\alpha}} - SW_{Frag} < 0$ at $\alpha = \alpha^*$, then "no monitoring" is chosen so that $\alpha = 1$ and the incumbent firm will choose the "fragmentation strategy" and the fake firm will not enter the market. Thus, in the situation of costly monitoring assembly-line FDI may take place in the LDC.

Finally, we consider the social-welfare maximizing value of α for $F^* < F < F_1^* - R$. For this range of plant installation cost in the LDC market, from Proposition 4, it is clear that the foreign firm chooses the complete LDC strategy if $\alpha \in [\mathrm{t}(2(\mathrm{a}-\mathrm{w})-\mathrm{t})/8]$; otherwise, the export strategy is chosen. Like the earlier case, there could be two cases.

First, we consider the situation where $\alpha_1^* < \hat{\alpha}$ along with $F^* < F < F_1^* - R$. In this case, the complete LDC strategy is chosen for $0 < \alpha < \hat{\alpha}$ (as the imitator does not enter the market in this interval). Hence, the comparison of social welfare under the export strategy and the complete LDC strategy gives the following result:

$$SW_{LDC}|_{\alpha \leq \hat{\alpha}} - SW_{export} = \{(\mathrm{a}-\mathrm{w})^2/8 - (\mathrm{a}-\mathrm{c}-\mathrm{t})^2/8\} - \mathrm{d}(\alpha). \tag{3.36}$$

Given Eqs. (3.20) and (3.21), this can be written as

$$SW_{LDC}|_{\alpha \leq \hat{\alpha}} - SW_{export} = \frac{1}{2}F_1^* - d(\alpha) = \frac{F^*}{2} + \left\{\frac{(2(a-w)-t)t}{8}\right\} - d(\alpha). \tag{3.37}$$

Given Proposition 4, under this situation, if $SW_{LDC}|_{\alpha \leq \hat{\alpha}} > SW_{export}$ for $\alpha = \hat{\alpha}$, the government chooses $\alpha_{opt} = \hat{\alpha}$ and the incumbent firm will choose the complete LDC strategy and the fake firm will not enter the market.

As $R < (4t(a-w) - 2t^2)/8$ at $\alpha = \hat{\alpha}$, we have

$$\frac{F^*}{2} + \left\{\frac{(2(a-w)-t)t}{8}\right\} - d(\widehat{\alpha}) > \frac{F^* + R}{2} - d(\widehat{\alpha}).$$

Thus, if $(F^* + R)/2 > d(\widehat{\alpha})$, then the social welfare under complete LDC strategy dominates that of the export strategy and the government chooses $\alpha_{opt} = \hat{\alpha}$.

Otherwise, $\alpha_{opt} = 1$ and given that $F^* < F < F_1^* - R$ the incumbent firm will choose the export strategy as the equilibrium strategy.

Next, we consider the case where $\alpha_1^* \geq \hat{\alpha}$ along with $F^* < F < F_1^* - R$. Given that $\alpha_1^* \geq \hat{\alpha}$, the incumbent firm chooses the complete LDC production strategy and

[29] For proof, see appendix A1.

acts as a monopolist in the interval $0 < \alpha \leq \hat{\alpha}$ and acts as a duopolist in the interval $\hat{\alpha} < \alpha \leq \alpha_1^*$. Given that social welfare under the complete LDC strategy increases with α, the government will choose $\alpha_{opt} = \alpha_1^*$ if $SW_{LDC}|_{\alpha>\hat{\alpha}} > SW_{\exp ort.}$ at $\alpha = \alpha_1^*$, otherwise, it chooses $\alpha = 1$,

$$SW_{LDC}|_{\alpha>\hat{\alpha}} - SW_{export.} = \frac{(a-w)^2(7\alpha+4)}{32} - \alpha A - C - d(\alpha) - \frac{(a-c-t)^2}{8}. \tag{3.38}$$

In this case, it can be shown that sufficient conditions that $SW_{LDC}|_{\alpha>\hat{\alpha}}$ is higher than SW_{export} at $\alpha = \alpha_1^*$ are $(F^* + R)/2 > d(\alpha_1^*)$ and $\alpha_1^* \geq 2/7$.[30, 31] If $\alpha_{opt} = \alpha_1^*$, the incumbent firm will choose the complete LDC strategy and the fake firm will enter the market as $\alpha_1^* \geq \hat{\alpha}$.

However, if $SW_{LDC}|_{\alpha>\hat{\alpha}}$ is lower than SW_{export} at $\alpha = \alpha_1^*$, the government chooses $\alpha_{opt} = 1$ and the incumbent firm chooses the export strategy and no FDI takes place.

Proposition 5 signifies the possibility of two interesting situations. First, there may be situations where the FDI will flow only in the assembly-line sectors, i.e., the firms are taking fragmented production strategies. Weak IPR restrictions, but low cost of foreign investment, leads to a situation where multinationals may shift the assembly-line activities in the LDCs. This situation will lead to transfer of embodied technology in the LDC, which does not lead to spillover of knowledge. Second, Proposition 5 also shows that there may be situations where the government is choosing an IPR protection rate that induces the entry of the fake firm in the market when the multinational is transferring the entire production process in the LDC. In this situation, disembodied technology is transmitted to less-developed economies thus leading the situation of knowledge spillover.

5 Concluding Remarks

Our model investigates how foreign firms' decision to produce in the LDC market depends on the IPR protection rate, fixed costs of plant installation, market size, and transport cost of transferring the finished product to the LDC. The impact of these parameters on the strategic entry decision of an MNC gives some interesting results. Summing up the results, we find that the entry decision of the MNC will initially depend upon the plant installation cost of the firm in the LDC. If the plant installation cost is sufficiently high, then the firm will find it more profitable to export the finished product to the LDC market. In such a case, the government will find it optimal to exercise no IPR restriction in the form of monitoring mechanism as assumed in the

[30] This in turn implies that $(a-w) - \sqrt{(a-w)^2 - 4((F+R) - F^*) - (a-w)^2/7} < t < (a-w) - 2\sqrt{A+F}$.

[31] For proof, see appendix A2.

model. Now if the plant installation cost to start off production in the LDC is below the critical value defined in the model then the choice of entry will be restricted between fragmentation or complete LDC production. In this case, the decision will depend on the probability of entry of the fake producer, i.e., IPR restrictions enforced by the government. In case of low probability of entry of the fake producer, the foreign firm will undertake complete LDC production. A basic assumption of the model is that a fake firm with counterfeit production can enter the market through technology spillover if and only if the foreign firm undertakes complete production in the LDC. The model also assumes that the government chooses the optimal rate of IPR restriction by maximizing the social welfare function. From the social welfare consideration, it has been observed that it is optimal for the government to impose some IPR restrictions given some sufficient conditions. However, even with some IPR restriction it has been observed that the fake firm may enter the market. Alternatively, if the extra plant installation cost for starting up complete production in LDC is high or the transport cost is low then the government does not choose any IPR restriction. In this situation, only assembly-line FDI takes place if the plant installation cost for starting up any production in the LDC is below a critical level.

The results enumerated briefly earlier find support from the recent policy change undertaken in the Indian Patent Act. Empirical analysis has also shown that MNCs are not willingly to transfer disembodied technology if they face the risk of product imitation in the LDC. Thus, in the presence of weak IPR the optimal policy for the MNC will be to transfer embodied technology and FDI will take place in the assembly-line sectors only.

6 Appendix A1

From Eq. (3.35),

$$\begin{aligned} SW_{LDC}|_{\alpha>\hat{\alpha}} - SW_{Frag} &= \frac{(a-w)^2(7\alpha+4)}{32} - \alpha A - C - d(\alpha) - \frac{(a-w-t)^2}{8} \\ &= \left[\frac{t(2(a-w)-t)}{8}\right] + \alpha\left[\frac{7(a-w)^2}{32} - A\right] - C - d(\alpha). \end{aligned}$$

This can be written as follows:

$$= \left[\frac{t(2(a-w)-t)}{8}\right] + \alpha\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha A + \left(\frac{7}{2}\alpha - 1\right)C - d(\alpha).$$

In this case, the first bracketed term is always positive and the second bracketed term is positive from footnote 12.

In this situation, $R < (4t(a-w) - 2t^2)/8$.

So,

$$\frac{(2t(a-w)-t^2)}{8} + \alpha\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha A + \left(\frac{7}{2}\alpha - 1\right)C - d(\alpha) >$$
$$R/2 + \alpha\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha A + \left(\frac{7}{2}\alpha - 1\right)C - d(\alpha).$$

At $\alpha = \alpha^*$,

$$SW_{LDC}|_{\alpha>\hat{\alpha}} - SW_{Frag} =$$
$$\frac{(2t(a-w)-t^2)}{8} + \alpha^*\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha^* A + \left(\frac{7}{2}\alpha^* - 1\right)C - d(\alpha^*) >$$
$$R/2 + \alpha^*\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha^* A + \left(\frac{7}{2}\alpha^* - 1\right)C - d(\alpha^*).$$

Sufficient conditions that the $SW_{LDC}|_{\alpha>\hat{\alpha}} \geq SW_{Frag}$ at $\alpha = \alpha^*$

$$R/2 \geq d(\alpha^*) \quad \text{and} \quad \alpha^* \geq 2/7$$

When $R < (4t(a-w) - 2t^2)/8(\text{or}(a-w) - \sqrt{(a-w)^2 - 4R} < t < (a-w))$, $\alpha^* \geq 2/7$, in turn, implies that a feasible range of t values defined as $(a-w) - \sqrt{(a-w)^2 - 4R - (a-w)^2/7} < t < (a-w) - 2\sqrt{A+F}$.

7 Appendix A2

$$SW_{LDC}|_{\alpha>\hat{\alpha}} - SW_{\text{export}}$$
$$= \frac{(a-w)^2(7\alpha+4)}{32} - \alpha A - C - d(\alpha) - \frac{(a-c-t)^2}{8}$$
$$= \frac{1}{2}F_1^* + \alpha\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha A + \left(\frac{7}{2}\alpha - 1\right)C - d(\alpha)$$
$$= \frac{F^*}{2} + \left\{\frac{(2(a-w)-t)t}{8}\right\} + \alpha\frac{7}{2}\left[\frac{(a-w)^2}{16} - A - C\right] + \frac{5}{2}\alpha A + \left(\frac{7}{2}\alpha - 1\right)C - d(\alpha).$$

At $\alpha = \alpha_1^*$, following the earlier logic as $R < (4t(a-w) - 2t^2)/8$, sufficient conditions that

$$SW_{LDC}|_{\alpha>\hat{\alpha}} > SW_{export.} \quad \text{at} \quad \alpha = \alpha_1^* \quad \text{are} \quad R + F^*/2 \geq d(\alpha_1^*) \quad \text{and} \quad \alpha_1^* \geq 2/7.$$

Now

$$\alpha_1^* = \frac{8(F^* - F)}{(a-w)^2} + \alpha^* = \frac{8(F^* - F)}{(a-w)^2} + \frac{4t(a-w) - 2t^2 - 8R}{(a-w)^2} \geq 2/7$$

$$\frac{8(F^* - F)}{(a-w)^2} + \frac{4t(a-w) - 2t^2 - 8R}{(a-w)^2} \geq 2/7$$

$$\frac{4t(a-w) - 2t^2 - (8(R+F) - F^*)}{(a-w)^2} \geq 2/7.$$

This, in turn, implies that

$$(a-w) - \sqrt{(a-w)^2 - 4((F+R) - F^*) - (a-w)^2/7} < t < (a-w) - \sqrt{A+F}.$$

References

Bascavusoglu E, Zuniga MP (2001) Intellectual property rights, technology and disembodied knowledge transfers cross borders: an empirical application. http://www.econ.kuleuven.ac.be/smye/abstracts/p502.pdf (Accessed 4 Jan 2014)

Banerjee DS, Banerjee (Chatterjee) T, Raychaudhuri A (2008) Optimal enforcement and anti-copying strategies to counter copyright infringement. Jpn Econ Rev 59(4):519–535

Eicher TK, Woo J (2005) Trade, FDI or acquisition: optimal entry modes for multinationals. J Dev Econ 77:207–228

Helpman E (1993) Innovation, imitation, and intellectual property rights. Econometrica 61: 1247– 1280

Javorcik SB (2004) The composition of foreign direct investment and protection of intellectual property rights: evidence from transition economies. Eur Econ Rev 48:39–62

Lai ELC (1998) International intellectual property rights protection and the rate of product innovation. J Dev Econ 55:133–153

Maskus KE (1998) The role of intellectual property rights in encouraging foreign direct investment and technology transfer. Duke J Comp Int L 9:109–161

Mattoo A, Marcelo O, Saggi K (2004) Mode of foreign entry, technology transfer, and FDI policy. J Dev Econ 75(1):95–111

Nahgavi A (2005) Strategic intellectual property rights policy and north south technology transfer. FEEM Working Paper No. 18.05 http://www.feem.it/Feem/Pub/Publications/WPapers/default.htm

Nicholson M (2000) Intellectual property protection, internalization and technology transfer. Discussion paper of Colorado University, working paper no. 00–11

Viswasrao S (1993) IPR and mode of technology transfer. J Dev Econ 78:474–493

Van Long N (2005) Outsourcing and technology spillovers. Int Rev Econ Financ 14:297–304

Wakasugi R, Ito B (2005) How enforcement of intellectual property rights affects the international technology transfer: evidence from Japanese MNCs. G-Sec working Paper No.1. June 2005. http://www.gsec.keio.ac.jp/wp/papers/wp1.pdf

Zigic K (1998) Intellectual property rights violations and spillovers in north-south trade. Eur Econ Rev 42:1779–1799

Zigic K (2000) Strategic trade policy, intellectual property rights protection and north-south trade. J Dev Econ 61:27–60

Chapter 4
An Example of Innovation-Inducing Tariff Protection

Swapnendu Banerjee

1 Introduction

Does trade liberalization foster innovation? This has been a debated issue both theoretically and empirically. In the Indian context, Desai (1980) and Lall (1989) argued quite strongly that tariff protection had eliminated innovation incentives whatsoever and made the domestic firms inward looking during the 1970s and 1980s. Whatever little innovation that took place were just minor innovations instead of at the frontiers of technology. Similar arguments have often been put forward for other developing countries in the countless debates over trade liberalization as an appropriate export promotion strategy. But casual empiricism observes a mixed experience in this regard. The lowering of tariffs has raised the level of research and development (R&D) in some countries, whereas it has lowered the level in others.

Theoretical analyses linking trade liberalization (or protection) and innovation have also remained inconclusive. Recent work by Acharyya and Banerjee (2012) has dealt with this issue in the vertical differentiation framework with some normative implications. In line with that paper this chapter illustrates with an example of how tariff protection can in fact induce quality innovation. I, however, refrain from making any normative predictions in this chapter.[1] That too much market power for the domestic firms is not conducive to innovation has also been argued by Porter (1990) and White (1974) among others. On the other hand, Rodrik (1992) derived just the opposite result. In a dynamic set-up, he demonstrated that liberalization slows down the pace of the productivity increase and delays technological catch-up since it shrinks the domestic monopolist's sales and thus reduces incentives to invest in cost-reducing technology. This essentially captures the Schumpeterian idea that monopoly power and innovation are positively related (Kamien and Schwartz 1982).

---

[1] Earlier, in one of his works, Acharyya (1995) showed how trade liberalization might foster quality innovation and, in certain situations, make it preferred to a cost innovation for the technologically constrained domestic firms.

S. Banerjee (✉)
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: swapnendu@hotmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_4,

This chapter provides a similar argument that protection fosters (quality) innovation in a vertically differentiated market. The present analysis considers a heterogeneous set of consumers with different preferences for the quality-differentiated good as in Mussa and Rosen (1978).[2] We, however, consider a discrete version of the standard model of quality choice developed in Cooper (1984) and used in Acharyya (1998) and Acharyya and Banerjee (2012).[3]

The rest of the chapter is organized as follows. Section 2 sets out the basic model and Sect. 3 derives the main results. Section 4 provides some concluding remarks.

2 Quality Choice with Discrete Consumer Types

We start with the Acharyya and Banerjee (2012) structure and consider a vertically differentiated good with observable quality indexed by $q \in [0, \overline{q}]$. Though the present state of scientific knowledge makes it possible to produce the good elsewhere over such a range of qualities, the domestic monopolist is technologically constrained in the sense that the technology he or she has access to allows him to produce only qualities within the range $[0, \tilde{q}_1]$ but not beyond that, where $\tilde{q}_1 < \overline{q}$. We shall later define quality innovation as a process whereby investing an exogenously given sum of F, he or she can learn the technical know-how to produce all $q \in [\tilde{q}_1, \overline{q}]$. With the marginal cost of production invariant with respect to the level of output but varying with the level of quality, the cost function in the pre-innovation stage for this domestic monopolist can be defined as:

$$\begin{aligned} C &= \bar{c}q^2 \qquad \forall q \in \left[0, \tilde{q}_1\right] \\ &= \propto \qquad \text{otherwise.} \end{aligned} \tag{4.1}$$

The domestic monopolist faces two types of consumers at home who differ in respect of their taste parameters (or what we shall define later as the marginal willingness to pay): $\alpha_2 > \alpha_1$. The number of consumers of each type is n_i. Each consumer buys, if at all, only one unit of the good. The net utility that type α_j consumer derives from the menu (q_j, P_j) is

$$U_j = u(q, \alpha_j) - P, \quad j = 1, 2. \tag{4.2a}$$

As is usually assumed in the literature, let $u(q, \alpha_2) > u(q, \alpha_1)$ and $u_q(q, \alpha_2) > u_q(q, \alpha_1)$, where $u_q(.)$ denotes the marginal utility of quality. Thus, the indifference curves between price and quality of the two types cross each other *only once*. Moreover, we assume that the marginal rate of substitution between price and quality (MRS_{pq}) is non-increasing in Z the quality level: $u_{qq}(q, \alpha_j) \leq 0$. Thus, the indifference curves

[2] Of late, incentives for quality innovation and choice of innovation type have been examined for a closed economy in a similar framework by Bandyopadhyay and Acharyya (2004) and Lambertini and Orsini (2000).

[3] See also Bandyopadhyay and Acharyya (2004).

are non-concave.[4] However, without any loss of generality, we assume a linear preference structure:

$$U_j = \alpha_j q - P. \tag{4.2b}$$

This follows all the assumed properties and as long as the cost function is sufficiently convex in the sense that $2\bar{c}\bar{q} > \alpha_2$, interior solution can be ensured,[5] i.e. qualities chosen by the monopolist at equilibrium will be strictly lower than $\overline{q}$. Note that α_j in this linear structure is the marginal utility of quality or the marginal willingness to pay for the α_j type.

A typical type α_j consumer participates in the market if

$$\alpha_j q \geq P. \tag{4.3}$$

and selects the menu (q_2, P_2) if the following self-selection constraint is satisfied

$$\alpha_j q_2 - P_2 \geq \alpha_j q_1 - P_1 \tag{4.4}$$

for $q_2 > q_1$.

Had there been no technological constraint, at a separating equilibrium the monopolist would have chosen the following qualities that maximize profit

$$\pi = n_1(\alpha_1 q_1 - \bar{c}q_1^2) + n_2\{\alpha_2(q_2 - q_1) + \alpha_1 q_1 - \bar{c}q_2^2\} : \tag{4.5}$$
$$\tilde{q}_1 = \frac{n_1\alpha_1 - n_2(\alpha_2 - \alpha_1)}{2n_1\bar{c}}, \tilde{q}_2 = \frac{\alpha_2}{2\bar{c}}.$$

Of course, we assume that the low end of the domestic market is sufficiently large, in the sense that $\frac{n_1}{n_2} > \frac{(\alpha_2 - \alpha_1)}{\alpha_1}$, to offer this separating menu (i.e. $\tilde{q}_1 > 0$). But given the technological constraint defined above, without investing in R&D, it is possible for the monopolist to offer at most $\tilde{q}_1$. In fact, he or she will offer a pooling menu (uniform price and quality to both types):[6] $(\tilde{q}_1, P_1 = P_2 = \alpha_1\tilde{q}_1)$.

However, the domestic consumers have the option to buy the good of higher quality from abroad. Suppose the world market for this quality-differentiated good is perfectly competitive and that producers abroad have identical cost of quality as the domestic monopolist except that they have no technological constraint. Thus, $C^* = \bar{c}q^2 \forall q \in [0, \bar{q}_1]$. Moreover, the preference structure and consumer types (or the marginal willingness to pay) are the same abroad as at home. These simplifying assumptions imply that competitive foreign producers will offer the same qualities in the domestic market under free trade (as well as under a specific tariff as we will argue later) as they do in the world market: $q_j{}^* = \frac{\alpha_j}{2\bar{c}}, j = 1, 2$. Thus, whereas the high quality is the same, the low quality offered by foreign competitive producers

[4] An important property of this type of preference structure is that the indifference curves (or, as we will define later, the self-selection constraints) are vertically parallel in the sense that MRS_{pq} is independent of the price level. The implication of this property will be made clear later.

[5] See Acharyya (1998).

[6] The same restriction on the distribution pattern that ensures existence of a separating menu will ensure that pooling menu is relatively profitable to the menu where the monopolist caters only to the high type by charging $(P_2 = \alpha_2\tilde{q}_1)$.

is greater than that offered by the domestic monopolist. This captures the quality distortion at the *bottom* under monopoly.[7] The world price of the low-quality variety being $P_l{}^* = \bar{c}q_1{}^{*2}$, from the self-selection constraint of the domestic consumers as defined in Eq. (4.4) it follows immediately that under free trade ($t = 0$), all will purchase the foreign good of quality $q_1{}^*$ even when the domestic monopolist charges the marginal cost price, $\bar{c}\tilde{q}_1^2$. What follows then is that without tariff protection the domestic monopolist cannot survive. Suppose that the import of the low-quality $\tilde{q}_1*$ is subject to a non-prohibitive per unit tariff 't_1'. This makes the tariff inclusive domestic price of the imported quality $q_1{}^*$ equal to $P_1^d = \bar{c}q_1{}^{*2} + t_1{}^8$. But the monopolist must charge a price $P_1(t_1)$ strictly less than $\bar{c}q_1{}^{*2} + t_1$. In particular,

Lemma 1 Given the technological constraint, to sell a strictly positive quantity the monopolist must charge a price lower than the tariff-inclusive import price,

$$P_1\,(t_1) = \alpha_1\tilde{q}_1 - \alpha_1 q_1{}^* + \bar{c}q_1{}^{*2} + t_1 - \varepsilon = (\bar{c}q_1^{*2} + t_1) - \alpha_1(q_1^* - \tilde{q}_1) - \varepsilon. \quad (4.6)$$

Proof Since the domestic monopolist in the pre-innovation stage does not have the technology to produce qualities beyond $\tilde{q}_1$, he or she can sell such a quality lower than the foreign quality only at a price $P_1(t_1)$ such that the self-selection constraint of the α_1 type consumers is satisfied. That is, the net utility from buying $\tilde{q}_1$ must be higher than that from $q_1{}^*$. A little manipulation of the self-selection constraint, $\alpha_1\tilde{q}_1 - P_1(t_1) > \alpha_1 q_1{}^* - P_1{}^d$, yields the price as defined in Eq. (4.6) which the monopolist can at most charge.

On the other hand, from the non-negative profit constraint of the monopolist, it is obvious that it will operate only if $P_1(t_1)$ covers the marginal cost of producing $\tilde{q}_1$. Given Lemma 1, this means there is a strictly positive tariff $\tilde{t}_1 = (q_1^* - \tilde{q}_1)[\alpha_1 - \bar{c}(q_1^* + \tilde{q}_1)]$ that must at least be offered to protect the domestic monopolist. That is, for all $t_1 \in [0, \tilde{t}_1]$, domestic production is zero. At the other extreme, the domestic government can set a prohibitive tariff t_{p1} on the low-quality import $q_1{}^*$ such that the monopolist can charge the monopoly price along the individual rationality constraint of the α_1 type consumers i.e. $P_1(t_1) = \alpha_1\tilde{q}_1$. Using Eq. (4.6) and for ε sufficiently close to zero we get the prohibitive tariff on the low-quality import as $t_{P1} = \alpha_1 q_1{}^* - \bar{c}q_1{}^{*2}$. These are illustrated in Fig. 4.1. Therefore, $\forall t \in [0, \tilde{t}_1 - \varepsilon]$, the low-type domestic consumers will buy the imported variety $q_1{}^*$, whereas $\forall t \in [\tilde{t}_1, t_{P1}]$ they will buy the variety $\tilde{q}_1$ offered by the domestic monopolist.

On the other hand, since later we will focus on the monopolist's incentive for quality innovation in response to trade liberalization, we assume that imports of higher quality variety (i.e. for $q > q_1{}^*$) are prohibited by a very high tariff level defined below. Given such a prohibitive tariff on high-quality variety t_{P2} (calculated similarly to t_{P1} had there been no technological constraint faced by the domestic monopolist),[8] the high-type domestic consumers must choose between the imported

[7] This is the standard result derived in the literature as long as the preference function follows the properties defined earlier. See Mussa and Rosen (1978) and Srinagesh and Bradburd (1989).

[8] Suppose that the home government imposes two different per-unit tariffs on the low-quality and high-quality varieties; t_1 on $q_1{}^*$ and t_2 on $q_2{}^* = \tilde{q}_2$.

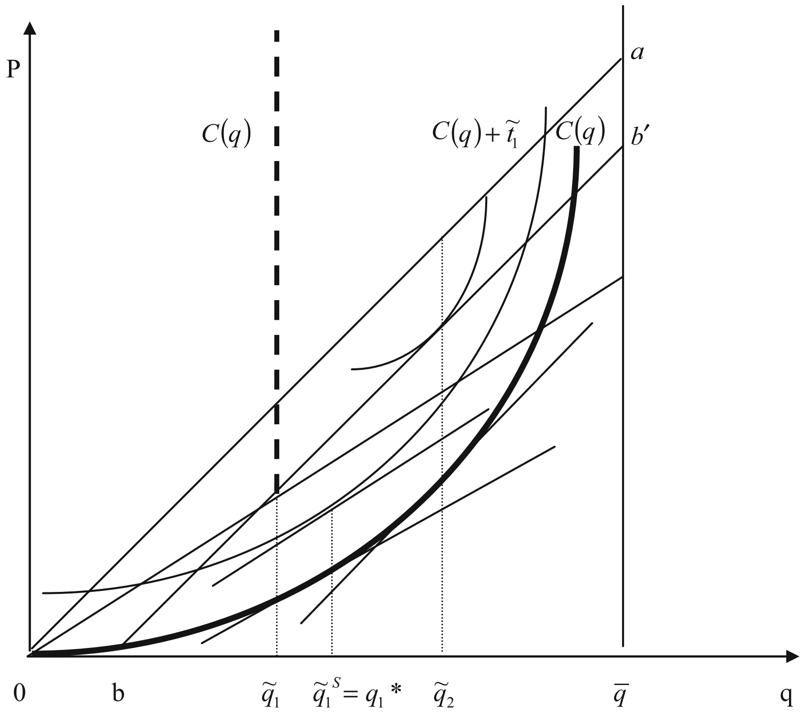


Fig. 4.1 Optimal qualities and tariff protection

variety $q_1{}^*$ at price $P_1^d = \bar{c}q_1{}^{*2} + t_1$ and the domestic variety $\tilde{q}_1$at$P_1(t_1)$ in the pre-innovation period. The following lemma specifies such a choice (given in Acharyya and Banerjee 2012):

Lemma 2 In the pre-innovation stage, with a prohibitive tariff t_{P2} on the high-quality imported variety, the high-type domestic consumers buy the imported variety $q_1{}^*\forall t_1 \in [0, t_{P1}]$ instead of the domestic variety of even poorer quality, $\tilde{q}_1$ (Acharyya and Banerjee 2012).

Proof Suppose, this is not the case. That is, suppose the high-type domestic consumers buy $\tilde{q}_1$. Then, by the self-selection constraint, the following must be true:

$$\alpha_2\tilde{q}_1 - P_1(t_1) > \alpha_2 q_1{}^* - P_1{}^d \quad \forall t_1.$$

By lemma 1, this boils down to,

$$(\alpha_2 - \alpha_1)\tilde{q}_1 > (\alpha_2 - \alpha_1)q_1^* - \varepsilon.$$

But $\tilde{q}_1 < q_1^*$and $\varepsilon > 0$. Hence, a contradiction.

But, of course, when the domestic monopolist can produce qualities beyond $\tilde{q}_1$ through quality innovation, for the tariff t_{P2} on $q_2{}^*$,, the high-type domestic consumers will buy the high quality $\tilde{q}_2 = q_2^*$ from the domestic monopolist. It is in this sense t_{P2} is termed as prohibitive tariff on high-quality imports. That is, t_{P2} becomes prohibitive only at the post-innovation stage. But in the pre-innovation stage, as Lemma 2 establishes, even such a high tariff on the high-quality imports fails to protect the high end of the domestic market for the domestic monopolist.

Thus, in pre-innovation, there will exist two quality-price bundles of the good X in the domestic market. One is $\tilde{q}_1$ at price $P_1(t_1)$, which is offered by the domestic monopolist to the α_1 type consumers, provided of course $t_1 \geq \tilde{t}_1$. The other is imported from the world market which is q_1^* at price $\bar{c}q_1^{*2} + t_1$ and these bundles are purchased by the α_2 type consumers. Therefore, in the pre-innovation stage, the domestic monopolist's profit will be

$$\prod (\Pr e - innov) = n_1 \left[\alpha_1 \tilde{q}_1 - \alpha_1 q_1^* + \bar{c} q_1^{*2} + t_1 - \bar{c}\tilde{q}_1^2\right]. \tag{4.7}$$

3 Tariff Protection and Innovation

Consider an innovation process whereby the monopolist attains the technology to produce all qualities within the range $[\tilde{q}_1, \bar{q}]$ by investing a fixed sum F in R&D. Once the technological constraint is removed through quality innovation, if undertaken at all, he or she can offer the profit-maximizing qualities to the low-type domestic consumers. Note that since with the world market being perfectly competitive, the monopolist cannot exercise his monopoly power in the domestic market for any non-prohibitive tariff on low-quality imports. He or she behaves *as if* a price taker and equates marginal cost with the marginal willingness to pay. Thus, instead of $\tilde{q}_1$, the profit-maximizing choice of quality is $\hat{q}_1 = \frac{\alpha_1}{2\bar{c}}$,which is the same as the foreign low quality.[9] Thus, the monopolist can now raise the price for α_1 type above $P_1(\mathrm{t}_1)$ defined in Eq. (4.6) to $(\bar{c}q_1^{*2} + t_1) - \varepsilon$ and can still induce them to buy from him.

On the other hand, given the prohibitive tariff t_{P2}, the domestic monopolist need not bother about the foreign high quality q_2* because even if he or she charges the maximum price $\alpha_2\hat{q}_2$ that extracts the entire surplus from the α_2-consumers, it is not incentive compatible for them to buy the same imported variety $q_2^*(= \hat{q}_2)$ at price $\bar{c}q_2^{*2} + t_{P2}$. But the monopolist must ensure that α_2-consumers do not purchase $\hat{q}_1(= q_1^*)$ offered to the α_1-consumers. That is, he or she must make $(\hat{q}_2, P_2)$ incentive compatible for them and this makes P_2 dependent on the level of tariff on low-quality imports. More precisely,

$$P_2\,(t_1) = \alpha_2\hat{q}_2 - \alpha_2 q_1{}^* + \bar{c}q_1^{*2} + t_1. \tag{4.8}$$

[9] Of course, $\hat{q}_2 = \tilde{q}_2$, i.e. the profit-maximizing quality offered to the high-type under price-taking behaviour is the same as that when he can exercise his monopoly power signifying the standard property of this class of models that there is no quality distortion *at the top*.

Thus, the monopolist's profit in the post innovative situation is

$$\prod(Post - innov) = n_1 t_1 + n_2 \left[\alpha_2 \hat{q}_2 - \alpha_2 q_1^* + \bar{c} q_1^{*2} + t_1 - \bar{c}\hat{q}_2^2\right]. \tag{4.9}$$

The relative (gross) gain from quality innovation is accordingly

$$RG(Q) = n_2 \left[(\hat{q}_2 - q_1^*)\left\{\alpha_2 - \bar{c}(\hat{q}_2 + q_1^*)\right\} + t_1\right] - n_1 \left[(q_1^* - \tilde{q}_1)\left\{\bar{c}(q_1^* + \tilde{q}_1) - \alpha_1\right\}\right]. \tag{4.10}$$

where

$$\frac{\partial RG(Q)}{\partial t_1} = n_2 > 0.$$

It is evident from the above expression that the net gain from quality innovation increases as tariff protection on the low-quality good is increased and the extent of such increase depends on the size of the high-type consumers. The reason for this is simple. A higher tariff on low-quality imports enables the domestic monopolist to extract greater surplus from the high-type consumer by offering $\tilde{q}_2$, in the post-innovation stage, at a higher price that leaves them with just the same net utility as they would have derived consuming the lower imported quality $\tilde{q}_1{}^*$. Obviously, total additional surplus extracted (which is in fact the incentive for quality innovation at the margin) depends on the number of the high-type consumers or the size of high-end of the domestic market. This is in essence a Rodrik (1992) type result where he observes that protection raises the incentive for cost innovation at the margin by ensuring a larger market for the incumbent firms.

For all, $t_1 \in [0, \tilde{t}_1]$, on the other hand, since $\prod(\text{Pr}\, e - innov) = 0$, relative (gross) gain from quality innovation equals

$$RG(Q) = n_1 t_1 + n_2 \left[\alpha_2 \hat{q}_2 - \alpha_2 q_1^* + \bar{c} q_1^{*2} + t_1 - \bar{c}\hat{q}_2^2\right], \tag{4.11}$$

with

$$\frac{\partial RG(Q)}{\partial t_1} = n_1 + n_2 > 0.$$

Thus, once again the net (gross) gain from quality innovation increases as level of protection (t_1) is raised. The only difference is that the incentive to upgrade quality now increases at a faster rate. Therefore,

Result 1 *Any specific or per unit tariff on import of low-quality variety, regardless of whether it is protective or not, raises the incentive for quality innovation.*

Proof Follows from Eq. (3.6) and Eq. (3.7).

The innovation decision, on the other hand, is illustrated in Fig. 4.2. Given the positively sloped $RG(Q)$ curve, we can define a level of (fixed) innovation cost, $\tilde{F}_q$ such that the relative (net) gain from quality innovation (for t_{P1}) is exactly zero. For all F_q greater than this, of course, the domestic monopolist will not innovate. But,

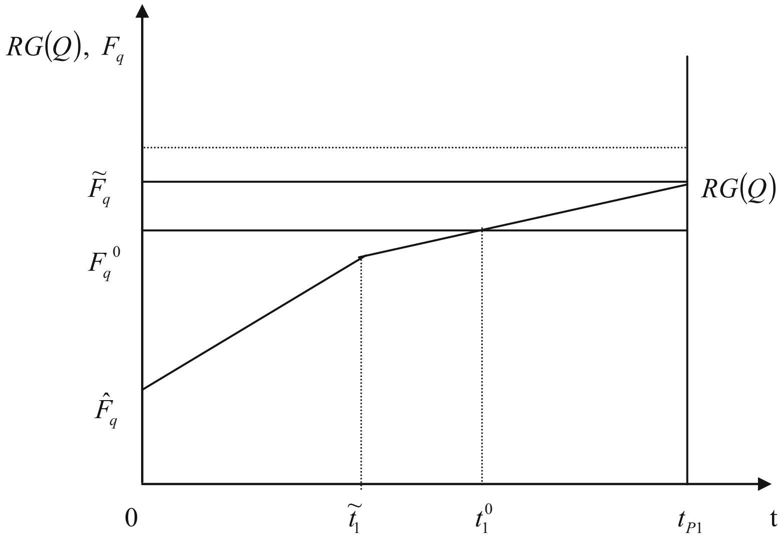


Fig. 4.2 Tariff and quality innovation

for $F_q \in [0, \tilde{F}_q]$, the innovation decision depends on the level of protection. For example, for $F_q = F_q{}^0$, the monopolist invests in R&D only for a $t_1 \in [t_1{}^0, t_{P1}]$. That is, given an initial situation where the tariff on low-quality import is smaller than $t_1{}^0$, the government can induce the domestic monopolist to innovate by raising the tariff rate beyond $t_1{}^0$.

4 Conclusion

In a vertically differentiated domestic market with a single domestic firm and competitive producers abroad, we have demonstrated that tariff protection raises the incentive for quality innovation. This corroborates Rodrik (1992) who found tariff is conducive to cost innovation in a dynamic set up with homogeneous good, but in sharp contrast to oft-quoted argument that trade liberalization fosters innovation.

The assumption of discrete consumer type with full market coverage rules out any demand effect whatsoever that may arise due to tariff protection. An obvious extension will, therefore, be to consider continuum of types along with partial market coverage *at the initial equilibrium* and examine implications of change in demand through change in the extent of market coverage.

Acknowledgement I thank Rajat Acharyya, Nobuhiro Kiyotaki, Sugata Marjit and Sugato Bhattacharya for their comments and suggestions on an earlier version of the chapter. I am solely responsible for any remaining errors.

References

Acharyya R (September 9, 1995) Liberalization, innovation and exports. Econ Polit Wkly 30(36):2269–2271

Acharyya R (1998) Monopoly and product quality: separating or pooling menu? Econ Lett 61: 187–194

Acharyya R, Banerjee S (2012) On tariff and quality innovation in a market with discrete preferences. Econ Model 29:917–925

Bandyopadhyay S, Acharyya R (2004) Process and product innovation: complementarity in a vertically differentiated monopoly with discrete consumer types. Jpn Econ Rev 55(2):175–200

Cooper R (1984) On allocative distortions in problems of self-selection. Rand J Econ 15:569–577

Desai AV (1980) The origin and direction of industrial R&D in India. Res Policy 9(1):74–96

Kamien MT, Schwartz NL (1982) Market structure and innovation—a survey. J Econ Lit 13(1):1–37

Lall S (1989) India's technological capacity: effects of trade, industrial, science and technology policies. In: Fransman M, King K (eds) Technological capability in the third world. Macmillan Press, London

Lambertini L, Orsini R (2000) Process and product innovation in a vertically differentiated monopoly. Econ Lett 68:333–337

Mussa M, Rosen S (1978) Monopoly and product quality. J Econ Theory 18:301–317

Porter M (1990) The competitive advantage of nations. Macmillan, London

Rodrik D (1992) Closing the productivity gap: does trade liberalization really help? In: Helleiner GK (ed) Trade policy, industrialisation and development: new perspectives. Clarendon Press, Oxford

Srinagesh P, Bradburd RM (1989) Quality distortion by a discriminating monopolist. Am Econ Rev 79(1):96–105

White L (December 1974) Industrial organization and international trade: some theoretical considerations. Am Econ Rev 64(6):1013–1020

Chapter 5
Import Restrictions, Capital Accumulation and Use of Child Labour: A General Equilibrium Analysis

Runa Ray

1 Introduction

The use of child labour in different activities mostly in developing countries has attracted attention in the academic and policy circles on the question of designing appropriate policy interventions to mitigate its incidence. The adoption of labour standards by the World Trade Organization (WTO) hovers around the question of whether the use of trade sanctions on the products that use child labour can be effective in reducing the incidence of child labour or not. As developing countries integrate into the world economy and increasingly rely on export markets to sell their products, rich countries can use the threat of trade sanctions to adopt policies that attempt to curtail child labour. In December 2001, European Union (EU) trade commissioner Pascal Lamy announced that EU foreign ministers have approved a preferential tariff scheme for countries that adhere to International Labour Organization (ILO) labour standards including child labour. Both the Clinton and Bush administrations have voiced a commitment to a similar policy and in 1999 the USA imposed quotas on Cambodia's garment industry because of its poor working conditions and use of child labour. Thus, trade sanctions on the exports of least developed countries (LDCs) by the rest of the world may be a vehicle for the reduction of the incidence of child labour in poor developing countries. However, a caveat may be noted. Many have argued for an international labour standard policy that requires the elimination of child labour for access to developed countries' markets. In some cases, the argument is simply providing the cover for standard issue of protectionism. However, it is often by a genuine concern for the welfare of children in developing countries. If this is indeed the motivation, the implementation of trade sanctions to enforce an international standard against the use of child labour is likely to have perverse consequences. Except in unusual cases, effective sanctions would make the families of child workers worse off. If sanctions are effective, they will generally have the consequence of lowering the price of the goods produced with child labour. This reduction in the prices would

R. Ray (✉)
Department of Economics, Vidyasagar College, Kolkata, India
e-mail: runa_maju@yahoo.co.in

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_5, © Springer India 2014

lead to lower wages for child workers. Those children who remain in the labour force are worse off because they are paid less. As a consequence of such low wage payment, some children will stop working and go to school. However, if child labour is indeed a means of coping with desperate poverty, families are sending children to work only when the current value of the income they earn is greater than the (discounted value) of the future benefits of the education. Lowering the wages of the child labour to induce the family to send the child to school makes the family worse off.

The integration of the world economy through the process of globalization has unleashed many forces that have affected the economic environment and the conditions of the people in both the developed and the developing world, and trade policy reforms through the WTO have accentuated such changes. One such area where there have been significant effects in the living conditions through the process of globalization-induced economic changes has been with respect to the *incidence of child labour in the production process of both developed and developing countries*. Although there have been significant changes in the size and type of child labour used in different parts of the globe in recent years,the problem still remains quite serious and policy options are being debated in the literature on economic policy and development economics regarding the optimality or otherwise of alternative measures to combat child labour. One such important area of policy intervention could be the use of trade policy instruments to minimize the incidence of child labour particularly in developing countries. In this chapter, we shall develop a general equilibrium framework for a less developed economy using child labour in one of its sectors, and examine the effects of policies of import restriction and of domestic capital accumulation on the incidence or use of child labour in such an economy.

2 Evidence

The Second Global Report (2006) presents trends in child labour in the world on the basis of new estimates that are fully comparable with those published in 2002. The newly emerging picture confirms that the incidence of child labour is declining. Some interesting patterns of trends about child labour in the world were noted by the Second Global Report on Child Labour (2006). First of all, the new estimates suggest that there were about 317 million economically active children in the age group of 5–17 years in 2004, of whom 218 million could be regarded as child labourers. Of the latter, 126 million were engaged in hazardous work. The corresponding figures for the narrower age group of 5–14-year-olds are 191 million economically active children, 166 million child labourers and 74 million children in hazardous work. Secondly, the number of child labourers in both age groups of 5–14 and 5–17 years fell by 11 % over the 4 years from 2000 to 2004, but the decline was much greater for those engaged in hazardous work: by 26 % for the 5–17 year age group and 33 % for 5–14-year-olds. It was estimated that the incidence of child labour (percentage of children working) in 2004 was around 13.9 % for the age group of 5–17 years, compared to 16 % in 2000. The proportion of girls among child labourers, however, remained steady over the years.

Thus, the global picture that emerged was that child work was declining, and the more harmful the work and the more vulnerable the children involved, the faster was the decline. Latin America and the Caribbean had made the greatest progress—the number of children at work had fallen by two thirds over the past 4 years, with just 5 % of children now engaged in work. The least progress was made in sub-Saharan Africa, where the rates of population growth, human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) infection and child labour remained alarmingly high. In fact, the involvement of children in work had declined in all three categories over 2002–2006—both in absolute and in relative terms—and among all age groups and both sexes. Two trends were discernible, the qualitative decline in child labour (the younger and more vulnerable the child and the more hazardous the work, the greater the decline) and the massive declines shown by the Latin America and Caribbean region, which put it on par with some developed and transition economies. Particularly impressive had been the decline in use of child labour in hazardous works.

The 2006 ILO Second Global Report on Child Labour showed an overall declining trend for child labour, but the Third Global Report on Child Labour (2010) indicated a mixed picture. On the positive side, there was a welcome decline of child labour among girls and among children in hazardous work, but overall indications were that progress was uneven, neither fast enough nor comprehensive enough to reach the goals that had been set. Since 2006, there has been a slowing down of the global pace of reduction: Child labour among boys and young people in the 15–17 age group has risen, and in sub-Saharan Africa progress has stalled. The bottom line is that some 215 million children across the world are still trapped in child labour and a staggering number of 115 million children are still exposed to hazardous work. The Asia-Pacific, Latin America and the Caribbean regions continue to reduce child labour, while sub-Saharan Africa has witnessed an increase in both relative and absolute terms with one in four children in the region being engaged in child labour. Interestingly, there have been increases in child labour among boys but considerable and welcome decreases among girls. There are now concerns that the global economic downturn will put a further brake on progress towards the 2016 goal for the elimination of the worst forms of child labour and render the challenge of achieving the Millennium Development Goals (MDGs) all the more difficult.

ILO's estimation of child labour trends for the period 2004–2008 shows the following:

- Globally, child labour continues to decline, albeit to a lesser extent than before. There are still 215 million children trapped in child labour.
- The number of children in hazardous work, often used as a proxy for measuring the extent of the worst forms of child labour, is declining, particularly among those below 15 years of age. The overall rate of reduction, however, has slowed. There are still 115 million children in hazardous work.
- Children's work is declining in the Asia-Pacific region and in Latin America and the Caribbean, but it is increasing in sub-Saharan Africa.

- Among girls, there is a significant decrease. Among boys and older children (age 15–17), however, the trends show some increase.
- Most child labourers continue to work in agriculture. Only one in five working children is in paid employment. The overwhelming majority are unpaid family workers.

From the above estimations, two major trends are discernible. First, among 5–14-year-olds, the involvement of children in work has declined across the board over the past 4 years, both in absolute and in relative terms. The number of children in child labour in this age group declined by 10 % and the number of children in hazardous work fell by 31 % over 4 years. This downward trend is in line with the previous estimates and confirms that child labour declines faster in its worst forms and among the more vulnerable. Second, the encouraging trends for the younger age group seem to have reversed in the case of older children, 15–17 years old. The results indicate that child labour in this age group has increased from 52 to 62 million corresponding to a change of 20 % from 2004 to 2008.

Although there are registered cases of children under 5 years working, almost all child labour occurs in the age group of 5–17. These children, roughly 70 % of all 'children in employment', are classified as child labourers because they are either under the minimum age for work or above that age and engaged in work that poses a threat to their health, safety or morals, or are subject to conditions of forced labour. The number of children in child labour has continued its declining trend, falling by 3 % between 2004 and 2008. The corresponding incidence rate declined from 14.2 to 13.6 %.

3 The Framework

The objective of the present chapter is to provide an analytical framework to investigate the effectiveness of trade and non-trade policies in the presence of child labour use. We take the trade policy interventions in this chapter as equivalent to the increase in protectionism in the import-competing sector in the home country. We take the expansion of domestic capital stock as the non-trade policy intervention. We consider a $3 \times 3 \times 3$ competitive general equilibrium framework of a small open less developed economy which exports products produced by child labour.

3.1 The Model

The economy we consider is a small, open less developed economy with the following characteristics:

1. There are three sectors in the economy:
   a. Among the three, two are informal sectors. The informal sector is comprised of 'plucky' micro entrepreneurs who choose to operate informally in order to avoid costs and effort of formal registration.

1. One of the informal sectors is the export sector producing agricultural product (X).
2. The service sector attempts to incorporate the other informal sector into the model. The service sector produces non-traded final commodity (Z).

b. The third sector is the formal sector and it is the import-competing sector producing a manufacturing commodity (Y).

2. Three kinds of inputs are available in the economy—adult labour, child labour and capital.
3. X and Z are produced using adult labour and child labour but Y is produced using adult labour and capital.

Thus, in our model, adult labour is perfectly mobile between informal and formal sectors but child labour is perfectly mobile within informal sectors and capital is the specific factor of production.

4. Owing to effective wage legislation and unionization of labour, the adult wage rate in the formal sector is fixed at W^*, which is greater than the competitive informal sector adult wage rate, W.
5. The adult labour allocation mechanism is of the following type:
   Adult workers first try to get employment in the formal manufacturing sector and those who are unable to find employment in the said sector are automatically absorbed in any one of the informal sectors, as there is complete wage flexibility in the latter sector.
6. Following Basu and Van (1998), we make the assumption of 'substitution' in the informal sector, which suggests that adult labour is a substitute for child labour. Gupta (2000) has done theoretical work in determining the wage rate of a child worker. In our model, it is assumed that an adult labour is equivalent to β units of child labour where $\beta > 1$. The adult labour in the informal sectors earns W, while a child labour earns $\frac{W}{\beta}$.
7. Production functions satisfy constant return to scale with positive but diminishing return to each factor.
8. All inputs are fully employed. Stock of capital and adult labour force are exogenously given.
9. Owing to the small open economy, assumption prices of X and Y are given internationally. Since Z is produced and consumed domestically, its price is determined within the country by demand and supply forces.
10. We assume each working family comprises of one adult and a certain number of child workers.

The following symbols are used in the formal presentation of the model:

a_{Lai} Adult labour–output ratio in the ith sector, $i = $ X, Y, Z,
a_{Lci} Child labour–output ratio in the ith sector, $i = $ X, Z,
a_{KY} Capital–output ratio in the formal sector (Y),
Pi World price of the ith good, $i = $ X, Y,
Pz Domestically determined price of Z,
W Adult wage rate in the informal sectors(X, Z),

$\frac{W}{\beta}$ Child wage rate in the informal sectors (X, Z),
W^* Institutionally given wage rate in formal sector (Y),
r Rate of return to capital,
L^A Total adult labour endowment,
L_C^S Aggregate supply of child labour,
K Fixed stock of domestic capital in the economy,
L_1 Number of adult workers engaged in the informal sectors,
I Aggregate income of the formal sector labourers and capitalists and
t Ad-valorem tariff rate on the import of Y.

4 General Equilibrium Analysis

A general equilibrium of the system is represented by the following set of equations:

$$a_{LaX}W + a_{LcX}\frac{W}{\beta} = P_X, \tag{5.1}$$

$$a_{LaY}W^* + a_{KY}r = P_Y(1+t), \tag{5.2}$$

$$a_{LaZ}W + a_{LcZ}\frac{W}{\beta} = Pz, \tag{5.3}$$

$$a_{LaX}X + a_{LaY}Y + a_{LaZ}Z = L^A, \tag{5.4}$$

$$a_{KY}Y = K, \tag{5.5}$$

$$a_{LaY}W^*Y + rK = I, \tag{5.6}$$

$$Z(Pz, I) = Z \text{ with } \frac{\partial Z}{\partial \mathrm{Pz}} < 0, \frac{\partial Z}{\partial \mathrm{I}} > 0, \tag{5.7}$$

$$a_{Lcx}X + a_{Lcz}D(Pz, I) = L_C^S(W, L_1) \text{ with } \frac{\partial L_C^S}{\partial W} < 0, \frac{\partial L_C^S}{\partial L_1} > 0, \tag{5.8}$$

where

$$L_1 = a_{Lax}X + a_{Laz}Z. \tag{5.9}$$

Equations (5.1)–(5.3) are the usual price-unit cost equality conditions in the three sectors of the economy. Equations (5.4) and (5.5) are the endowment equations of adult labour and capital. Total income of the formal sector consists of wage income of the formal sector workers and the rental income of the capital owners. It is shown in Eq. (5.6). The demand for non-traded final commodity (shown by Eq. (5.7)) produced by child and adult workers comes from the richer section of the society. The demand function obeys the usual price and income effect. Finally, child labour market equilibrium condition is shown by Eq. (5.8). The left-hand side reveals the demand for child labour which depends on the output of X and Z. The aggregate supply function is assumed to depend on the adult wage rate as well as on the number of people engaged in the informal sector. If the wage rate of the adult worker rises, they send fewer children to the job market. On the other hand, the supply of child labour varies positively with the number of the informal sector adult workers because these workers earn low wage income and send their children to work.

The working of the model is described as follows: There are eight endogenous variables in the system and eight independent equations. The endogenous variables are W, r, P_Z, X, Y, Z, I and L_C^S. The parameters in this model are: P_X, P_Y, β, L^A, K and W^*. We should note that the system possesses the decomposition property since the two unknown input prices W (hence $\frac{W}{\beta}$) and r can be determined from the price system alone, independent of the output system. Once the factor prices are known, factor coefficients a_{ij}'s are also known. P_Z is to be obtained from Eq. (5.3). Since Z is non-traded, its price is determined by the demand–supply mechanism—the supply of Z is constrained by the demand for Z in the formal sector. Y can be obtained from Eq. (5.5) and substituting the value of Y and r in Eq. (5.6) we get I. Similarly, once P_Z and I are determined, Z can also be obtained from Eq. (5.7). X will be solved from Eq. (5.4) and finally, from Eq. (5.8) L_C^S will be derived.

5 Comparative Static Exercises

We now investigate the impact of trade and non-trade policies on child labour supply. The trade policy undertaken by the domestic economy is protectionism in the country's import-competing sector, whereas the non-trade policy constitutes economic expansion or growth due to expansion of the capital stock of the economy.

A. Protectionism in Import-Competing Sector and Impact on Child Labour Supply Much of the recent policy debate and controversy surrounding globalization and the WTO have been focused on the issue of child labour in poor countries. On one hand, opponents of market integration argue that globalization may increase the wages paid to the working children or increase the earning opportunities of children in poor economies, thereby increasing child labour. Some further suggest that rich countries should restrict the sale of goods from developing countries that lack or do not enforce child labour laws. Theoretical models by Maskus (1997) and Ranjan (2001) show that trade sanctions or import tariffs against countries that use child labour do not necessarily reduce the incidence of child labour. On the other hand, increase in household income and increased availability of schooling opportunities in low-income countries could help in reducing child labour (Basu (1999)).

To find out the impact of increased protectionism on child labour supply, we will first examine the impact on factor prices and composition of output.

Due to protectionism, there will be no change in the adult as well as child wage rate as revealed by Eq. (5.1). However, the rental rate will change. To find out the impact on rental rate we differentiate Eq. (5.2):

$$W^* d(a_{L_{aY}}) + a_{KY} dr + r da_{KY} = P_Y(dt)$$

$$\Rightarrow \theta_{L_{aY}} \hat{a}_{LaY} + \theta_{KY} \hat{r} + \theta_{KY} \hat{a}_{KY} = dt$$

$$\Rightarrow \theta_{KY} \hat{r} = dt$$

[Since from the condition of cost minimization $\theta_{LaY} \hat{a}_{LaY} + \theta_{KY} \hat{a}_{KY} = 0$]

$$\Rightarrow \hat{r} = \frac{dt}{\theta_{KY}} > 0. \tag{5.10}$$

Proposition 1 A rise in price of import-competing good produces a Stolper–Samuelson effect and raises the rental rate leaving the wage rates unchanged.

Now, we try to find out the impact of protectionism on the output of import-competing sector. In our model, capital is specific to the import-competing sector. Differentiating Eq. (5.5) we get

$$\hat{Y} = -\hat{a}_{KY} = \theta_{LaY} \sigma_Y \hat{r} = \theta_{LaY} \sigma_Y \frac{dt}{\theta_{KY}} > 0, \tag{5.11}$$

where σ_Y is the elasticity of substitution, given by $\sigma_Y = \left(\frac{a_{K\hat{Y}} - a_{L_{\hat{a}}Y}}{\hat{W}^A - \hat{r}} \right)$ and cost minimization condition for the producer of Y entails that $\theta_{KY} a_{K\hat{Y}} + \theta_{La\hat{Y}} = 0$

$$\therefore a_{La\hat{Y}} = -\theta_{KY} \sigma_Y \left(\hat{W}^A - \hat{r} \right) \text{ and}$$

$$a_{K\hat{Y}} = \theta_{LaY} \sigma_Y \left(\hat{W}^A - \hat{r} \right).$$

In our model since $\hat{W}^A = 0$, we have $\therefore a_{K\hat{Y}} = -\theta_{LaY} \sigma_Y (\hat{r})$. We have proposition 2 next.

Proposition 2 Due to increased protectionism in the import-competing sector, formal sector (import-competing sector) of the economy must expand.

The intuitive explanation is as follows: As the price of an import-competing good has increased, profit-maximizing producers will be interested to produce more units of Y leading to an expansion of the formal sector. Alternatively, producers will curtail the use of capital in per unit Y production as its price increases. This is also reflected in Eq. (5.11). To ensure full employment of capital, output of the import-competing sector must expand.

Taking total differentiation of Eq. (5.6) we get,

$$W^* Y \left(da_{LaY} \right) + W^* a_{LaY} dY + K dr = dI$$

or

$$\left(W^*Ya_{LaY}\right)\hat{a}_{LaY} + (W^*Ya_{LaY})\hat{Y} + Kr\hat{r} = dI$$

or

$$\left(W^*a_{LaY}Y\right)\left(\hat{a}_{LaY} + \hat{Y}\right) + Kr\hat{r} = dI$$

or

$$\left(W^*a_{LaY}\right)\left(\hat{a}_{LaY} - \hat{a}_{KY}\right) + Kr\hat{r} = dI$$

or

$$(W^*a_{LaY}Y)\sigma_Y\hat{r} + Kr\hat{r} = dI \text{ } (\because \hat{a}_{La}Y = \sigma_Y\theta_{KY}\hat{r} \text{ and } \hat{a}_{KY} = -\sigma_Y\theta_{LaY}\hat{r} < 0)$$

$$dI = \frac{[W^*a_{LaY}Y\sigma_Y + Kr]dt}{\theta_{KY}} > 0. \tag{5.12}$$

Thus, the income of the richer section of the society increases. Differentiating Eq. (5.7) we get,

$$dZ = Z_I dI = \frac{Z_I[W^*a_{LaY}Y\sigma_Y + Kr]dt}{\theta_{KY}} > 0. \tag{5.13}$$

Thus, the output of the non-traded final good-producing sector goes up.

Finally, to find out the impact on the export sector, we differentiate Eq. (5.4) and get,

$$a_{LaX}\frac{dX}{dt} + a_{LaY}\frac{dY}{dt} + Y\frac{da_{LaY}}{dt} + a_{LaZ}\frac{dZ}{dt} = 0$$

or

$$a_{LaX}\frac{dX}{dt} = -\frac{a_{LaY}\left(Y\theta_{LaY}\sigma_Y\right)}{\theta_{KY}} - \frac{Y\left(a_{LaY}\theta_{KY}\sigma_Y\right)}{\theta_{KY}} - \frac{a_{LaZ}Z_I}{\theta_{KY}}\left(W^*a_{LaY}\sigma_Y Y + Kr\right)$$

or

$$a_{LaX}\frac{dX}{dt} = -a_{LaY}\frac{Y\sigma_Y}{\theta_{KY}}\left(\theta_{LaY} + \theta_{KY}\right) - \frac{a_{LaZ}Z_I}{\theta_{KY}}\left(W^*a_{LaY}\sigma_Y Y + Kr\right)$$

or

$$\begin{aligned}\frac{dX}{dt} &= -\frac{a_{LaY}Y\sigma_Y}{a_{LaX}\theta_{KY}} - \frac{a_{LaZ}Z_I}{a_{LaX}\theta_{KY}}\left(W^*a_{LaY}\sigma_Y Y + Kr\right) \\ &= -\frac{a_{LaY}Y\sigma_Y}{a_{LaX}\theta_{KY}}\left[1 + a_{LaZ}Z_I W^*\right] - \frac{aL_{aZ}Z_I Kr}{a_{LaX}\theta KY} \\ &= -\frac{1}{a_{LaX}\theta_{KY}}\left[\left\{\left(a_{LaY}Y\sigma_Y\right)\left(1 + a_{LaZ}Z_I W^*\right)\right\} + \left(a_{LaZ}Z_I Kr\right)\right] < 0. \end{aligned} \tag{5.14}$$

Thus, the export sector will contract.

Proposition 3 Due to protectionism in the import-competing sector, one of the informal sectors (non-traded final good producing sector) will expand and the export sector (another informal sector) will contract.

The intuitive explanation is quite straightforward: As the income of the richer section of the population increases due to the expansion of the formal sector, demand for the non-traded good boosts up leading to an expansion of that sector. Since adult labour endowment is fixed and adult labour is used in the production of all three commodities, the export sector must contract to release the required demand for adult labour generated in the other two sectors.

To find out the aggregate impact on the informal sector of the small open economy we proceed as follows:

Let,

$$M = X + Z \tag{5.15}$$

$$\therefore dM = dX + dZ$$

$$= -\frac{dt}{a_{LaX}\theta_{KY}}\left[\left(a_{LaY}\sigma_Y Y\right)\left(1+a_{LaZ}Z_I W^*\right)+\left(a_{LaZ}Z_I Kr\right)\right]+\frac{Z_I dt}{\theta_{KY}}\left(W^* a_{LaY}\sigma_Y Y+Kr\right)$$

$$= -\frac{\left(a_{LaY}\sigma_Y Y\right)dt}{\theta_{KY}}\left[\frac{\left(1+a_{LaZ}Z_I W^*\right)}{a_{LaX}} - \left(W^* Z_I\right)\right]+\frac{Z_I Krdt}{\theta_{KY}}\left(1-\frac{a_{LaZ}}{a_{LaX}}\right)$$

$$= -\frac{(a_{LaY}\sigma_Y Y)dt}{\theta KY}\left[\frac{1}{a_{LaX}} - W^* Z_I\left(1-\frac{a_{LaZ}}{a_{LaX}}\right)\right]+\frac{Z_I Krdt}{\theta_{KY}}\left(1-\frac{a_{LaZ}}{a_{LaX}}\right). \tag{5.16}$$

Suppose,

$$\frac{a_{LaZ}}{a_{LaX}} < 1. \tag{5.17}$$

Now, for simplicity we assume

$$a_{LCX} = a_{LCZ}, \tag{5.18}$$

$$\frac{a_{LaZ}}{a_{LCZ}} < \frac{a_{LaX}}{a_{LCX}} \tag{5.19}$$

i.e. Z is a child labour-intensive commodity.

In this case, overall impact on the informal sector is ambiguous. On the other hand, suppose

$$\frac{a_{LaZ}}{a_{LaX}} > 1, \tag{5.20}$$

$$\text{then}\,\frac{a_{LaZ}}{a_{LCZ}} > \frac{a_{LaX}}{a_{LCX}} \tag{5.21}$$

i.e. Z is the adult labour-intensive commodity. In this case, the informal sector will unambiguously contract. Thus, we have proposition 4 next.

Proposition 4 Due to protectionism in the import-competing sector, the informal sector must contract when the exported product is child labour-intensive in nature and the result becomes ambiguous when the exported product is adult labour-intensive in nature.

Now, $\left[a_{LaX}\frac{dX}{dt}+a_{LaZ}\frac{dZ}{dt}\right]$

$$= -\frac{a_{LaX}}{a_{LaX}\theta_{KY}}\left[\left(a_{LaY}\sigma_Y Y\right)\left(1+a_{LaZ}Z_I W^*\right)+a_{LaZ}Z_I Kr\right]+\frac{a_{LaZ}Z_I}{\theta_{KY}}\left(W^* a_{LaY}\sigma_Y Y+Kr\right) \tag{5.22}$$

$$= -\frac{a_{LaY}\sigma_Y Y}{\theta_{KY}} < 0 \tag{5.23}$$

$$\therefore \frac{dL_C S}{dt} = \frac{\partial L_C S}{\partial\left(a_{LaX}X + a_{LaZ}Z\right)}\left(-\frac{a_{LaY}\sigma_Y Y}{\theta_{KY}}\right) < 0, \tag{5.24}$$

$$\frac{\partial L_C S}{\partial\left(a_{LaX}X + a_{LaZ}Z\right)} > 0. \tag{5.25}$$

Proposition 5 Stringent protectionism in the import-competing sector must lead to a decline in the child labour supply.

In our model, there are two determinants of child labour supply—one is the adult wage rate which is unchanged due to protectionism in the import-competing sector. But the number of adult workers engaged in informal sector must contract, since some part of the released adult labour from X product will be recruited in the formal sector and they will not send their children to the job market. Hence, child labour supply in the economy must fall.

B. Impact of Non-Trade Policy (Economic Expansion) on Child Labour Supply Non-trade policy in our model is captured by economic expansion due to an increase in domestic capital stock. Interesting works in this area have been done by Gupta (2003) and Swaminathan (1998). In this case, without tariff on the country's importable commodity *Y*, Eq. (5.2) of our model is rewritten in the form

$$a_{LaY}W^* + a_{KY}r = P_Y. \tag{5.26}$$

As the size of the capital stock of the economy goes up, the factor prices remain unaffected due to the decomposition property of the system but the product mix changes. Also, there will be no change in the price of the non-traded good since there will be no change in factor prices. Thus, we have the following proposition:

Proposition 6 Economic expansion by increase in domestic capital stock will not lead to any change in factor prices.

To find out the impact on output of different sectors, taking total differentiation of Eq. (5.5) we get,

$$a_{KY}dY = dK(\because da_{KY} = 0 \text{ as } dW = dr = 0). \tag{5.27}$$

or

$$dY = \frac{1}{a_{KY}} dK > 0$$

or

$$\frac{dY}{Y} = \frac{dK}{K} \left(\because a_{KY} = \frac{K}{Y} \right)$$

or

$$\hat{Y} = \hat{K} > 0 \tag{5.28}$$

Proposition 7 Capital expansion will lead to enhance the size of the import-competing sector.

Since capital is treated as a specific factor in the import-competing sector of the economy, full utilization of the expanded capital stock must lead to an expansion of the import-competing sector.

Taking total differentiation of Eq. (5.6) we get,

$$dI = a_{LaY} W^* dY + rdK(\because da_{LaY} = 0)$$

or

$$dI = \frac{a_{LAY} W^*}{a_{KY}} dK + rdK$$

$$= \left[\frac{a_{LaY} W^*}{a_{KY}} + r \right] dK > 0. \tag{5.29}$$

Taking total differentiation of Eq. (5.7) we get,

$$dZ = \frac{\partial Z}{\partial I} dI$$

$$= Z_I \left[\frac{a_{LaY} W^*}{a_{KY}} + r \right] dK = \left(\frac{Z_I}{a_{KY}} P_Y \right) dK > 0. \tag{5.30}$$

Taking total differentiation of Eq. (5.4) we get,

$$a_{LaX} dX + a_{LaY} dY + a_{LaZ} dZ = 0$$

$$\Rightarrow a_{LaX} dX = - \left(\frac{a_{LaY}}{a_{KY}} \right) dK - \left(\frac{a_{LaZ} Z_I}{a_{KY}} P_Y \right) dK$$

$$\Rightarrow \quad dX = - \left[\left(\frac{a_{LaY}}{a_{KY} a_{LaX}} \right) + \left(\frac{a_{LaZ} Z_I}{a_{KY} a_{LaX}} \right) P_Y \right] dK < 0. \tag{5.31}$$

Proposition 8 Expansion of domestic capital stock of the economy will adversely affect the export sector. However, the non-traded good producing sector will expand.

Thus, one of the informal sectors will expand and the other will contract. The intuitive explanation can be given as follows: Since supply of capital has increased and capital in our model is a specific factor of production, the capital-using sector must expand. Again, as income of the people engaged in the formal sector will rise due to expansion of the import-competing sector, demand for the non-traded good must go up implying a rise in output. Since the total labour supply is assumed to be constant, the export sector has to contract to release the additional supply of adult labour required for increase in output of Y and Z.

In our model, we have assumed that,

$$L_C{}^S = L_C{}^S(W, a_{LaX}X + a_{LaZ}Z) \text{ with } \frac{\partial L_C{}^S}{\partial W} < 0, \frac{\partial L_C{}^S}{\partial L_1} > 0 \tag{5.32}$$

$$\frac{dL_C^S}{dK} = \frac{\partial L_C^S}{\partial(a_{LaX}X + a_{LaZ}Z)}\left(a_{LaX}\frac{dX}{dK} + a_{LaZ}\frac{dZ}{dK}\right). \tag{5.33}$$

Now, $\left(a_{LaX}\frac{dX}{dK} + a_{LaZ}\frac{dZ}{dK}\right)$

$$\begin{aligned} &= -a_{LaX}\left[\frac{a_{LaY}}{a_{KY}a_{LaX}} + \frac{a_{LaZ}Z_I}{a_{KY}a_{LaX}}P_Y\right] + \frac{a_{LaZ}Z_IP_Y}{a_{KY}} \\ &= -\frac{a_{LaY}}{a_{KY}} - \frac{a_{LaZ}Z_IP_Y}{a_{KY}} + \frac{a_{LaZ}Z_IP_Y}{a_{KY}} \\ &= -\frac{a_{LaY}}{a_{KY}} \end{aligned} \tag{5.34}$$

$$\therefore \frac{dL_C^S}{dK} = \frac{-\partial L_C^S}{\partial(a_{LaX}X + a_{LaZ}Z)}\left[\frac{a_{LaY}}{a_{KY}}\right] < 0. \tag{5.35}$$

Proposition 9 Due to economic expansion by increase in domestic capital stock, child labour supply in the economy will fall.

The intuitive explanation is as follows: Due to the fall in output of X, adult labourers will be released from X production. Now some part of the released labour force will be employed in the formal sector and only some part in the non-traded final commodity-producing informal sector. Hence, the number of adult labourers engaged in the informal sector which is given by $(a_{LCX}X + a_{LCZ}Z)$ goes down as illustrated earlier. Now, in our model there are two determinants of child labour supply: one is adult wage rate, which is unchanged due to economic expansion. But, the number of people engaged in the informal sector, which is another determinant of child labour supply goes down. Hence, child labour supply will unambiguously decline. Thus, the labour reallocation effect in this case works positively on the supply of child labour.

6 Conclusion

In a labour–surplus, less developed economy, it is presumed that employers would prefer adults by paying them marginally higher wages than they would to children. This argument assumes that paying lower wages is the only reason why children are preferred to adults. The fact of the matter is that children, like women, are a docile labour force. They are unlikely to demand higher wages on efficiency ground (Swaminathan 1998), willing to work longer hours without question. Various studies have shown that children are preferred because they are easy to control. Children who are working outside the formal labour market have been defined as 'nowhere children'.

This chapter concludes that the child labour problem in a less developed economy can be reduced by adopting policies favourable to economic growth through capital accumulation or by adopting a restrictive trade policy in the sector where child labour is not used. One is directly altering the use of child labour, another is through changes in factor process. Thus, quantity effects and price effects are both important. We have not, however, enquired in this chapter about the effects of these two types of policies on national welfare. A ban on child labour is not necessary, and it creates the problem of compliance. Thus, we conclude that a country should not always follow every aspect of the WTO-prescribed policy package. The effectiveness of domestic policies in reducing the incidence of child labour will depend on the institutional and technological characteristics and the trade pattern of the relevant country.

Acknowledgement I am privileged to be able to contribute to this volume in honour of my teacher and mentor, Professor Biswajit Chatterje, who taught and guided me for my Ph.D. in the field of international trade.

References

Basu K, Van PH (1998) The economics of child labour. Am Econ Rev 88(3):412–427
Basu K (1999) Child labour: cause, consequence, and cure, with remarks on international labour standards. J Econ Lit 37(3):1083–1119
Gupta MR (2000) Wage determination of a child worker: a theoretical analysis. Rev Dev Econ 4(2):219–228
Gupta MR (2003) Child labour and economic development: a theoretical analysis. Arthaniti 2: 97–107
ILO (2006) Second Global Report on Child Labour
ILO (2010) Third Global Report on Child Labour
Maskus KE (1997) Should core labour standards be imposed through International Trade Policy? World Bank, Policy Research Working Paper, No.1817
Ranjan P (2001) Credit constraints and the phenomenon of child labour. J Dev Econ 64(1):81–102
Swaminathan M (1998) Economic growth and the persistence of child labour: evidence from an Indian city. World Dev 26(8):1513–1528

Chapter 6
Direction of Trade, Exchange Rate Regimes, and Financial Crises: The Indian Case

Rajendra Narayan Paramanik and Bandi Kamaiah

1 Introduction

In international trade research, the gravity model (of different variants) has become popular due to its empirical appeal and success in explaining bilateral trade flows. This model provides a cogent approach to test the role of possible influential variables affecting bilateral trade flows. Since the seminal work of James Anderson in the late 1970s, research in the area of bilateral trade flows using the gravity model gathered significant momentum. Much before Anderson (1979), Tinbergen (1962) and Linnemann (1966) proposed the idea of the gravity model in its simple format where bilateral trade flows are specified to be proportional to the product of the mass (gross domestic product, GDP) of the trading nations and inversely related to the cost (spatial distance). Later, the model is used in its augmented form by incorporating different sets of explanatory variables that affect the trade flows. Over the years, the gravity model has been a powerful tool to explain the bilateral trade flows, estimating trade potentials, identifying the impact of trade groups, explaining the trade pattern, and assessing the cost of border trade (Lin and Wang 2004; Liu and Jiang 2002; Sheng and Liao 2004).

In its extended form, the gravity model covered new explanatory variables. These variables may be classified into two groups, namely exogenous and dummy variables. Exogenous variables include per capita GDP, population, etc. The dummy variables incorporate factors like preferential trade agreement, integration, or participation in any group or organization Shi et al. (2005). For example, Aitken (1973) added a new variable to estimate the impact of the European Economic Community (EEC) on trade of its member nations. Frankel and Wei (1993) found that level of economic development, captured by per capita gross national product (GNP), plays a vital role in shaping a nation's trade flow.

R. N. Paramanik (✉) · B. Kamaiah
Department of Economics, University of Hyderabad,
Hyderabad, Andhra Pradesh 500046, India
e-mail: raju.purulia@gmail.com

B. Kamaiah
e-mail: kamaiahbandhi@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_6,

Traditionally, the gravity equation has been estimated for cross-section data using ordinary least square (OLS) method. The conventional cross-section approach without the inclusion of country-specific effects was found to be misspecified which has led to biased estimates (Matyas 1997). For instance, in a pure cross-sectional study, GDP of the home country becomes constant, and if the log-linear form is considered for the model, then log of the home country's GDP gets confounded with the intercept term. To overcome this limitation, panel data analysis has been suggested as a way out, provided that time series component is not small enough to lead to near multicollinearity with the intercept term. Egger and Pfaffermayr (2003) recommended the use of a three-way model with effects for importer, exporter, and time or explicit introduction of country-pair effects (to account for country-pair heterogeneity).

The trade gravity approach has been frequently used to analyze the trade pattern of different nations with their corresponding trading partners. For instance, in a recent study, Cieślik (2007) utilized the gravity approach to identify the trade effects of free trade agreements (FTAs) during the period 1992–2004. The estimated model, apart from standard variables, contained factors such as geographical proximity, language, common history, and FTA dummies. Cieślik found the impact of bilateral and regional trade agreements to be positive and statistically significant. The analysis was performed on a pooled panel of data with time effects and estimated with heteroscedasticity-adjusted OLS method. The impact of various agreements on trade differed. Cieślik found the regional agreements to be more trade-creating than bilateral agreements. At the same time, the trade effects associated with the European Union (EU) were found to be far greater than the effect of FTAs. Furthermore, the positive impact on trade seems to appear only several years after establishment of liberalized trade arrangement.

The gravity model has been used for empirical analysis in the Indian context too. Batra (2004) used cross-section data for the year 2000 and suggested that the magnitude of India's trade potential is highest with the Asia-Pacific region followed by Western Europe and North America. Countries like China, UK, Italy, and France reveal maximum potential for expansion of trade with India. Similarly, using panel data analysis Bhattacharya and Banerjee showed how factors like colonial heritage and size of the trading partners' economy play a vital role in determining India's direction of trade (DOT). In both the studies, apart from traditional variables (such as GDP and population), some common explanatory variables such as common language, colonial heritage, etc. are used.

So far, in the Indian studies, issues like exchange rate regimes and effect of financial crises on trade have not been addressed within the framework of the gravity model. In the recent past, the world experienced two global meltdowns, one in 1997–1998 (Asian crisis) and another in 2008 which affected the global economy. Though, in terms of its impact, the Asian crisis was confined within the vicinity of its genesis arena, the 2008 crisis has a wider and deeper impact all over the world; in both the cases, Indian trade relation with its trading partners got affected to a considerable extent. Apart from that, it is known that exchange rate regime of the trading partners also affects trade relations since volatility of exchange rate affects trade negatively which is contingent upon the exchange rate regime of the nation under consideration.

In the light of this background, it would be of interest to examine India's DOT with selected major trade partners, by addressing the issues, namely impact of financial crises and exchange rate regime changes. For this purpose, India's 25 major trade partners (in terms of total trade volume) are considered for analysis which constitutes 75–80 % of India's total trade since 1997–1998 to 2009–2010.

2 Theoretical Background

The gravity equation, as a tool for explaining bilateral trade patterns, was originally proposed by Tinbergen in the year 1962 (Tinbergen 1962). Despite its unquestionable success in empirical studies, it was often criticized for insufficient theoretical foundations. This drawback has been more than eliminated in the past 20 years with the rise of new trade theory with its rich microfoundation. It is worth to stress, that the gravity equation can be formally derived within an imperfectly competitive set-up with increasing returns to scale and firm-level product differentiation as well as within a perfect competition set-up with product differentiation at the national level. The gravity equation in the simplest form postulates that bilateral trade between two countries is directly proportional to economic size of the trading partners and inversely proportional to the distance between them, thus resembling the famous Newton's gravity law. The economic size of the partners is usually given by real income (Y). In mathematical notation, the simple gravity equation has the following structure:

$$\mathrm{TT_{ij}} = \left(A^{*} Y_{\mathrm{i}}^{\alpha *} Y_{\mathrm{j}}^{\beta}\right) / {D_{\mathrm{ij}}}^{\varphi},$$

where $\mathrm{TT_{ij}}$ denotes total trade flow between the nations, Y denotes the economic mass (GDP), D_{ij} denotes the distance between them, and i and j denote the two nations, respectively.

α, β, φ denote the parameters. A is some constant, known as gravity parameter.

Log linearizing the above equation yields the following form:

$$\ln \mathrm{TT_{ij}} = \ln A + \alpha \ln Y_{\mathrm{i}} + \beta \ln Y_{\mathrm{j}} + \varphi \ln D_{\mathrm{ij}}.$$

The basic gravity equation is frequently extended to incorporate other factors affecting (stimulating or hindering) bilateral trade flows. These could include, for instance, incomes per capita of trade partners. The gravity model implies that the larger, more prosperous and closer two countries are to each other, the more likely they are to trade. The model could be further augmented to incorporate cultural and linguistic proximity, historical links, and various barriers to trade. In the popular set-up, two different components of barriers to trade are often included which have a spatial and nonspatial dimension. Apart from the impact of distance, the spatial exogenous barriers severely affecting transport cost are, for instance, given by common border (adjacency) or landlockedness. The removal of nonspatial barriers (trade liberalization) is commonly proxied by dummies for regional or bilateral trade agreements.

3 Data and Methodology

The data used in this study are annual data covering India's major 25 trade partners which consistently constituted 75–80 % of India's total trade volume over the time period 1997–1998 to 2009–2010. There are altogether 338 observations, 13 time series components (years), and 26 cross-sectional entities (26 nations including India). The 25 nations along with their respective exchange rate regime are listed in the Appendix. Data on GDP is collected from Penn World Table and they are purchasing power parity (PPP) adjusted. To get the data, we actually multiplied the term per capita GDP (cGDP) of the nations with their corresponding population, both of which are available in the above mentioned data source. Trade volume data are collected from the website of the Ministry of Commerce and deflated by the corresponding year's consumer price index (CPI) of the USA, so as to get real value of the trade flow. CPI of the USA is obtained from the US Department of Labor. Great circle distance between the nations' capital is measured as a proxy for distance variable and data are obtained from Centre d'Etudes Prospectives et d'Informations Internationales (website: http://www.cepil.fr/anglaisgraph/bdd/distances.htm), under the file name "dist_cepii.dta" (in kilometres). The distance is calculated using great circle calculation. Here, two types of dummy variables are used, one to capture the effect of exchange rate regime of the trade partners and another is a time dummy to grasp the effect of the two major financial crises over India's trade volume. For the determination of exchange rate regimes of the trading partners, we have taken the help from different sources like working papers of International Monetary Fund (IMF; e.g., de facto exchange rate arrangements and anchors of monetary policy, 31 July 2006), World Bank publications, etc. For the time dummy, we assigned the values 1 for the post-crisis periods like 1998 (i.e., after the Asian crisis) and for the recent global meltdown, we also assigned the value 1 for 2008–2009 as well as 2009–2010 since the impact of this crisis is wider in its persistence unlike the previous one. Though India is believed to have escaped the crisis internally, the adverse effect of it is experienced globally which in turn affected the trade relation of India.

The estimated gravity model has the following log-linear form:

$$\begin{aligned}\ln \mathrm{trade}_{\mathrm{ijt}} = \alpha + \beta \ln \mathrm{gdp}_{\mathrm{it}} + \gamma \ln \mathrm{gdp}_{\mathrm{jt}} + \varphi \ln \mathrm{dis}_{\mathrm{ij}} + \Upsilon \ln \mathrm{pop}_{\mathrm{it}} + \Omega \ln \mathrm{pop}_{\mathrm{jt}} \\ + \phi D_{\mathrm{ik}} + \lambda T_t + e_{\mathrm{ijt}}. \end{aligned} \tag{6.1}$$

where α is constant term and common to all years and all nations, trade means the total bilateral trade between i th trade partners and j(India), $\mathrm{gdp}_{\mathrm{it}}$ signifies the gross domestic product of ith trade partner in the tth year, $\mathrm{gdp}_{\mathrm{jt}}$ means the gross domestic product of India in tth year, similarly pop signifies population of trading nation and India respectively, D denotes the dummy variable of exchange rate regime of ith nation, T is a time dummy which assumes the value 1 when year = 1998, 2008, 2009 and 0 otherwise, e_{ijt} is the error term. In Eq (6.1), 'ln' refers to logarithm of the variables and k = 1, 2.

Another version of the above equation has been empirically tested by replacing GDP with cGDP, i.e., per capita GDP of the respective nations so as to test whether the level of development of the countries has any significant impact over trade relation between nations or not.

$$\ln \text{trade}_{\text{ijt}} = \alpha + \beta \ln cg\text{dp}_{\text{it}} + \gamma \ln \text{cgdp}_{\text{jt}} + \varphi \ln \text{dis}_{\text{ij}} + \Upsilon \ln \text{pop}_{\text{it}} + \Omega \ln \text{pop}_{\text{jt}} + \phi D_{\text{ik}} + \lambda T_t + e_{\text{ijt}}. \quad (6.2)$$

Choosing an appropriate estimation technique is of prime importance in this context. Generally, in case of panel data with dummy variables, fixed-effect or random-effect models are used without further scrutinizing the problems of possible existence of nonspherical error which violates the basic assumptions of the models and leads to imprecise estimation. Though our model yields satisfactory results in case of fixed-effect over random-effect models which has been tested with the help of Hausman test, disturbances are tested to be heteroscedastic (each country has its own variance). In such circumstances, Winsten regression with panel-corrected standard errors (PCSE) is accepted as a useful technique to resolve the above mentioned problem. Papazogulou (2006) and Marques (2008) also suggested use of the PCSE in such circumstances and empirically validated their arguments. Apart from that, the balanced panel data sets fulfill the required desiderata of being moderately long in terms of time dimension (number of years = 13). PCSEs are similar to White's heteroscedasticity-consistent standard errors for cross-sectional estimators, but are better because they take advantage of the information provided by the panel structure of the data. Through Monte Carlo studies, Beck and Katz (1995) demonstrate that PCSEs produce more reliable standard errors than feasible generalized least squares Feasible Generalized Least Square (FGLS) methods.

4 Estimation Results

To start with, Eqs. (6.1) and (6.2) are estimated as fixed-effect and random-effect models. The Hausman test suggests superiority and applicability of the fixed-effect model over its counterpart since probability of chi square is obtained as 0.0006 which leads to rejection of the null hypothesis of no systematic difference in coefficients of the two models.

Once the fixed-effect model is tested to be applicable, we check the possible existence of cross-sectional heteroscedasticity since every trading partner has a differential impact over trade relation with India and there is a fair chance of possible heterogeneity among the cross-sectional units, i.e., countries. Modified Wald test to test group-wise heteroscedasticity in panel data context shows that there is a high degree of heteroscedasticity among the cross-sectional levels. So we test the two models (Eqs. 1 and 2) as PCSE model and the absence of contemporaneous correlation of the error terms is tested with the help of Pesaran's test which suggests poor contemporaneous correlation of cross-sectional error terms. This particular finding supports the applicability of the PCSE model.

The estimates of the first equation are shown in Table 6.1, from which we can see that the estimated coefficients and their respective influence over the trade volume are

Table 6.1 Estimated coefficients of the first model

	Coefficients	Standard error	Z value	$p > \mathrm{I\,Z\,I}$	Confidence interval	
$\mathrm{ldis}_{\mathrm{ij}}$	– 0.0554969	0.0364087	−1.52	0.127	– 0.0158629	0.1268567
$\mathrm{lpop}_{\mathrm{it}}$	0.3684939	0.0869616	4.24			0.1980522
$\mathrm{Lpop}_{\mathrm{jt}}$	0.289156	0.0577158	5.01	0.000	0.6423419	
$\mathrm{lGDP}_{\mathrm{it}}$	0.4763637	0.0946438	0.000	0.5389355	0.661822	
$\mathrm{lGDP}_{\mathrm{jt}}$	0.647689	0.107056	6.05	0.000	0.2789651	0.6901256
D2	0.415815	0.0961075	4.33	0.000	0.2274477	0.6041822
D3	0.2952705	0.1038122	2.84	0.049	– 0.0581976	0.3487386
T_t	– 0.0253897	0.1765531	−0.14	0.000	– 0.3714274	0.320648
Constant	1.604799	1.218702	−1.32	0.188	– 3.99341	0.7838122

Here, the suffix i represents the trading nation and j represents India

in accordance with their expectation of the traditional gravity model. On an average, over the years, GDP of the home country (India) has greater impact on trade volume than its trading partners and it indicates that economy size of India has significant impact on trade since the Z statistic is very high.

Similarly, population size of both the nations has a statistically significant impact on trade volume. A 1 % increase in population size of India causes a 0.28 % increase in trade of India on an average. But a similar change in population of the trading partner has a greater impact than that of India. Since the model is in log-linear form, we interpret the coefficients as measures of elasticity. A 1 % change in India's GDP in turn makes a positive change of around 0.65 % in trade volume, whereas a similar change of the trading nation's GDP has a little smaller influence since the coefficient value is 0.4763. A positive sign of both the coefficients signify similar impact on trade volume between India and its trade partners. A possible economic rationale behind this might be the increasing demand for Indian products abroad which is propelled by the increasing market size of the trading partners.

One apparently astonishing finding is the impact of the distance variable which is no longer influential since it is not statistically significant, though it has marginal negative impact on trade volume (coefficient value is – 0.055 but Z value is not high enough). Over time, technological advancement has greater impact on trade and exchange across the globe and it reduces the transaction or shipping costs that used to affect the trade relation between nations which are spatially distant. Banerjee and Bhattacharya (2006) also found that India trades less with their neighboring nations compared to the other nations of the world.

Exchange rate regimes of the trading partners also play a significant role in determining their trade relation with India. We have broadly classified the exchange rate regimes of the trading partners into three categories. The first one covers the nations with independently floating or managed-floating exchange rate. The second one covers those nations which come under the fixed exchange rate regime or exchange arrangement with no legal separate tender, and the third one comprises the nations with crawling peg, etc. We have considered two dummy variables to capture the impact of exchange rate regimes. Dummy variable D2 incorporates the nations with exchange rate arrangement with no separate legal tender and in this

Table 6.2 Estimated coefficients of the second model

	Coefficients	Standard error	Z value	$p > \mathrm{I\,Z\,I}$	Confidence interval	
$\ln cGDP_{it}$	0.376754	0.198645	1.897	0.062	– 0.0651837	0.3465595
$\ln cGDP_{jt}$	0.467834	0.232456	2.015	0.051	– 0.067165	0.3556378
D2	0.399682	0.095028	4.205	0.000	0.2174482	0.5941876
D3	0.2852406	0.110712	2.576	0.045	– 0.0571872	0.3456389
T_t	– 0.0276748	0.1765531	– 0.1567506	0.000	– 0.3815276	0.320648

Other variables like population and distance are not tabulated, since they show a similar kind of impact on trade volume as shown in Table 6.1

case, the nation adopting such regimes implies the complete surrender of the monetary authorities' control over domestic monetary policy (e.g., Germany and Italy). D2 also incorporates the nations with fixed exchange rate regime (e.g., China) where the center has significant control over monetary policy. Another dummy D3 is used for the nation which has either floating exchange rate regime, where exchange rate is solely determined by the marker forces (e.g., developed countries like USA, UK, and Japan and emerging economies like Taiwan, etc.), or managed-floating exchange rate regime like India, where monetary authority directly or indirectly intervenes the foreign exchange market according to the need of the economy.

That India's trade potential is there with all the nations of both the groups of different exchange rate regimes as both the dummy coefficients show low p value but its potential with fixed and exchange rate arrangement with no legal tender countries is more prominent. Coefficients of the dummy D2 is highly significant since the corresponding p value is very low. But coefficient of the dummy D3 is marginally significant and it has lesser incremental impact on trade volume. Nations with fixed peg or exchange rate arrangement with no legal tender have an average positive impact of about 0.42 %, whereas that of floating or managed-floating exchange rate nations is only around 0.29 %. This reveals that India's trade potential with those nations is stronger than their counterparts. India's recent (last one decade) trade relation with nations like China, Saudi Arabia, and Kuwait supports the above finding. Since in the recent past developed nations (those under floating exchange rate regime along with some emerging economies like Thailand) have gone through some economic crises, even though their trade relation (trade volume) with India in absolute terms has increased on an average, growth rate has not been that impressive compared to the major trading partners belonging to the other regime.

The time dummy "T" which is introduced to capture the impact of the two major crises on trade volume shows expected signs. Occurrence of those crises had an adverse impact on trade and the coefficient has been statistically highly significant. On an average, the crises impacted 0.025 % of trade volume over the period under consideration.

In the second model, the per capita income of the trading partners as well as India is not observed which shows that the so-called level of development (as measured by per capita GDP) does not have any impressionable impact on trade, whereas economic mass (GDP) has significant impact which is inferred from the result of the first model. It can be seen from Table 6.2 that coefficients of cGDP of India as well as other nations are statistically insignificant because of their low Z value. Other factors like distance and population of both nations have retained the influence on trade and exchange rate regimes as it is observed from the first model.

5 Conclusion

The present chapter investigates the extent of empirical success of the traditional gravity model in explaining India's DOT with its major 25 trading partners from 1997–1998 to 2009–2010. The study employs the widely used regression model (PCSE) developed by Paris and Winstern, so as to get unbiased and robust estimates of the gravitational trade model. The empirical findings suggest that apart from expected influence of the traditional variables like GDP, population, etc., other factors like exchange rate regime and financial crisis also play a vital role in shaping the bilateral trade relation of India with its major 25 trading nations in the last decade. Factors like geographical distance and cGDP have not emerged as influential variables to affect the trade relation of the nation to a considerable degree, whereas nations under fixed exchange rate regime as well as fixed peg or exchange rate arrangement with no legal tender are found to have more potential to trade than the nations under floating or managed-floating exchange rate. The two major crises, namely the Asian crisis (1997) and the recent (2008) US housing market crash, are also found to a have profound negative impact on Indian trade relations with its partners. The study may be extended further and carried out at a disaggregated level to discriminate between import and export.

Appendix

Table 6.3 Exchange rate regimes of countries

Country	Exchange rate regime	Monetary policy target	Country	Exchange rate regime	Monetary policy target
Australia	Independent or floating	Inflation targeting framework	Brazil	Independent or floating	Inflation targeting framework
Belgium	Exchange arrangement with no separate legal tender	Euro area policy framework	China	Fixed peg	Exchange rate anchor
France	Exchange arrangement with no separate legal tender	Euro area policy framework	Germany	Exchange arrangement with no separate legal tender	Euro area policy framework
Hong Kong	Floating	Other	Iran	Crawling peg	Monetary aggregate target
Indonesia	Floating	Monetary aggregate target	Italy	Exchange arrangement with no separate legal tender	Euro area policy framework
Japan	Floating	Other	South Korea	Floating	Inflation targeting framework
Malaysia	Managed floating with no predetermined path of exchange rate	Other	Kuwait	Fixed peg arrangement against US dollars (2001 onwards)	Exchange rate anchor
Russia	Managed floating with no predetermined path for the exchange rate	Other	Taiwan	Floating or independent	Other
Thailand	Managed floating with no predetermined path for the exchange rate	Inflation targeting framework	Singapore	Managed floating with no predetermined path for the exchange rate	Other
Switzerland	Floating or independent	Other	Saudi Arabia	Fixed peg arrangement	Exchange rate anchor
UK	Floating or independent	Inflation targeting framework	USA	Floating or independent	Other
UAE	Fixed peg arrangement	Exchange rate anchor			

"Other" signifies that corresponding countries have no explicitly stated nominal anchor, but rather monitor various indicators in conducting monetary policy

References

Aitken N (1973) The effect of the EEC and EFTA on European trade: a temporal cross-section analysis. Am Econ Rev 63:881–892

Anderson JE (1979) A theoretical foundation for the gravity equation. Am Econ Rev 69(1):106–116

Banerjee T, Bhattacharya R (2006) Does the gravity model explain India's direction of trade? A Panel Data Approach. IIM-A working paper

Batra A (2004) India's global trade potential: the gravity model approach, working paper no: 151. Indian Council for Research on International Economic Relations (ICRIER)

Beck N, Katz JN (1995) What to do (and not to do) with time-series cross-section data. American Polit Sci Rev 89:634–647

Cieślik A (2007) The impact of free trade agreements on Poland's foreign trade in 1992–2004. Bank i Kredyt 6:100–113

Egger P, Pfaffermayr M (2003) The proper panel econometric specification for the gravity equation: a three-way model with bilateral interaction effects. Empir Econ 28(3):571–580

Frankel J, Wei S-J (1993) Is there a currency bloc in the Pacific? In: Wignall AB (ed) The exchange rate, international trade and the balance of payment. Reserve Bank of Australia, Sydney

Lee J-W, Kwanho S (2004) Exchange rate regimes and economic linkages. Korean Int Econ Assoc 24(1):1–23

Lin L, Wang Y (2004) Empirical test of China's bilateral trade with the trade gravity model and its policy implications. World Econ Study 7:56–60

Linnemann H (1966) An econometric study of international trade flows. North-Holland Publishing Company, Amsterdam

Liu Q, Jiang S (2002) Study on the Chinese bilateral trade arrangement using the gravity model. Zhejiang Soc Sci 6:17–20

Marques H (2008) Asymmetries in heterogeneous integrated areas: evidence from intra-EU sectoral trade. Paper presented at 2008 EEFS conference in Prague

Matyas L (1997) Proper econometric specification of the gravity model. World Econ 20:363–368

Papazoglou Ch. et al (2006) A gravity model forecast of the potential trade effects of EU enlargement: lessons from 2004 and path-dependency in integration. The World Economy, 29:1077–1089

Sheng B, Liao M (2004) China's trade flows and export potential: the gravity model approach. World Econ Study 27:4–10

Shi C, Gu H, Qin X (2005) A survey of the theoretical basis of gravity model used in international trade. Nankai Econ Stud 2: 39–40

Tinbergen J (1962) Shaping the world economy: suggestions for an international economic policy. Twentieth Century Fund, New York

Chapter 7
Global Crisis, Financial Institutions and Reforms: An Indian Perspective

Dilip M. Nachane

1 Introduction

The relationship between finance and economic growth has witnessed a lively controversy and sharp divisions among schools of economic thought. Firstly, there are those who regard financial development as a critical precondition for economic growth (e.g. Goldsmith 1969; Hicks 1969; Schumpeter 1912; etc.). Secondly, an influential group of economists (e.g. Seers 1983; Lucas 1988) believes that the role of financial institutions is incidental to economic development and hugely 'over-stressed' in the conventional literature, while a third group sees finance as passively adapting to developments in the real sector (most notably Joan Robinson 1952).[1]

In recent years, there has been a marked shift in attitude towards financial development among economic growth theorists. The earlier scepticism has given way to a growing realization that financial markets and institutions play a defining role in the economic evolution of societies. Empirical evidence based on both cross-country as well as micro-level studies lends support to the view that financial development crucially affects the speed and pattern of economic development. This it does both by influencing the composition and pace of capital accumulation as well as by promoting technological innovation.

The financial system is traditionally viewed as performing the following five functions (see Levine 1997; Archer 2006; etc.):

1. Resource allocation
2. Mobilization of savings
3. Expanding goods and services markets

[1] Several well-known tracts on development economics frequently make no reference to the financial system at all (see Meier and Seers 1984; Stern 1989; etc.).

D. M. Nachane (✉)
Indira Gandhi Institute of Development Research (IGIDR),
Gen. A.K. Vaidya Marg, Goregaon East, Mumbai 400065, India
e-mail: nachane@hotmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_7, © Springer India 2014

4. Facilitation of risk pooling, hedging and diversification
5. Monitoring managers and exercising corporate control

   To this list, one must append an extra function, which has assumed a great deal of importance in recent years in emerging market economies (EMEs), viz,

6. Provision of credit to the informal sector (rural as well as urban) via microfinance institutions (MFIs).

However, even within the broad consensus recognizing the role of financial systems for economic development, important areas of disagreement persist, viz. the type of financial system most conducive to growth, private versus public ownership of financial institutions, the degree of regulation and supervision, the role of financial innovations and the pace and extent of financial liberalization. The Latin American crises of the 1980s and 1990s, the Asian financial crisis and the current global recession have once again brought the critical role of financial institutions under the scanner and introduced some important caveats to the consensus. The present chapter aims to take stock of some of these issues in the Indian context. While it is certainly not being claimed that the Indian experience is representative of the entire South Asian region, it is nevertheless felt that some of the lessons drawn here would have some relevance transcending their immediate context.

2 Indian Financial System

The financial system in India comprises the Reserve Bank of India (RBI) at the apex, numerous financial intermediaries, money market, debt market, foreign exchange market and equity market. Financial intermediaries include commercial banks, cooperative banks and non-bank financial institutions (NBFIs). Commercial banks constitute the largest segment of India's financial system, and a characteristic feature of this sector is the dominance of the public sector banks (PSBs) in terms of both branch offices and banking operations. Other types of banks include regional rural banks, local area banks and cooperative banks. Cooperative banking is also an integral component of India's banking system. It comprises two major segments, viz. urban cooperative banks (UCBs) and rural cooperative credit institutions (RCCIs). Of these, RCCI has a far more extensive branch network and a more diverse and complex structure than UCBs that maintain a single-tier structure. NBFIs are an important segment of India's financial system, embracing a heterogeneous group of diverse institutions including development finance institutions (DFIs), insurance companies, non-bank financial companies (NBFCs), primary dealers (PDs) and capital market intermediaries such as mutual funds. NBFIs offer a variety of products and services and play an important role in providing access to financial services to a vast section of the population. Recent years have also witnessed a phenomenal growth in the number of microfinance institutions (MFIs).[2]

[2] My forthcoming paper with Shahidul Islam of the Institute of South Asian Studies (ISAS) provides a comprehensive description of the financial system in the South Asian region.

Table 7.1 Size of the Indian financial system

Year	CBA (Rs. trillion)	MC (Rs. trillion)	B (Rs. trillion)	Total financial assets (Rs. trillion)	S
March 2000	8.1687	19.3327	5.2091	32.7104	1.69
March 2005	19.1143	32.8401	12.6841	64.6384	2.07
March 2008	34.3756	99.9614	22.3368	156.6737	3.33

The RBI plays an instrumental role in the Indian financial sector. Being the country's monetary authority, it formulates, implements and monitors India's monetary policy. As a prime regulator and supervisor of India's financial system, it uses and prescribes broad parameters of banking operations within which the country's banking and financial system functions. The RBI supervises, among others, commercial banks, cooperative banks, DFIs, and NBFCs. Through its monetary policy, it aims to secure stability in the internal and external value of the Indian currency and manages the foreign exchange market. It is also the banker to the government. It provides merchant banking services to both central and state governments. The RBI also does other traditional central banking activities such as currency issuance, promotional functions, etc.

Till the early 1990s, the Indian financial system was characterized inter alia by administered interest rates guided by the social concerns, high intermediation costs, a low base of capital, directed credit programmes for the priority sectors, high degree of non-performing assets (NPAs), low intensity of technologies, stringent entry barriers for new entrants and excessive regulations. Since the early 1990s, financial sector reforms have been initiated with the explicit objectives of developing a market-oriented, competitive, well-diversified and transparent financial system. In broad terms, the reforms have been focussed on the following six areas: (i) removing the restrictions on pricing of assets; (ii) building of institutional and technological infrastructure; (iii) strengthening the risk management practices; (iv) fine-tuning of the market microstructure; (v) changes in the legal framework to remove structural rigidities; and (vi) widening and deepening of the market with new participants and instruments. (For an extended review and critique of this process, kindly refer Nachane and Islam (2009)).

With a view to providing some perspectives on the evolution of the financial sector in India, we present in Table 7.1 a few basic indicators. Our first indicator is the size of the financial system defined as

$$S = \frac{[CBA + MC + B]}{GNP \text{ at market prices}}$$

where *CBA* means commercial banks' assets,

MC equity market capitalization and

B bonds outstanding

Financial systems differ according to the primacy of the sources of corporate finance, i.e. as to whether firms are financed mainly through capital markets (as in the USA) or through bank loans (as in Germany). To get an idea about the relative composition of alternative sources of funding, we present two alternative indicators, viz.

Table 7.2 Composition of the Indian financial system

Date	India	USA	UK	Germany	Japan	Korea	China
March 2000	0.4225	0.377244	0.722975	7.128747	0.025747	2.97648	3.19797
March 2005	0.5820	0.487954	0.946397	8.161254	0.017327	1.15719	5.248572
March 2008	0.3439	0.949475	2.298705	10.23022	0.026746	2.193102	2.503276

Table 7.3 Composition of the Indian financial system

Date	India	USA	UK	Germany	Japan	Korea	China
March 2000	0.3328	0.197723	0.541355	2.993587	0.008826	0.875696	2.492968
March 2005	0.4199	0.22998	0.673839	2.866789	0.006009	0.558997	2.800088
March 2008	0.2811	0.308487	1.215864	2.792327	0.006415	0.688694	1.492713

$$F_1 = \frac{CBA}{MC} \text{ and } F_2 = \frac{CBA}{(MC + B)}.$$

Tables 7.2 and 7.3 present these indicators for India over the last decade. For comparison, the indicators are also presented for a few other countries.

Another important characteristic of any financial system is the ownership pattern of bank assets, i.e. the percentage of bank assets that are state owned. For this, we define a ratio

$$0_1 = \frac{\text{Assets of Public Sector Banks}}{\text{Total Bank Assets}}.$$

It is also of interest to define another indicator of ownership, viz. the percentage of total bank assets that are owned by foreign banks. This is given by the ratio

$$0_2 = \frac{\text{Assets of Foreign Banks}}{\text{Total Bank Assets}}.$$

One final consideration in assessing any country's financial system is the extent of its concentration. For this, two indices are usually employed, viz. the share of the top five commercial banks (private or public) in total bank assets, C_B, and the share of the top five listed companies in total market equity capitalization, C_E. The ownership and concentration statistics are presented in Table 7.4 (the last column representing total market equity capitalization, C_E, represents the share of the top five listed companies on the two prominent stock exchanges of India, viz. the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE)).

3 Global Crisis: Triggering and Aggravating Factors

The global financial crisis which took the USA and much of the rest of the world by storm in late 2007 has led to a wide-ranging reassessment of the entire gamut of issues bearing on the financial systems of advanced capitalist economies. Detailed post-mortems of the crisis (see, e.g. Brunnermeier 2009) distribute the blame in more or less equal measure among the following four causal factors:

Table 7.4 Ownership and concentration pattern of Indian financial system

Date	O_1	O_2	C_B	C_E
March 2000	0.8023	0.0746	0.3969	0.3076
March 2005	0.5122	0.0443	0.2768	0.1463
March 2008	0.6990	0.0840	0.3884	0.2815

1. ***Great Moderation***—A generic term used to describe the global situation from the 1990s onwards characterized by (i) low inflation and correspondingly low short-term interest rates, (ii) steady growth rates in USA and the Eurozone, (iii) high growth in EMEs (China and India in particular) and (iv) rising housing prices across the globe.
2. ***Global Savings Glut***—This refers to several simultaneous (and mutually correlated) phenomena, viz.: (i) rise in savings rates in China, Japan, Organization of the Petroleum Exporting Countries (OPEC) and East Asia; (ii) a matching decline in savings and huge current account (a/c) deficits in the USA and the Eurozone and (iii) flow of global investment into US Treasury securities leading to (iv) low real long-term interest rates in the USA and a correspondingly high demand for credit.
3. ***Lax Monetary Policy***—It is now generally agreed that the Greenspan years were characterized by an unwarranted easing of monetary policy in response to the Dotcom bubble of 2000–2001. Firstly, there was the aggressive cutting of short-term interest rates in 2000 and 2001, and this was followed by a regime of loose monetary policy[3] over the entire period 2001–2004 leading to the home price bubble.
4. ***Home Price Bubble***—The unprecedented house price boom over 1996–2005 in the USA and worldwide was sustained partly by the lax monetary policy of the US Fed (and other Organisation for Economic Co-operation and Development (OECD) central banks) over 2001–2004 and partly by a belief that housing prices would continue their one-way movement riding on the global growth story. The phenomenal growth of sub-prime mortgages (covering borrowers with poor credit history and scores) especially over the period 2002–2006 (when they rose from US$ 160 to 600 billion) fed this bubble almost till it burst in September 2007.

From a historical perspective, the above-mentioned pattern might look very similar to several past recessionary episodes but there were several new factors which sharply enhanced the magnitude of the crisis fallout (see Ashcraft and Schuermann 2008; Gorton 2008; Claessens 2009; etc.).

The first of these factors was *increased opaqueness* of the financial system brought about by complex financial instruments such as mortgage-based securities (MBSs), collateral debt obligations (CDOs) and so on. The process of securitization spread particularly rapidly during the period 2000–2007.[4] The second factor has been

[3] The policy was *loose* also in a more formal way—policy rates were consistently less than those predicted by various Taylor-type rules.

[4] For example, securitization of non-conforming mortgages (i.e. Alt-A and sub-prime) increased from 35 % in 2000 to 70 % in 2007.

financial integration and interlocking brought about by capital account openness, international financial harmonization and the increasing presence of foreign intermediaries in several banking systems. This factor accentuated the transmission of financial shocks across borders. The final factor was the *high degree of leveraging* of financial institutions either directly through the commercial banking system (as in Europe) or through the *shadow* banking system[5] (as in the USA). The combination of these factors resulted in a crisis of unmitigated proportions, necessitating huge bailout and stimulus packages and resulting in huge welfare losses.

4 Global Crisis: Reassessment of Financial Systems

Given the scale, extent and severity of the current crisis, it is perhaps inevitable that several established theories would be challenged and several entrenched notions reviewed. The intellectual underpinnings of the advocacy of financial sector reforms in least developed countries (LDCs) and EMEs by multilateral institutions and the developed world generally are rooted in two interrelated economic theories, viz.: (i) the *efficient markets doctrine* and (ii) the MacKinnon–Shaw thesis.

A. Efficient Markets Hypothesis (EMH) As is well known, this hypothesis is very much in the Chicago School tradition (or what is now being increasingly termed as the *freshwater* view). It posits that current market prices of financial assets embody rationally all the known information about prospective returns from the asset. Future uncertainty is of the 'white noise' kind and 'noise traders' (speculators) may succeed in pushing the markets temporarily away from equilibrium; but with the market clearing continuously, 'rational traders' will bring the system back to equilibrium, by taking countervailing positions and imposing heavy losses on those speculators who bet against the fundamentals. Equilibrium asset prices will therefore be altered only when there are 'shocks' to the fundamentals, and while supply shocks are inevitable, the severity of demand shocks can be tempered by policy aimed at giving more access to information about fundamentals to market participants, and avoiding 'policy surprises' or attempts to control asset prices. This approach has been the basis of virtually all past responses to financial crises—a response which was fundamentally skewed in that international financial institutions who usually precipitated such crises by their indiscriminate lending, rolling over of credit and

[5] Shadow banking institutions are typically intermediaries between investors and borrowers. For example, an *institutional investor* like a *pension fund* may be willing to lend money, while a corporation may be searching for funds to borrow. The shadow banking institution will channel funds from the investor(s) to the corporation, profiting either from fees or from the difference in interest rates between what it pays the investor(s) and what it receives from the borrower. By definition, shadow institutions do not accept deposits like a depository bank and, therefore, are not subject to the same regulations. Familiar examples of shadow institutions included *Bear Stearns* and *Lehman Brothers*. Other complex legal entities comprising the system include *hedge funds*, *Special Investment Vehicles (SIVs)*, *conduits*, *money funds*, *monolines*, *investment banks* and other NBFIs.

tax avoidance strategies were seen in the role of victims, whereas the major blame was apportioned to the crisis-affected countries for their bungled macroeconomic management (current account deficits, overvalued exchange rates, loose monetary policy, etc.) and for 'misleading' investors by withholding key information about fundamentals. Such governments were then administered 'bailout' packages with strong attached conditionalities as part of the International Monetary Fund (IMF)'s 'tough love' treatment.

The inappropriateness of the EMH as a description of actual trading strategies of forex traders has always been strongly suspected. However, they are in systematic violation of rational market behaviour. Behavioural theories of human decision-making (see Kahneman and Tversky 1984; Rabin and Thaler 2001; etc.) argue that in the face of complex uncertain situations, individuals do not proceed via maximizing expected utility but using *cognitive heuristics*. Such heuristics is an aid to reducing a complex task to a manageable proportion but often introduces systematic biases. The bulk of the econometric evidence on financial markets is also contra the EMH (see, e.g. Shiller 1981; LeRoy and Porter 1981; Shleifer and Summers 1990; etc.).[6]

In the wake of the current crisis, economists are increasingly turning to the so-called *saltwater* view, which is essentially a resurrection of the 1930s Keynesian description of financial markets as being 'casinos' guided by 'herd instincts' (see the public utterances of highly regarded economists such as Buiter 2009, De Long 2009, Krugman 2009, etc). In the Keynesian view, investors in financial assets are not interested in a long-term perspective, but rather in speculating on short-run price behaviour. Far from basing their expectations on prospective behaviour of the underlying fundamentals, such investors are more likely to base their opinions on market sentiments (i.e. the opinion of the other members of their group). This lends a dangerous edge of volatility to financial markets as any 'news' if it affects market sentiment strongly (in either direction) is likely to produce mood swings in market sentiment, even if the 'news' in question is unlikely to alter long-term fundamentals.

B. MacKinnon–Shaw Thesis This thesis essentially views financial liberalization as an integral component of overall liberalization, in the twin beliefs that (i) liberalization in the real sector could not proceed satisfactorily in the absence of financial liberalization and (ii) financial liberalization was an 'enabling condition' of faster economic growth, as it increased competition, transfer of know-how and transparency.

The McKinnon–Shaw case for financial liberalization was based on the perception that *financial repression* (arising as a consequence of government control over important parameters of bank behaviour) tended to result in low (and often negative) real interest rates and an excess demand for credit. The resultant credit rationing led to credit allocation to favoured sectors by administrative fiat rather than through the purview of a market mechanism. Following financial liberalization, real interest rates would rise to their natural levels and economic growth would result from an

[6] Long before the current crisis, Warren Buffet once famously remarked, '*I'd be a bum in the street with a tin cup if the markets were efficient.*'

increased quantum of domestic savings and a rise in total factor productivity (TFP) due to an improvement in the quality of bank credit for investment purposes. As we have already noted in Sect. 1, by and large financial liberalization does seem to promote growth in the long run. But the short- and medium-term consequences may not always be benign. Several authors have underscored the likely harmful effects of financial liberalization in triggering financial crises and in misdirecting the allocation of capital (see Saidane 2002; Eichengreen 2001; etc.). Demirguc-Kunt and Detragiache (1998), for example, argue that financial liberalization intensifies competition among banks, who in their eagerness to preserve market shares could indulge in indiscriminate and risky credit operations (moral hazard problem). During bullish periods, debt leveraging can augment the expected return from financial position-taking by corporate borrowers. Wider asset price movements also erode the ability of banks and other financial institutions to adequately collateralize their loans, while competition restrains them from raising the risk premia on loans.

Recent studies such as those by Rodrik et al. (2002), Alcala and Ciccone (2004) and Kaufmann et al. (2007) clearly indicate the importance of institutional features such as corruption, rule of law and general governance issues (such as political accountability, quality of bureaucracy, etc.) in determining whether the outcomes of financial liberalization would be beneficial or otherwise. This could be an important part of the explanation as to why liberalization usually succeeds in the developed countries but often fails in the developing world. *It also throws up in retrospect the fallacy implicit in the reform advocacy of the 1990s which urged developing countries with weak institutions to undertake economic reforms, under the implicit assurance that political progress and good governance would follow as a consequence.*

5 Global Crisis: Coordinated Policy Response

One distinctive feature of the current crisis has been that for possibly the first time since the abandonment of the gold standard in 1926, there has been an almost instinctive recognition by policymakers globally of the need for a coordinated approach to the crisis.[7] In the early stages of the crisis, the coordination efforts were confined mainly to the G5 countries (France, UK, USA, Germany and Japan) and covered three major areas, viz. liquidity provision by central banks, fiscal stimulus and the cleansing of bank balance sheets.

---

[7] This is not to deny that in the immediate wake of the crisis, individual country policies tended to be insular, with countries acting in an uncoordinated manner to expand lender of last resort facilities, increase protection of creditors and depositors and recapitalize banks with public funds. This lack of coordination had some destabilizing effects, at least in the short term. Two cases, in particular, stand out, viz. the Lehman bankruptcy and the collapse of the Icelandic banking system. When Lehman fell, countries moved immediately to ring-fence assets in their own jurisdictions. The case of Iceland was similar. Although Icelandic banks had a large number of non-resident depositors, the authorities failed to coordinate with the countries in question. Some of these countries ended up seizing Icelandic bank assets to protect their own depositors.

Thus, to begin with, there was an unprecedented coordinated cut in policy rates by six major central banks in October 2008—by 50 basis points. Second, on the liquidity provision front, the Fed authorized temporary foreign exchange swap lines with 14 different monetary authorities.[8] This unique arrangement was designed to alleviate the global shortage of dollar funding.

On the fiscal stimulus, let me quote from a recent speech[9] by a former managing director of the IMF:

'Fiscal stimulus is less effective in more open economies, as some of the spending feeds through to imports, benefiting output and employment in other countries. This is why collective action is so important, why countries must act in unison. If more countries act, the burden on each individual country is lessened.

It happened. Countries acted in a coordinated manner. Moving together, they delivered a global fiscal stimulus of 2 % of GDP in 2009, exactly what we asked for a year ago. Although, the coordination was not explicit, policymakers all did the same thing at the same time for the same reason. This was unprecedented, even if countries did not always receive due credit for this achievement. We are already seeing the payoff—IMF analysis suggests that the fiscal expansion boosted growth by between 1 and 3 percentage points this year, and up to a third of the gain comes explicitly from coordination.'

However, as the crisis unfolded and its persistent and pervasive nature became clear, it was recognized that the coverage of the coordination process had to be considerably widened to include the G20 group and the scope extended to include not only monetary and fiscal policy but also financial sector regulation and supervision. The main partners in such a coordinated approach would be

1. National regulatory and supervisory authorities
2. IMF
3. Financial Stability Board (FSB) and other international standard-setting bodies—Basel Committee On Banking Supervision (BCBS), International Organization Of Securities Commissions (IOSCO), etc
4. Influential groups like G20

6 Role of National Regulatory and Supervisory Authorities

The role of national regulatory and supervisory authorities was debated extensively first in the de Larosiere Group (February 2009) in the EU and then in the Working Group 1 of the G20 (March 2009). The deliberations threw considerable light on the existing deficiencies in the global financial system and suggested several measures to

[8] Today, the central banks in the UK, the euro area, Switzerland and Japan all have access to unlimited swap lines across different maturities.

[9] See the speech 'Crisis Management and Policy Coordination: Do We Need a New Global Framework?' by Dominique Strauss-Kahn made at Oesterreichische Nationalbank, Vienna, on May 15, 2009.

mitigate the possibility of recurrences of a crisis of such an amplitude. The suggested measures embraced the following five distinct areas:

1. Expanding the scope of regulation
2. Issues related to the leverage of financial institutions
3. Pro-cyclicality of capital requirements
4. Reducing costs of financial failures
5. Devising market incentives for prudential behaviour

A. Scope of Regulation It was felt that regulation not only has to be strengthened but its scope has to be extended considerably as well. For strengthening regulation, two measures seem to be in order, viz.: (i) entrusting a special regulatory authority (either an existing one or a newly constituted one) with an explicit financial stability mandate and (ii) ensuring coordination between different regulatory authorities. The scope of regulation needs to be extended to include credit-rating agencies and private pools of capital (including hedge funds) via a system of registration, disclosure requirements and oversight.

B. Leverage of Financial Institutions An important amplification factor for the current crisis has been not only the high degree of leveraging of many financial institutions but also that this leveraging has very often been quite opaque. Simpler leverage measures are necessary to supplement the Basel Tier I and II requirements (such as a minimum unweighted leverage ratio for bank capital). Considering the phenomenal growths of off-balance sheet activities (e.g. contingency banking and financial engineering products, securitization of bank assets, loan sales, etc.), there is a need for more accurate measures of balance sheet exposures. As part of the latter, one should expect a stronger focus by regulators on loan-to-value ratios (especially for mortgages) and higher loan-loss provisioning norms. Stress-testing exercises should be conducted periodically to monitor leveraging on an ongoing basis, accompanied by improved disclosure requirements for complex structured products.

C. Pro-cyclicality of Capital Requirements A fact well known to economists (see, e.g. Ghosh and Nachane 2003) but consistently ignored by policymakers in the pre-crisis era is the fact of capital adequacy requirements being pro-cyclical and, hence, a possible accentuating factor in any crisis. As the current crisis runs its course, there is a greater realization among central bankers globally that ways have to be found for countering this pro-cyclicality. The first step in such an endeavour would be to encourage the build-up of capital buffers during economic expansions. These could then be unwound in times of recession to forestall the adverse impacts of *fair valuation*, leverage and maturity mismatches. There is also the necessity of imposing higher capital requirements on *systemically important* financial institutions. Borel III proposes to take most of these considerations on board.

D. Reducing Cost of Financial Failures The welfare costs of financial crises are generally severe and fall disproportionately on disadvantaged groups in any society, and the current crisis is hardly an exception. An early warning diagnostic system can contribute considerably towards containing collateral damage. In this regard, the Prompt Corrective Action (PCA) scheme introduced by the RBI in December 2002 is noteworthy.[10] Another important measure relates to the prevalence of orderly closure rules for important financial institutions (as prevalent in the USA for banks under the Federal Deposit Insurance Corporation (FDIC) Improvement Act and Competitive Equality Banking Act). Under exceptionally turbulent circumstances, the use of credit ratings by private agencies could be temporarily suspended in favour of regulators' ratings. A final and critical step could be to establish clearing houses in OTC derivatives markets and make such central counterparties subject to transparent and effective oversight.

E. Devising Market Incentives for Prudent Behaviour Market incentives can play an important supplementary role in ensuring prudent behavior by financial institutions. It is generally recognized that an important triggering factor in the current crisis has been the unregulated corporate compensation framework, which provided perverse incentives for excessive risk taking resulting in a serious moral hazard syndrome. The solution to this proposed by Working Group 1 of the G20 is prudential oversight of financial executive compensation schemes. If this is found difficult to implement, a softer option seems to be to put in place a deferred compensation plan to replace the existing practice of paying bonus up-front to top management. To reduce the opaqueness associated with securitization, originators of securitized products may be required to take an equity slice in the products that they sell/distribute. Credit-rating agencies face important conflicts of interest between their ratings and consultancy activities and, hence, there is a need for better separation of these two functions. A very interesting suggestion for market induced discipline in financial institutions is the so-called Chicago Fed Plan (see Keehn 1989), which argues for the inclusion of a mandatory subordinated debt component in bank capital (detailed discussion of this Plan and several variants may be found in Calomiris and Powell 2000, Evanoff and Wall 2000, etc.).[11]

7 Revised Role of the IMF

In the aftermath of every crisis, there is always in evidence a general dissatisfaction of the IMF role as the sole global resolution mechanism. This dissatisfaction is particularly acute among LDCs and EMEs, but is not confined to them alone. Even

---

[10] Under the PCA, the RBI will initiate certain *structured* as well as *discretionary* actions in respect of banks, which have hit certain trigger points in terms of capital adequacy ratio (CAR), net NPAs and return on assets (ROA). Thus, if a bank's CAR falls to less than 9 %, but is equal to or more than 6 %, then the RBI will initiate the following structured actions: (1) submission and implementation of capital restoration plan by the bank, (2) bank will restrict expansion of its risk-weighted assets, (3) bank will not enter into new lines of business, (4) bank will not access/renew costly deposits and Certificates of Deposits (CDs) and (5) bank will reduce/skip dividend payments. In addition, it could also initiate any of the following discretionary actions: (1) RBI will order recapitalisation, (2) bank will not increase its stake in subsidiaries and (3) bank will reduce its exposure to sensitive sectors like capital market, real estate or investment in non-SLR securities.

[11] In India, there is no mandatory requirement for subordinate debt, but there is a ceiling ($<$ 50 % of Tier I capital). Such a debt is part of Tier II capital.

the advanced group of countries would like to see some reforms getting under way and, of course, there are several reform areas which exercise both groups in equal measure. By and large, the various suggestions for IMF reforms may be grouped under the following headings:

1. Replacement of ex-post (crisis) negotiations on IMF loan conditions with ex-ante rules for IMF membership This would imply that only those countries which agree to certain conditions ab initio would qualify for IMF assistance in the event of a crisis. The rationale for this measure would be that it would induce countries to adopt preventive measures well in advance of crisis situations, thus mitigating their severity or even forestalling their occurrence. Such a solution, if implemented would strictly isolate economies which, for some reason, refused to be signatories to this arrangement and would, in effect, be tantamount to ostracizing certain economies from the global financial community. The ex-ante conditions suggested include:

i. Subordinated debt requirements
ii. Minimum reserve ratios for banks and other financial institutions
iii. (Risk-based) deposit insurance (for banks and other financial institutions)
iv. Free entry to foreign banks
v. No regulatory bias in favour of *off-balance sheet* activities
vi. Regulation and supervision by national regulator to cover all *systemically important* financial institutions
vii. No scope for *regulatory arbitrage*, which arises when some components of the financial system are regulated with a 'lighter touch' relative to others (as happens, for example, with NBFCs in India)

2. A vigorous enforcement of IMF guidelines on Exchange Rate Surveillance This will be with special reference to emerging global imbalances, protracted currency undervaluation, currency mismatches, etc.

3. Lender of 'last resort' function of the IMF This is essentially a revival of Walter Bagehot's nineteenth-century proposal of what a lender of last resort should ideally do in the event of a banking crisis. It is proposed that the IMF should lend freely during crises on good collateral at a penalty rate and for short periods (say, less than 90 days) with limited rollover possibilities. As much as 25 % of the collateral could be in the form of foreign government securities, and the penal rate could be 2 % above the value-weighted yield on the bundle of securities offered as collateral (see Calomiris 2007; Goldstein 2008; Brunnermeier and Pedersen 2009; etc.).

4. More adequate representation to the point of view of LDCs and EMEs There has been a long-standing and simmering discontent among the LDCs and EMEs that the IMF does not provide adequate representation to their point of view. Their main demands are threefold:

1. Radical changes in access, pricing and conditionality for IMF borrowers, in particular the introduction of *flexible credit lines*
2. Raising quotas/votes of EMEs and LDCs as a group[12]
3. Negating the US veto on crucial IMF decisions

It is to be noted that the Committee on IMF Governance Reform (under the chairmanship of Trevor Manuel) which submitted its Report in March 2009 makes an honest effort to address several of these concerns, though whether they will be finally incorporated in the IMF Charter is as yet unclear. Among the major recommendations of the Report are the following:

a. An accelerated quota revision process to be concluded by April 2010
b. Expansion of the Fund surveillance mandate beyond exchange rates to macro-prudential issues, financial spillovers and capital account policies
c. Lowering of the voting threshold on critical decisions from 85 % to 70–75 %. This would in effect annul the US veto (as the USA has 17 % voting power)
d. Extending double majority to a wider range of decisions
e. Greater transparency and role for merit in the appointment of the managing director and deputy managing directors

8 Role of the FSB and Other International Standard-Setting Bodies

The Financial Stability Forum (FSF) was established in 1999, by the finance ministers and central bank governors of the G7 to promote international financial stability through enhanced information exchange and international cooperation in financial market surveillance. The FSF being heavily dominated by G7 representation excluded the concerns of much of the rest of the world including key emerging Asian economies such as China, India, Indonesia, Korea, Malaysia and Thailand. In April 2009, the FSF gave way to the FSB with an extended membership of countries (G20 + Spain + European Commission) and an expanded mandate. Several roles are envisaged for the FSB, of which the following four seem to be of particular importance:

i. FSB should alert standard-setting bodies about loopholes in existing regulatory structures. The bodies like BCBS, IOSCO, etc. can then devise specific operational guidelines for incorporation into national regulatory and surveillance frameworks.
ii. FSB can monitor and advise on market developments and their implications for regulatory policy.

[12] A proposed tripling of *basic* votes (number of votes every country has *qua* member) would increase developing country votes from 32.3 to 34.4 % (the corresponding World Bank figure is 42.6 % proposed to be raised to 43.8 %). Birdsall (2009) makes a particularly relevant suggestion in this context, viz. *double majority voting* on selected issues—a majority of weighted votes (as currently) plus a majority of countries. The system prevails at the Inter-American Development Bank, ADB and African Development Bank for election of a new president/head.

iii. FSB can develop an early warning system on emerging systemic risks when the situation so warrants (Brunnermeier et al. 2009). This could be done in collaboration with the IMF.
iv. FSB to engage contingency plans for cross-border crisis management particularly with respect to systemically important multinational firms.

8.1 Role of the G20

In recent years, the G20 has emerged as an influential group for directing the thrust of globally coordinated policy, though it must be mentioned that it is purely a consultative and advisory forum with no operational status. In spite of this last limitation, the G20 did contribute substantially to toning down some of the more serious consequences of the current crisis. Among its major achievements since the inception of the current crisis have been (i) succeeding in securing a substantial increase in IMF resources (US$ 750 billion + US$ 250 billion special drawing rights (SDR) allocation) as also of the Multilateral Development Banks (MDBs) (US$ 100 billion); (ii) ensuring a greater degree of flexibility in IMF support programmes (*flexible credit lines*); (iii) strengthening financial supervision and regulation (regulatory oversight of credit-rating agencies, action against non-cooperative jurisdictions and tax havens, improving accounting standards, establishment of a new FSB, etc.); (iv) supporting growth in EMEs and LDCs by helping to finance counter-cyclical spending, bank recapitalization, infrastructure, etc.; (v) countering rising protectionism and finally (vi) reaffirmation of millennium development goals.

An interesting proposal, which may at some stage be usefully espoused by the G20, pertains to the group insurance scheme (for G20 members) proposed by E. Prasad (May 2009). The broad contours of the scheme are as follows: (i) Participants are to be offered a short-term credit line in the event of a crisis. (ii) Entry fee is between US$ 10 and 20 billion. (iii) Premium is to depend on the level of insurance desired and a suggested value is 1 % of the face value of policy. (iv) Countries following policies that enhance global risk (such as large budget or current a/c deficits) would face higher premia. (v) Premia are to be invested in the USA, euro area and Japanese government bonds. In return, the central banks of these countries would top up the lines of credit in the event of a global crisis. (vi) The scheme is to be administered by the FSB rather than the IMF since the voting in the former is not based on country quota subscriptions (as is the case with the latter).[13]

[13] The G20 Insurance Solution has several points of similarity with the Chiang Mai Initiative of the ASEAN + 3.

9 Official Thinking on Future Course of Financial Sector Reforms in India: A Critique

Indian policymakers in the highest circles have been reaffirming their commitments to financial sector reforms, particularly so after the United Progressive Alliance (UPA) government's return to power in July 2009. These statements are too vague to offer any direct clue about the actual course of reforms over the next few years; but one has reason to suppose that the future road map of reforms will closely follow the recommendations outlined in the reports of two recent committees—*The Committee on Making Mumbai an IFC (International Finance Centre)* under the chairmanship of Percy Mistry and *The Committee on Financial Sector Reforms (CFSR for short)* under the chairmanship of Raghuram Rajan. The latter report, in particular, is a detailed examination of the Indian financial sector and makes a number of wide-ranging recommendations.

Let me begin by mentioning that there are several issues taken up by the CFSR with which I am in broad agreement, such as those related to broadening access to finance (Proposals 3 and 4), level playing field (Proposal 8), developing credit infrastructure (Proposals 29 and 30) and improvement of land registration and titling (Proposals 31 and 32).

However, I have serious reservations on some of the substantive issues that have been raised in the CFSR. I will group my comments under four major headings:

1. The general philosophy about financial markets espoused by the Report
2. The macroeconomic framework
3. Principles versus Rules-based regulation and
4. Regulatory and supervisory independence (RSI)

On the first of these, it needs to be emphasized that the entire CFSR approach is strongly grounded in the *new classical* (or *freshwater)* view of financial markets, about which several reservations have been noted in Sect. 4, particularly in the wake of the current crisis.[14] Hence, these need not be repeated here. Let me, however, briefly touch upon the remaining issues.

1. Macroeconomic Framework The two aspects of the macroeconomic framework suggested by the CFSR, which have attracted the maximum attention, are *inflation targeting* (IT) and *capital account convertibility* (CAC).

The main policy recommendation of the CFSR as regards a suitable monetary policy regime for India pertains to the announcement of a 'low inflation objective—a number, a number that can be brought down over time, or a range—over a medium term horizon (say 2 years) as the primary goal of monetary policy'. In the execution of this objective, the RBI would be granted full operational independence and simultaneously held fully accountable. This recommendation is in tune with current mainstream academic thinking, so in a way the CFSR is only reiterating the orthodox position. To many including the top brass in the US Fed and the ECB (including such

[14] My own reservations about this approach have predated the crisis (see Nachane 2007).

notables as Alan Greenspan, Otmar Issing, Donald Kohn, etc.), IT appears an idea whose time is yet to come, but even those who regard it as a desirable long-term goal admit that the devil lies in the details. Where the CFSR falls short of expectations is in its failure to convince the reader of the superiority of IT to the existing (discretionary) monetary policy regime in India and the lack of attention to the specific difficulties that would need to be overcome for operationalizing such a procedure (in the Indian context).

IT central banks typically assume that financial stability is a by-product of price stability. The recent sub-prime crisis in the USA, following a long period of steeply appreciating equity and real estate prices, occurring in a period of sustained price stability, should serve as a telling refutation of this position. However, this case is by no means unique. The literature (e.g. Bordo et al. 2000; Borio and Lowe 2002) furnishes several such instances—Japan (1985–1989), Korea (1990–1997), USA (1925–1929), etc. Even the Indian situation of 2005–2007 appears similar. The fact of this possible disconnect between asset prices and general inflation could mean that typically an IT central bank may allow credit to expand and feed an asset price boom for too long. Ultimately, when the asset price boom feeds into general inflation (via the wealth effect) the central bank would be forced to apply the brakes abruptly, which could result in a prolonged asset price deflation (to wit Japan's 'lost decade' of the 1990s) and a general recession. Thus, an IT central bank will always intervene too late to prevent a crisis.

There also seems to be an implicit supposition in the CFSR that adoption of IT guards against balance of payment crisis. This need not necessarily be so. IT does not insulate a country against balance of payment crises (see Calvo and Vegh 1999; Mendonza and Uribe 2001; Kumhof et al. 2007; etc.). Such a vulnerability could arise from a weak fiscal revenue base, implicit financial bailout guarantees, contingent government liabilities, etc. In short, if fiscal discipline is relatively lax, then achieving macroeconomic stability by strict monetary discipline can be counterproductive. The Fiscal Responsibility and Budget Management (FRBM) Act in India does seem to promise an era of fiscal discipline in the future, but in general fiscal discipline seems to be far more difficult to achieve than monetary discipline.[15]

Finally, the empirical evidence on the success of IT regimes is mixed.[16] Ball and Sheridan (2003) come up with the finding that '. . . there is no evidence that inflation targeting improves performance', whereas Levin et al. (2004), Hyvonen (2004) and Vega and Winkelried (2005) report a lowering of inflation persistence and an anchoring of inflationary expectations for ITers.[17]

Let me now turn to the other major issue regarding the macroeconomic framework, viz. CAC. Here, the CFSR makes a strong pitch for an accelerated move in the

[15] The current total fiscal deficit—both central and state and including several contingent liabilities—at 12 % of GDP may perhaps be regarded as exceptional and likely to be moderated as the fiscal stimulus is wound down.

[16] Countries using some form of IT currently include Australia, Brazil, Canada, Chile, Colombia, Finland, Mexico, New Zealand, Poland, Sweden, UK, etc.

[17] For a fuller discussion of this viewpoint, kindly refer Nachane (2008).

direction of CAC.[18] As we have mentioned above, there has been in the aftermath of the current crisis a resurgence of interest in the Keynesian view of financial markets (or the so-called *saltwater* view). A logical corollary of this tilt towards Keynesianism has been a great deal of rethinking on the issue of capital controls—within the academic community as well as in several official circles. Most significant in this context is the revised stance of the IMF on its pet bugbear of capital controls. Capital controls, which had (particularly since the 1980s) been an anathema to the IMF's thinking, are now not only admitted, but even actively promoted. As a matter of fact, when Iceland's banking system collapsed in September 2008, a key component of the IMF reform package was 'controls on capital outflows'. Several countries in central and eastern Europe (including Turkey, Russia, Kazakhstan, Ukraine, Poland, Bulgaria, etc.) and Africa have either introduced some form of capital controls or are on the verge of doing so.[19]

Indian policymakers, right from the inception of reforms, have shown tremendous enthusiasm for accelerated capital account liberalization. The two committees appointed to examine the issue (Tarapore I and II) have laid out a detailed road map for full CAC. It is important to stress that the line taken by several apologists for CAC that the risks of financial instability are negligible and hence more than compensated for by the benefits ignores the magnitude of the potential costs of a crisis which have been carefully noted by Rakesh Mohan (*Banque de France Seminar* 14 June 2007). The total welfare costs would be substantially higher given the fact that the poor and vulnerable sections of the society have to bear a disproportionately large share of the costs.

Surprisingly, the pronounced swing of opinion against unfettered capital account liberalization which has occurred among a majority of academic economists, as well as several foreign governments and multilateral institutions (the IMF not excluded), in the light of the recent financial upheavals seems to have completely bypassed Indian policy circles. Indian policymakers need to remember that empirical studies fail to demonstrate any clear and convincing evidence of a favourable impact of CAC on total factor productivity, economic growth and poverty reduction—even where such effects are in fact detected, they are circumscribed by a host of conditioning factors including levels of economic development, institutional quality, sequencing of reforms, etc. (see, in particular, the summary evaluation in the recent report of the Bank of International Settlements (BIS) Committee on the Global Financial System 2009 under the chairmanship of Rakesh Mohan). On the other hand, CAC poses very real threats to financial stability and monetary policy autonomy, especially for

---

[18] Rajan's strong advocacy of CAC contrasts strangely with the pragmatic (and far more nuanced) approach to capital account liberalization that he espouses (along with Prasad ES) in the *Journal of Economic Perspectives* (Summer 2008).

[19] Academic thinking, in the highest circles, is also veering strongly towards the need for capital controls to cite but two opinions from a long list. First, Paul Krugman (*The New York Times* 12 September 2009) has this to say on capital controls 'Back in 1998, in the midst of the Asian financial crisis, I came out in favor of temporary capital controls . . . At the time it was regarded as a horribly unorthodox and irresponsible suggestion . . . Today, that wild and crazy idea is so orthodox it's part of standard IMF policy.' Second, in a recent lecture De Long talks about 'the intellectual bankruptcy of the Chicago School' (*6th Singapore Economic Review Public Lecture,* 7 January 2009).

countries with weak regulatory mechanisms and undeveloped financial markets (see, in particular, Y.V. Reddy's recent book *India and the Global Financial Crisis* 2009).

There are several further issues specifically germane to the Indian situation. Firstly, a substantial opening of the capital account has already taken place over the past decade, and (since CAC can only convey short-term growth effects (see Henry 2007)), whatever benefits of opening the capital account that were due, must have already accrued. Any further opening up of the capital account can only convey marginal benefits, while increasing the risks of financial instability substantially. Participatory notes (PNs) also present several problems most notably related to anti-money laundering and terrorist funding—see the speech by M.K. Narayanan, former National Security Advisor, Government of India at the *43rd Munich Conference on Security Policy* (2007). While they do lend considerable liquidity to the stock market, PNs can hardly be viewed as a source of getting long-term funds into India.

Irrespective of whether India decides to go for full CAC or otherwise, management of capital inflows will remain an important issue for some time into the future. One rational policy response would then be to examine a minimal set of capital account restrictions that will mitigate the probability of financial crises of the order of the Asian crisis (1997–1998), the Long-Term Capital Management (LTCM) crisis (1998) or the Russian crisis (1998). Various proposals for managing capital inflows have been made including Tobin taxes, variable deposit requirements, interest equalization taxes, group insurance, etc. Of these, the trip-wire speed bump approach (TW-SB) is the one which I find particularly appealing. The essence of this approach is simple. Certain basic indicators (trip wires or TWs) are defined and as and when these indicators deteriorate (below a threshold), certain safety measures (relating to capital account transactions)—speed bumps or SBs—are 'triggered off' (see Nachane 2007). The TW-SB approach has several points of commonality with the early warning system advocated at the recently concluded Pittsburgh Summit of the G20.

2. Principles versus Rules-based Regulation The Principles vs Rules mode of regulation controversy was first brought into the picture (in the Indian context) by the Committee on making Mumbai an IFC. The Committee chastised the RBI for the plethora of rules that financial institutions are required to follow and strongly advocated a switchover to a *principle-based* system. The CFSR reiterates the same position but with less rhetoric and greater attention to detail. Principle-based regulation involves greater reliance on 'principles and outcome-focused, high-level rules as a means to drive at the regulatory aims we want to achieve, and less reliance on prescriptive rules' (FSA 2007). Essentially, two concepts are involved, viz. principles restraining regulatory discretion and general guidelines that might supplant the existing detailed rules for auditors and regulated entities. It is the CFSR's contention that the current rule-based system in India displays 'low tolerance for innovation and excessive micro-management' (Chap. 6, p. 2). It therefore recommends a gradual but time-bound movement in the direction of principle-based regulation.

There are several imminent problems with the adoption of a principle-based approach for India. At least a few of these deserve to be mentioned:

a. In a principle-based system, the interpretation of principles is often with the regulator, in contrast to a rule-based system, where interpretation lies equally with the regulator and the regulated, with well-defined mechanisms for resolving conflicts of interpretation. Thus, ironically, a principle-based system places a greater discretion at the disposal of the regulator. This can lead often to arbitrary regulation, but the greater danger is of attempts by powerful corporate interests at *regulatory capture* and blocking of competition (the recent Wal-Mart case in the USA is an example—see *Financial Times* 5 July 2007).
b. It is not exactly clear whether the CFSR is suggesting a principle-based system only for financial institutions or for the entire corporate sector. Either interpretation is fraught with difficulties. If principle-based regulation is to apply for the entire corporate sector, the details spelt out in the CFSR for implementing such a gigantic scheme are extremely sketchy, being confined to generalities rather than going into specificities. Too much stress on self-regulation and expecting firms to appreciate their own regulatory responsibilities is an unwarranted utopian view of the Indian corporate mentality. If, on the other hand, the CFSR is making the more modest suggestion of a principle-based system for the financial sector, with the existing system largely intact for the non-financial sector, the procedure is more feasible but will create problems arising from the interface of two distinct systems. In a litigious country like India, the arbitration/judicial system will be overwhelmed with public-interest litigations (PILs), right to information (RTI) queries and *private class actions*.
c. Finally, as noted by Wallison (2007), there is the *safe haven* effect of a rule-based system. Compliance with rules, which are fully transparent, gives the regulated entities a sense of absolution, which is never present in a principle-based system. There is besides no question of discrimination between different regulated entities in a rule-based system, a problem which is never totally absent in a principle-based system.

All this is hardly to say that the existing rule-based system in India is without defects, and several of these are highlighted effectively in the CFSR Report. Instead of a switchover to a principle-based system, a far better alternative is to impart flexibility and dynamism to the existing rules, making them more transparent and installing a system of quick incentives/penalties for compliance/non-compliance.

3. Regulatory and Supervisory Independence (RSI) The issue of RSI is often confused with central bank independence (CBI), though as stressed in the literature (see Lastra 1996; Taylor and Fleming 1999; Quintyn and Taylor 2002), the two are conceptually distinct and need not necessarily coexist even when the regulation and supervision functions and the monetary policy functions are vested in the same authority. In a sense, RSI is to financial stability what CBI is to monetary stability. Unfortunately, the academic literature on regulation has been almost exclusively focused on CBI, to the virtual neglect of RSI. The CFSR also fails to touch on this aspect at all, though of course it does pay a great deal of attention to the issue of single versus multiple regulators.[20]

The neglect of RSI assumes importance when one considers the fact that almost all episodes of financial distress have been associated with a weak RSI.[21] RSI refers to independence of the regulatory and supervisory structure from not only the government but also the industry and financial markets. In India, the financial regulatory and supervision functions are distributed between the RBI (banks and NBFCs), state governments (for cooperative financial institutions jointly with RBI) and the National Bank for Agriculture and Rural Development (NABARD), for regional rural banks (RRBs). For the purposes of this discussion, let us confine ourselves to the regulation and supervision of the banking sector and the NBFCs. The RBI discharges this function under the guidance of the Board for Financial Supervision (BFS), which comprises four directors from the RBI's central board, the RBI governor (as chairman) and four deputy governors.

So far as independence from the government on the regulatory and supervisor fronts is concerned, this is ensured to a large extent by the fact that the RBI (acting under the guidance of the BFS) is authorized to issue directives in all areas of regulation and supervision. However, this realization has to be tempered by the fact that an element of indirect control of the government does exist by virtue of the fact that the RBI directors (from whom four of the BFS members are drawn) are appointed by the central government. Incidentally, the CFSR's recommendation to set up the Financial Development Council under the chairmanship of the finance minister 'for macroeconomic assessment and development issues' (Proposal 26), has already been implemented. Some fears have however been expressed that this council will strongly limit the existing independence of the regulators and supervisors, as it will provide a legitimate platform for the Finance Ministry to intervene in several matters, and further exacerbate the coordination problems between the RBI and the Finance Ministry.

The other major dimension of RSI, viz. independence from markets is equally important but has not received the attention it deserves. In the words of a very famous US central banker, '. . . it is just as important for a central bank to be independent of markets as it is to be independent of politics' (see Blinder 1997). Independence from markets is more difficult to ensure than independence from politicians, since the forces operative here are extremely subtle. This can occur primarily through two channels, both of which have been operative in the Indian context. Firstly, there is an overwhelming preponderance of financial sector and industry representatives in official committees and bodies, concerned with the designing of the regulatory architecture. This usually takes place at the instance of a government strongly committed to reforms and is usually done with the ostensible purpose of taking on board

[20] The CFSR's preferred regulatory architecture (Proposals 23–28) is one where all depository institutions come under the supervisory purview of the RBI, with a separate agency for supervising large systematically important financial conglomerates.

[21] See De Krivoy (2000) for the Venezuelan experience of the mid-1990s, Lindgren et al. (1999) for the East Asian experience, Hartcher (1998) for Japan, etc.

the 'financial industry' point of view. Secondly, large sections of the media are strongly aligned with corporate sector interests and, hence, by extremely ingenious propaganda manage to create a general impression that national interests are closely conflated with financial sector interests. A *grading system* is then set up, whereby supervisors and regulators are rated publicly on how friendly they are to markets. We are treading on thin ground here. On the one hand, financial stability is a public good, and financial market development contributes to real development. Yet, it is undeniable that exuberance, animal spirits and general short-termism strongly pervade financial markets. A regulatory authority overtly sensitive to financial market demands could be a classic case of what Stigler (1971) has termed *regulatory capture*. Unfortunately, the CFSR observes a deafening silence on this vital issue.

10 A Suggested Agenda

At the outset, let me emphasize that there is a fundamental difference between the crises in the USA, EU and India. In the USA, the crisis originated endogenously within the financial system and then spread from *Wall Street to Main Street* (to use President Obama's famous expression). In the EU and other Western countries, the financial system was first affected largely by contamination from the US financial system and then the crisis spread to the real economy. In India (and some other Asian countries), the primary source of contagion has been via the trade channel, so that the real sector has been affected to a great extent but the financial system has been intact. It is true that sporadic evidence of exposure of domestic private sector banks (in India), as well as some nationalized banks and foreign subsidiaries, to the so-called toxic assets and CDOs in the USA and EU has come to light from time to time, but the extent of total exposure is likely to be limited to something like US$ 1.5 billion (on a mark-to-market basis). A substantial share of the credit for the robustness of the Indian financial system must go to the former RBI Governor Y.V. Reddy, for carefully monitoring the securitization process in India and forestalling the emergence of asset bubbles feeding on indiscriminate credit expansion. *The New York Times* (20 December 2008) came closest to the mark when it described him as 'the right man in the right job at the right time'. Given that the recessions in the USA, EU and India represent three distinct patterns, the nature of the Indian policy response should not necessarily track theirs but should be specially designed to account for the specificities of our situation. In particular, three concerns should be paramount in the Indian context, viz.:

1. Revival sans Stagflation Firstly, there is, of course, the need to revive the real economy without, in the process, unleashing forces that could trigger a future asset and/or commodity price inflation.

2. Firewalls Around the Financial Sector As mentioned above, since the Indian crisis is largely an imported one (primarily via the trade and investment routes) and, further, since the financial sector has been more or less secure so far, the policy should emphasize the insulation of the Indian financial sector from adverse shocks originating either in the Indian real sector or in the financial systems of the USA and EU.

3. Safety Nets for the Vulnerable Sections As International Labour Organization (ILO)'s *Global Employment Report* (January 2009) points out, in a classic moral twist to the global crisis tale, those who had the least to do with the perpetration of the crisis (namely, the vulnerable sections of society across the globe) are being forced to bear the brunt of the consequences. India is no exception and hence a necessary third pillar of any anti-recessionary strategy should be to build extensive safety nets for those at maximum risk of exposure to collateral damage.

Indian policy in the aftermath of the crisis[22] has been addressed almost exclusively to the first objective, with attention to the second confined mainly towards bank capitalization (see Footnote 16), while the last objective has largely languished in the domain of rhetoric.

A policy package, in consonance with the three objectives set out above, would be one that would include the following specific measures:

1. Any further monetary policy easing or fiscal stimulus runs the grave danger of laying the foundation for a future high inflation phase. Given the long lags in monetary and fiscal policy[23] (between two and three quarters), the effects of the policy measures taken so far are likely to take effect towards the end of 2009, just about the time when the excess capacity in the economy is estimated to be working itself out.
2. There is no denying that the failure of credit delivery to micro, small and medium enterprises (MSMEs) is having systemically important effects in delaying the recovery and in aggravating social distress due to job losses etc. It is now time to strike out boldly by attempting measures like government guarantees of loans to MSMEs along the lines of the Mandelson Plan in UK. (Lord Mandelson, Business Secretary, UK government has announced plans to guarantee bank loans to

[22] The policy measures so far adopted in India may be summed up in a single phrase—easy money and fiscal stimuli. On the monetary policy front, there has been a flurry of activity—the repo rate was reduced in a succession of steps from the level of 9 % in September 2008 to 5 % in March 2009 (with a corresponding reduction in the reverse repo rate from 6 to 3.5 %), the cash reserve ratio (CRR) was also reduced from 9 to 5 % over the same period, whereas the statutory liquidity ratio (SLR) was brought down by 1–24 %. Altogether, it has been estimated that these measures have released more than Rs. 4,000,000,000,000 (US$ 80 billion approximately) of liquidity into the system. There have also been three successive fiscal stimuli packages amounting to a total cost of Rs. 801,000,000,000 (US$ 16.3 billion) to the Exchequer. Fiscal stimulus I (7 December 2008) mainly comprised an across-the-board cut of 4 % in excise duty (estimated cost of Rs. 310,000,000,000). Fiscal stimulus II (2 January 2009) comprised Rs. 200,000,000,000 towards bank capitalization over the next 2 years, as well as providing greater market borrowing access to state governments as well as the IIFCL (India Infrastructure Financing Co. Ltd.) (estimated cost of Rs. 700,000,000,000). The final stimulus III (24 February 2009) provides a 2 % reduction in both the excise duty and the service tax and an extension of the previous excise duty cuts beyond 31 March 2009 (estimated cost of Rs. 291,000,000,000). The total burden on the Exchequer at Rs. 810,000,000,000 amounts to nearly 1.82 % of the 2008–2009 GDP (at current prices) or 2.57 % (at constant prices).

[23] No systematic estimates of these lags are available in the Indian case. Some work in progress currently by the author estimates the lags in monetary policy at around 8 months and for fiscal policy around 12 months. However, these estimates have yet to be firmed up.

SMEs with sales of up to $\leq$ £ 500 mn). Actually, there is provision for credit guarantees under the Deposit Insurance and Credit Guarantee Corporation (DICGC) Act 1961. However, this has now become defunct (see DICGC Annual Report 2007–2008, p. 1). The DICGC needs to be strengthened with an infusion of funds and entrusted with the responsibility of administering such a scheme.

3. There already exist provisions for special treatment of risk weights on loans to MSMEs under Basel II. The provisions envisage exemption of loans to MSMEs from capital requirements (or at least assigning these loans a lower risk relative to larger size firms in the same rating category).
4. As a purely temporary measure, for the duration of the crisis, loans above a certain limit to industries in sensitive sectors can be tied to some employment protection guarantees.
5. Encouraging innovative schemes like SMECARE & SMEHELP (initiated by the State Bank of India, SBI) for adoption by other banks on an extensive scale.[24]
6. Financial crises affect vulnerable sections of society (including labour) far more than non-vulnerable sections. Hence, in the interests of such sections, ensuring against financial contagion should receive top priority. The general monetary and fiscal measures undertaken so far contribute very little to this objective, with the possible exception of the bank capitalization provisions under Fiscal Stimulus II. But bank capitalization is not an insurance against a crisis—it is at best a damage-limitation measure in the event of a crisis actually occurring. Hence, there is need for several prudential and 'fire-fighting' measures such as:
   a. A switchover to a system of risk-based deposit insurance relying on a system of fair value accounting.[25]
   b. A rise in the deposit insurance coverage from the current Rs. 100,000 to Rs. 500,000. This will provide a much-needed safety net for the savings of the middle classes.[26]
   c. The role of rating agencies in the perpetration of the current crisis has come under heavy scrutiny from economists like Buiter (2008), Portes (2008), Giovanni and Spaventa (2008), etc. Such criticism has prompted the Financial Stability Forum (now FSB), through the IOSCO to offer a code of conduct for credit-rating agencies. The RBI should see that this code of conduct is accepted and adhered to, by major credit-rating agencies in their Indian operations.
   d. A strict monitoring of off-balance sheet items and structured product vehicles (SPVs) of banks and financial institutions.

[24] Under SMECARE, MSME borrowers (with fund-based limits of up to Rs. 100,000,000) can avail additional working capital of up to 20 % of their existing fund-based limits, whereas under SMEHELP a 5-year tenured loan is extendable for MSMEs with a liberal margin of 15 % for financing capital expenditure. Both schemes offer loans at a concessional rate of 8 % during the 1st year. These concessional schemes are mainly in the areas of pharmaceuticals, food processing and light engineering goods.

[25] Such fair value accounting could be on the lines of the SFAS No.133 issued by the US Financial Accounting Standards Board in 1998.

[26] The concept of middle class used here corresponds to that employed in Sengupta et al. (2008).

e. There is a need to recognize that substantive capital flows (in either direction) have potentially strong destabilizing consequences. In such circumstances, it is necessary to reserve for ourselves the right to impose key capital controls on a pre-announced basis for specified periods of time. The extent and duration of these controls could be related to the setting off of certain macroeconomic triggers. This TW-SB approach is elaborated at length in my earlier paper (Nachane 2007).[27] The fact that Brazil introduced a Tobin tax of 2 % on capital inflows shows that at least some EMEs are convinced by the merits of this approach.
f. One of the most effective safety nets for the poor has been suggested by the National Commission for Employment in the Unorganized Sector (NCEUS). This involves the setting up of a National Fund for the Unorganized Sector (NAFUS). The Fund is proposed to have an authorized capital of Rs. 10,000,000,000. and would be designed to provide (i) refinance to banks and other financial institutions to supplement their efforts to provide credit to unorganized sector enterprises with investment in plant and machinery less than Rs. 2,500,000, but with a special focus on enterprises with investment less than Rs. 500,000; (ii) microfinance support through NGOs/self-help groups (SHGS)/MFIs, etc. and (iii) venture capital for innovative enterprises in the unorganized sector. This suggestion of the NCEUS requires an extremely serious consideration from the government, though there are no signs that this is happening.

11 Conclusion

While there is no denying the fact that financial system development is an integral component of overall development, there are important caveats to this general statement. The current financial crisis has exposed some clear fault lines in unchecked financial innovation and deregulation. In particular, opinion seems to be swinging away from the pristine view of free markets evident in classical laissez faire to the more nuanced view of Keynes. This shift in thinking has challenged several established orthodoxies, and as economists grapple to resolve their controversies, policymakers are struggling to find solutions to hitherto unencountered problems. Robert Posner's recent article (*The New Republic* 23 September 2009) is an honest

[27] In direct contrast to this view, we have a substantially influential group of Indian economists who see in the crisis an opportunity for introducing capital account convertibility. Thus, Lahiri in his P. T. Memorial Lecture (16 January 2009) says, 'The current crisis may provide an opportunity for introducing capital account convertibility. The dominant worry about introducing convertibility has been an upsurge of capital flows with large upward pressure on the exchange rate of the rupee followed by a sudden sucking out of such a capital, precipitating a crisis. Risk aversion on the part of international investors is an all-time high now, and the risk of large inflows is limited.' Not all of us may be persuaded by this somewhat convoluted logic, though it seems to have provided considerable grist to the mill for several professional bloggers.

admission of the profession's confusion, wherein he says, 'We have learned since September that the present generation of economists has not figured out how the economy works. The vast majority of them were blindsided by the housing bubble and the ensuing banking crisis; and misjudged the gravity of the economic downturn that resulted ... By now a majority of economists are in general agreement with the Obama administration's exceedingly Keynesian strategy for digging the economy out of its deep hole.'

As the global economy is slowly emerging from the crisis, certain things are becoming clear—in particular the inconsistencies in regulatory systems across countries and clear conflicts of interests between regulators across borders as well as between regulators and financial markets. A new era of global financial coordination to deal with global systemic risk seems to be dawning. However, this will have to contend with four formidable and fundamental issues, viz.: (i) *the coordination of regulations,* (ii) *coordination of resolution tools,* (iii) *coordination in depositor and investor protection* and (iv) *enhanced information sharing*.

The global coordination process would essentially involve four main partners, viz.:

1. National regulatory and supervisory authorities
2. IMF
3. FSB and other international standard-setting bodies—BCBS, IOSCO, etc.
4. Influential groups like G20

The success of the global coordination process would depend upon how sincerely these four main partners execute their respective mandates.

It is interesting to note that in several EMEs, the fact that the consequences of the crisis have been relatively muted seems to have lulled policymakers into a sense of security and convinced them that the financial liberalization process can continue in the same vein as before. The Indian case is particularly noteworthy where the revolutionary change in thinking now taking place globally has completely bypassed official policy thinking, which, from all apparent signs, seems to find it difficult to rid itself of the pre-crisis euphoria about financial liberalization, as encapsulated in the two Tarapore Committee reports as well as the more recent Percy Mistry and Raghuram Rajan reports. India (along with other countries in South Asia) has the unique opportunity to benefit from the lessons learnt from the current crisis, most of whose fallout has been on the developed economies of the West. I cannot resist quoting an old Confucian adage: 'Any fool can learn from his own mistakes. It takes a truly wise man to learn from the mistakes of others.' Will policymakers in these countries show that kind of wisdom?

References

Alcala F, Ciccone A (2004) Trade and productivity. Quart J Econ 119(2):612–645

Archer D (2006) Implications of recent changes in banking for the conduct of monetary policy. BIS Papers No 28, p. 123–15

Ashcraft A, Schuerman T (2008) Understanding the securitization of subprime mortgage credit. Staff Report No. 318, Federal Reserve Bank of New York (March)

Ball LM, Sheridan N (2003) Does inflation targeting matter? IMF Working Paper No 03/129

Birdsall N (2009) IMF governance reform committee Report leaves much to the imagination. Centre for Global Development, 6 April 2009
BIS (Committee on the Global Financial System—CGFS) (2009) Capital flows and emerging market economies. CGFS Paper No 33
Blinder AS (1997) What central bankers could learn from academics—and vice versa. J Econ Perspect 11(2):3–19
Bordo M, Dueker M, Wheelock D (2000) Aggregate price shocks and financial stability: an historical analysis. NBER Working Paper No 7652
Borio C, Lowe P (2002) Asset prices, financial and monetary stability: exploring the nexus. BIS Working Paper No 114
Brunnermeier MK (2009) Deciphering the liquidity and credit crunch 2007-2008. J Econ Perspect Am Econ Assoc 23(1, Winter):77–100
Brunnermeier M, Crockett A, Goodhart C, Persaud A, Shin H (2009) Fundamental principles of financial regulation. *Geneva Reports on the World Economy*, 11 Jan 2009
Brunnermeier MK, Pedersen LH (2009) Market liquidity and funding liquidity. Review of Financial Studies, Oxford University Press for Society for Financial Studies, Vol 22(6), pp. 2201–2238, June
Buiter W (2009) Regulating the new financial sector. http://www.voxeu.org/index.php?q=node/3232. 9 March 2009
Buiter WH (2008) Housing wealth isn't wealth, NBER Working Paper No 14204
Calomiris CW (2007) Bank failures in theory and history: the great depression and other contagious events. NBER Working Papers No. 13597, National Bureau of Economic Research
Calomiris CW, Powell A (2000) Can emerging market bank regulators establish credit discipline? The case of Argentina 1992-99 NBER Working Paper No 7715
Calvo G, Vegh C (1999) Inflation stabilization and BOP crises in developing countries. In Taylor JB, Woodford M (ed) Handbook of macroeconomics, Vol 1C, Elsevier-North Holland, Amsterdam
Claessens S (2009) The financial crisis and financial nationalism. Joint World Bank-CEPR Conference, Brussels, 26–27 May
De Krivoy R (2000) Collapse: the Venezuelan banking crisis of 1994, Group of Thirty, Washington, DC
De Larosiere Group (2009) Report of the high-level group on financial supervision in the EU, Brussels, Feb 2009
De Long JB (2009) What has happened to Milton Friedman's Chicago School? 6th Singapore Economic Review Public Lecture, 7 Jan 2009
Demirguc-Kunt A, Detragiache E (1998) Financial liberalization and financial fragility, IMF Working Paper No. 98/83
Eichengreen B (2001) The EMS crisis in retrospect. CEPR Discussion Paper No. 2704
Evanoff D, Wall L (2000) Subordinated debt and bank capital reform, FRB Atlanta Working Paper No. 2000–24
G20 (2009) Enhancing sound regulation and strengthening transparency, Final Report of the G20 Working Group 1
Ghosh S, Nachane D (2003) Are Basel capital standards pro-cyclical? Some empirical evidence from India. Econ Polit Wkly XXXVIII(8, Feb.22–28):777–784
Giovanni A, Spaventa L (2008) Filling the information gap. In Felton A, Reinhart C (ed) The first global financial crisis of the 21st century. http://www.voxeu.org/index.php?q=node/690
Goldsmith R (1969) Financial Structure and Development. Yale Univ Press, New Haven
Goldstein M (2008) A proposal to improve banks' regulatory liquidity. *Financial Times*, 22 May 2008
Gorton GB (2008) The subprime panic. NBER Working Paper No 14398, October 2008
Hartcher P (1998) The ministry. How Japan's most powerful institution endangers world markets, Harvard Business School Press, Boston
Hicks J (1969) A theory of economic history. Oxford University Press, Clarendon
Hyvonen M (2004) Inflation convergence across countries. Reserve Bank of Australia Research Discussion Paper No 2004–04

Kahneman D, Tversky A (1984) Choices, values and frames. Am Psychol 39(4):341–350
Kaufmann D, Kraay A, Mastruzzi H (2007) Governance matters VI: governance indicators for 1996-2006. World Bank Policy Research Discussion Paper 4280
Keehn S (1989) Banking on the balance: powers and the safety net. Monograph, Federal Reserve Bank of Chicago
Krugman P (2009) A dark age of macroeconomics (wonkish). The New York Times, 27 Jan 2009
Kumhof M, Li S, Yan I (2007) Balance of payments crises under inflation targeting. IMF Working Paper No. 84
Lastra RM (1996). Central banking and banking regulation, Financial Markets Group, LSE
Leroy SF, Porter RD (1981) The present value relation: tests based on implied variance bounds. Econometrica 49(3):555–574
Levin AT, Natalucci F, Piger J (2004) The macroeconomic effects of inflation targeting. FRB of St. Louis Review, Vol 86(4), p.51–80
Levine R (1997) Financial development and economic growth: Views and Agenda. J Econ Literat XXXV(June):688–726
Lindgren C, Balino T, Enoch C, Gulde A, Quintyn M, Teo L (1999) Financial sector crisis and restructuring: lessons from Asia. IMF Occasional Paper No.188
Lucas R (1988) On the mechanics of economic development. J Monet Econ 22(1):3–42
Meier G, Seers D (1984) (ed) Pioneers in development. Oxford University Press, New York
Mendonza E, Uribe M (2001) The business cycles of balance-of-payments crises: a revision of the Mundellian framework. In Calvo G et al (eds) Money, capital mobility and trade: essays in honour of Robert A Mundell. MIT Press, Cambridge
Mohan R (2007) Capital account liberalization and conduct of monetary policy: the Indian experience. Paper presented at a seminar on Globalization, Inflation and Financial Markets (Banque de France, Paris, France) 14 June 2007
Nachane D (2007) Liberalisation of the capital account: perils and possible safeguards. Econ Polit Wkly XLII(36, Sept. 8–14):3633–3643
Nachane D (2008) Committee on financial sector reforms: a critique. Econ Polit Wkly XLIII(32, Aug 9–15):3633–3643
Nachane D, Islam MS (2009) Financial sector reforms in South Asia: a perspective (forthcoming)
Narayanan MK (2007) Address at the 43rd Munich conference on security policy, 11 Feb 2007
Portes R (2008) Ratings agency reforms. In Felton A, Reinhart C (eds) The first global financial crisis of the 21st century. http://www.voxeu.org/index.php?q=node/690
Posner R (2009) How I became a Keynesian. The New Republic, 23 Sept 2009
Prasad E (2009) The risk of a resurgence of global imbalances. Economic Times, 14 Sept 2009
Prasad ES, Rajan R (2008) A pragmatic approach to capital account liberalization. J Econ Persp 2(3):149–172
Quintyn M, Taylor MW (2002) Regulatory and supervisory Independence and financial stability. IMF Working Paper No. WP/02/46
Rabin M, Thaler RH (2001) Anomalies: risk aversion. J Econ Perspect 15(1):219–232
Reddy YV (2009) India and the global financial crisis. Orient Blackswan, Delhi
Robinson J (1952) The rate of interest and other essays. Macmillan, London
Rodrik D, Subramanian A, Trebbi F (2002) Institutions rule: the primacy of institutions over geography and integration in economic development. NBER Working Paper 9305
Saidane D (2002) The role of financial liberalization in development: weaknesses and corrections. Savings Develop 26(3):259–276
Schumpeter J (1912) The theory of economic development (trans by Redvers O) Harvard Univ Press, Cambridge (1934)
Stigler G (1971) The theory of economic regulation. Bell J Econ Manage Sci 6(2) 114–141
Seers D (1983) The political economy of nationalism. Oxford Univ Press, New York
Shiller RJ (1981) Do stock prices move too much to be justified by subsequent changes in dividends? Am Econ Rev 71(3):421–436
Shleifer A, Summers LH (1990) The noise trader approach to finance. J Econ Perspect 4(1):19–33

Stern N (1989) The economics of development: a survey. Econ J 99:597–685
Taylor M, Fleming A (1999) Integrated financial supervision: lessons of Northern European experience. World Bank Policy Research Paper No. 2223
Vega M, Winkelried D (2005) Inflation targeting and inflation behaviour: a successful story? Int J Cent Bank 1(3):153–175
Wallison PJ (2007) America will prefer to rely on rules, not principles. Financial Times, 6 July 2007

Chapter 8
Global Capital Flows and Payment Imbalances

Rameshwar Tandon and Shariq Mohd.

1 Introduction

Alongside a tendency for the globalization of production through foreign direct investment, we have witnessed a parallel tendency towards globalization of foreign capital and indeed, a period of revolutionary innovations in the way global financial capital conducts its business. 'Such innovations involve often quite complex structured financial products, which aim to diversify risk globally' observed George Argitis and Christos Pitelis (Argitis and Pitelis 2008).

Whether recession or depression, 'The present crisis is a crisis of the capitalist system, like the East Asian crisis a decade ago or the Japanese crisis before that and it is the first crisis of the system since the 1930s. It is global in scope, so there are no fast growing regions of the world, which can provide support for the resumption of growth in the crisis affected regions', Robert Wade put it so tersely (Wade 2009b, p. 39).

Wade continues further, 'The current financial and economic crisis that has forced the likes of Greenspan to question the coherence of dominant conceptual framework is unprecedented in global reach and systematic gravity. Rather global imbalances have had an important causal role, not at the international level in the form of 'currency' recycling, but at the domestic level in the form of 'credit recycling' to the agents, spending more than their income, who are the other end of the external deficit. The breakdown occurred in the 'credit' recycling mechanism.

Nobel Laureate Paul Krugman is so worried, 'This looks an awful lot to the beginning of a second Great Depression . . . Further, recent economic numbers have been terrifying, not just in the United States, but around the world. Manufacturing in

---

This paper was presented at a national Seminar at the IISE, Lucknow and also in the Lucknow University, India.

R. Tandon (✉)
International Institute of Special Education, Lucknow, 220007, India
e-mail: rameshwartandon@rediffmail.com

S. Mohd.
e-mail: shariqmohd2008@yahoo.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_8, © Springer India 2014

particular is plunging everywhere. Banks are not lending; business and consumers are not spending' (Krugman 2009).

Long back, Charles Kindleberger held that financial markets are likely to be intrinsically unstable and particularly, with 'excess' liquidity, unable to self-correct (Kindleberger 1986). In fact, instead of automatic stabilizers, they can easily be derailed even further by automatic 'de-stabilizers' (Stiglitz 2003). As a result, borrowers and lenders can accumulate far more risk than is privately efficient—let alone socially efficient. Hence, financial crises can occur for more reasons than irresponsible government deficits, moral hazards due to government guarantees, bad luck, mob psychology, crony capitalism, misguided policy or other 'exogenous' factors, interfering with the otherwise perfectly efficient allocation of resources by financial markets. The current financial crisis is a paradigmatic case of this, observed J.G. Palma (2009).

As Wade asks angrily, 'Are we moving away from the normative primacy of the liberal capitalist economy, in which the state acts in a regulatory and facilitating role and towards a "government" capitalist economy (with the state more involved in sponsoring structural changes), or a "coordinated" capitalist economy, to buffer the costs of change and the framework for organized actors to coordinate their investment decisions.' It seems that today a page has been turned; policies excoriated by the right—nationalization, higher taxes, Keynesian economics, financial regulations—are suddenly back on the agenda of liberal capitalist economics. So the big question is whether we will add 2008 after 1945 and 1980 (Wade 2009a).

We find it is not a single phenomenon, with a single cause. In the Anglo-American heartland, it began and remained until September 2008, primarily a financial crisis, though with 'real economy' causes (a structural deficit in the production of tradable goods and services) and also 'real' economy effects. In Japan, Germany and much of the periphery, it began later, primarily as an export crisis, in response to slowdown in the Anglo-American heartland, reflecting a structural surplus capacity to produce tradable goods and services but the export crisis then fed through finance and wider growth problems.

2 Another Tipping Point?

Serious observers like Robert Wade fear that the 'world economy will probably hit another tipping point—similar to the one in September 2008, with the collapse of Lehman brothers—in the summer-autumn of 2010. The tipping point will be caused by rising general awareness throughout Europe, America, Asia and South America that hundreds of millions of people are experiencing rapidly falling consumption; that the crisis is getting worse, not better, and that it has escaped the control of public authorities. More big banks and non-financial companies are likely to fall, raising the level of fear in the economy and prompting banks to contract still more. More Icelands may be coming in East and central Europe, from the Baltics down to the Balkans and Turkey. Small countries, with their own currencies and loads of foreign debt, are especially vulnerable, as "investors", speculators flee into the relative safety of the well-resourced central currencies' (Wade 2009a, p. 40).

Moreover, one of the most dramatic portents of more trouble ahead is the bankrupt California State Government's resort to paying employees with its own I owe you's (IOUs), because it has no more dollars. This is part of the inner erosion of the US$, now heavily dependent on the goodwill of the Chinese Government. Last year, the Chinese Government lent the USA more than US$400 billion, equal to more than 10 % of China's gross domestic product (GDP); opposition is growing within China to its continued bank-rolling of the fabulously rich USA (Dyer 2009). Since a dollar crash would be the metaphorical equivalent of a hydrogen bomb exploding in the existing world economic order, it is lucky for the rest of the world that China is an authoritarian state, which can overrule popular opposition. Hence, global recovery will take several years. In the five worst 'post-war' financial crises in the Organisation for Economic Co-operation and Development (OECD) world, house price fell, on average, by 35 % over 6 years, equity prices by 55 % over 3.5 years and output by 9 % over 4 years, observed Reinhart and Rogoff (2008). As Rogoff further observed, the Hubristic belief in America that 'we do not have financial crisis', is now obviously false, since the banking crises have been almost as common in rich economics, as the developing ones. The Reinhart–Rogoff results make depressing reading indeed. Downturns that follow a financial crisis are typically long and deep on average; GDP per person falls by more than 9 % from its peak and takes almost 2 years to reach bottom. The misery in the jobs market tends to last far longer. The unemployment rate increases by an average of 7 percentage points after a severe meltdown and reaches a peak almost 5 years after its rise began. If the gauge is accurate, unemployment in America is set to rise to an alarming rate of 11–12 % in coming years. The house burst is unlikely to end quickly either. House prices take an average of 5 years to reach their nadir and fall by 36 % in real terms. Equities take less time to reach rock bottom, but lose more than half of their value by the time they get there. Rogoff further observed, 'The most astounding result is the effect on public finances. The real government debt rises by an average of 86 % in countries afflicted by severe crisis. But the damage has little to do with the costs of bailing out banks. Rather ballooning debt reflects a collapse in the receipts, as a consequence of recession and in most countries, a big increase in public spending to shore up the economy. It is really chilling that such huge deterioration in public finances is still not enough to prevent deep and prolonged down turns' (Reinhart and Rogoff 2008).

The numbers are not real guides to the future; one obvious shortcoming is the range of outcomes. While declines in home and equity prices were remarkably uniform after past crises, GDP per person fell and unemployment rose—by much less than the average in some episodes and by far more in others. America's recession could be milder than the average post-crisis downturn, but it could also be much deeper. But in fact the recession has barely begun and will be long and deep for some more years.

'Globalization of financial markets affects assets and debts, securities, bank loans and deposits, titles to land and physical capital. Trades in these assets and debts are much easier to globalize than trades in commodities and labour. Indeed their globalization has progressed most rapidly. Nothing is involved in financial transactions, beyond exchanging pieces of paper or making entries in electronic ledgers, James Tobin put it so bluntly'.... He further observed, 'As these have been liberalized in country after country, international financial flows have flooded national security markets and inundated banking systems all over the world. These flows could be the vehicles, by which savings in the advanced capitalist democracies are channeled into productive capital investments in the developing countries of Asia, Africa and Latin America. Or they could be causes of currency crises, recessions and depressions, unemployment and deprivation in these countries, or both' (Tobin 2000).

The economic rationale for internationalization of asset markets is that it can assist the movement of productive capital from wealthy, developed countries to poorer, developing countries. But what matters are the 'net' flows of capital, not the 'gross' volume of transactions. The emerging economies of East Asia as well as some in Latin America and eastern Europe are beneficiaries of foreign business investment. But much of their capital flows have taken the form of short-term loans of hard currencies from banks in financial centres such as Tokyo, New York and Frankfurt, Banks in Korea, Thailand and Indonesia. Crisis came when the lenders, viewing the growing deficit, became distressful and refused to renew the loans.

3 Financial Crisis: Here and There

We should say, the explosion of international financial intermediation after the 1980s and the rising incidence of financial crisis, with cross-border affects, were obviously related. For policy makers, the principal question was what it has long been—when 'real' economy growth rates were sought in excess of those capable of being generated by domestic savings, how were the benefits and costs of financial openness to be distributed? In principle, inward flows of privately owned capital make it possible for real economies to grow more rapidly than if they rely solely on domestic resources. In practice, the extra costs associated with crisis-induced capital outflows, bailouts, and the lost confidence of investors occasionally, threaten to undermine the real economies and set back the process of industrialization and disrupt underlining political and social order. The immediate costs of financial crisis can be huge, their social and political effects insidious and lingering', observed worried Louis Pauly (2008).

Looking back, we can say, the inherent instability of capitalist economies and their propensity to crises have been highlighted so well by veterans like Karl Max, J.M. Keynes, Michael Kalecki, Karl Polanyi, Hyman Minsky, Robert Solow, Paul Krugman, Axel Leijonhufvud, etc. The financial and economic crisis can hardly be understood within the dynamic stochastic general equilibrium (DSGE) framework, since these models derive macroeconomic outcomes from explicit choice-theoretic micro-foundations, where agents are assumed to optimize inter-temporally under rational expectations. The models are stochastic, because they allow random exogenous shocks to the system, whose probability distribution is known to the representative agent. 'It is difficult to believe that a financial crisis can result from decisions of agents that know the probability distribution of future events', observed Robert Frenkel and Martin Rapetti (Frenkel and Rapetti 2001). The contrast between the current crisis and the world pictured by modern macroeconomic theory is striking. The existence of asymmetric information between the lenders and borrowers is important to understand financial market behaviour. The factors that trigger the booming phase preceding a financial crisis are different in developed and developing countries. The conditions that have led to financial crises in poor countries typically arose from the implementation of macroeconomic policies, which created incentives that ended up generating the boom and bust cycles. But on the contrary, in developed countries, the elements that trigger the booming phase have developed 'endogenously' within the domestic financial system.

4 Minsky, Hyman and the Crisis

Minsky's query that 'can it happen again' has been answered by the markets all over. It is not surprising that observers of financial markets have brought Minsky's ideas back from an almost total intellectual exile (Minsky 1977, 1986). The conditions that caused and helped to develop the current crisis in the USA correspond so neatly to Minsky's model of crisis. His model stresses that unregulated market economies are not 'dynamically' stable systems that converge to full employment equilibrium, but systems that are cyclical in nature in which crises are not unusual events.

Minsky worked out the financial side of business cycles more thoroughly than anybody else because in his theory, finance was the cause of the instability of capitalism. To Minsky, the capitalist economy had an ever-present inherent tendency to general speculative booms. The potential endogenity of financial crisis as well as the potentially detrimental effects of imposition of financial interests on other interests have been looked into so well. While Minsky always drew inspiration from Keynes, this 'upward instability' hypothesis stands in contrast to the economy's tendency in Keynes' theory to gravitate to a state of 'unemployment' equilibrium. This contrast between Keynes and Minsky has been emphasized by E. de Antoni, Bellofiore and Ferri (de Antoni 2009; Bellofiore and Ferri 2001). Kindleberger (1978) also draws heavily on Minsky's work and Wray interprets the current crisis in the light of Minsky's theory.

The concepts of hedge, speculative and Ponzi financing are central to Minsky's theory of a systemic fragility. A unit is hedged if expected cash flow from operations substantially exceeds its debt-servicing commitments. It is engaged in speculative finance, if it has to depend on periodically refinancing debt. A Ponzi unit has to constantly borrow more in order to meet its debt-servicing commitments. Minsky argued that a prolonged period of stability would induce some unit to migrate from hedge to speculative and others from speculative to Ponzi finance. This makes the system as a whole increasingly fragile. In a highly fragile economy, no identifiable 'exogenous' stock is needed to unleash a crisis. Some random event can be the trigger (Minsky 1977).

In Minsky's view, stability germinates the seeds of instability. 'The recently highly touted "Great Moderation", which has now come to a crashing end, fits his theory perfectly. And the unmasking of Bernard Madoff and a host of smaller crooks has made the Minsky story almost too perfect', Leijonhufvud (2009a) put it so well. He further observes, 'It can not be doubted that the "low" interest policy of the Federal Reserve System gave significant impetus to the speculative boom. So there is a definite Austrian element to the events of the last few years and Austrian Economists might argue that it makes the Minsky hypothesis of the upward instability of capitalism "otiose". But central bank misjudgment does not, by itself, explain the blossoming of credit default swaps or "squared", "cubed" and multi-sectoral collateralized debt obligations (CDOs). They fit more naturally into the Minsky story—as does Bernard Madoff. Both themes have some validity' (Leijonhufvud 2009a).

5 J.M. Keynes and Financial Side of Recessions

Keynes saw the crisis as being 'endemic' to the system, not an aberration in its financing, as one of its essential characteristics, as opposed to a symptom of its failure. He attributed the crisis to a 'sudden' collapse in the marginal efficiency of capital, which, in turn, was related to the phenomenon of speculation. He defined speculation as distinct from enterprise as follows: 'If I may appropriate the term speculation for the activity of forecasting the psychology of the market and the term "enterprise" for activity of forecasting the prospective yield of assets over their whole life, it is by no means always the case that speculation predominates over-enterprise. Speculators are concerned, not with what an investment is really worth to a man who buys it for keeps, but what the market will value it at, under the influence of mass psychology 3 months or a year. Hence, the warning not to allow the 'real' economy to be governed by the machinations of a casino may be well taken, but once you have ignored it to your peril, then what do you do? The logic of Keynes' liquidity preference theory (LPT) indicates that the primary function of financial markets is to provide liquidity, not efficiency (Keynes 1936, p. 155; see also Keynes 1930/1972).

Peter Bernstein also argues that the LPT and not the efficient market theory (EMT) is the relevant theory for the world in which we live. The fatal flaw in the EMT hypothesis is that there is no such thing as an 'efficient' equilibrium price. A market can never be efficient unless equilibrium prices exist and are known (Bernstein 1996, 1998). To Keynes, speculation did not give rise to asset market stabilization, but to bouts of 'euphoria' or speculation excitement, as he called it. In this, he was right; speculation in real life is far from being asset-price stabilizing. Speculation acts as a 'super multiplier' (to use Hicks' term) or compound multiplier (as Lange put it), upon the real economy. Speculation itself does not engender the term, but it contributes to a prolongation of the boom by the euphoria it generates.

One of the crucial issues here, as Keynes' liquidity theory points out, is that decision makers do not really know and cannot really know the future outcome of the current financial decisions—the future is uncertain and not merely probabilistically risky (Davidson 2007).

What makes this analysis relevant in today's context is that it describes a process of general 'deleveraging' as part of a business downturn. Causally, in Keynesian theory, it is the decline of investment expectations and the consequent contraction of output that prompt deleveraging. Today, we are faced with the converse problem, where the deleveraging that the financial sector is rather desperately trying to carry through has driven the economy into the worst recession since the 1930s (Leijonhufvud 2009a).

As Bernstein found that since World War II, the number of stock markets around the world has grown from 50 to more than 125; even the Chinese, nominally still socialists' have seen it fit to establish stock markets in their territory (Bernstein 1998). If financial markets are, as Minsky suggests, so fragile and destabilizing, why are so many emerging economies using them. We should say, the writing might well have been on the wall. The nature of money which cannot be readily 'commodified' (Polanyi 1944), the potential perils of a Casino capitalism (Keynes 1936), the potential endogenity of financial crisis (Minsky 1982) and the potentially detrimental effects of imposition of financial interests on other interests (including those of industrialists) have all been well rehearsed (Argitis and Pitelis 2006). We are so worried for the increasing freedom of financial capital and its ability to create new, 'more' complex ways to promote its goals, could exacerbate financial crisis, impacting on the real economy.

While Kindleberger acknowledges that a speculative mania cannot be predicted, he identifies two factors that make them more likely. First, speculation is linked with positive economic expectations, in particular, in new and emerging markets and market segments. Second, ample liquidity makes hyperbolic investments more likely (Schnabl and Hoffman 2008). Both factors seem relevant for the world economy since the mid-1980s. On the one hand, the central banks in large industrial countries have tended to provide ample liquidity in response to financial turmoil and the threat of recession. On the other hand, economic prospects have been very positive in new stock, real estate and financial market segments in the industrialized countries, as well as in an increasing number of emerging market economics. In particular, the US housing markets as well as East Asia and central Europe which are heading to become new hubs of industrial production have become economic focuses (Kindleberger 1978).

The resulting new international world of finance made the exchange rate itself an object of speculation; utilizing new computer technology, financial capital could speed around the globe at the speed of light. Since the mid-1970s, international financial transactions have grown 30 times as fast as the growth in international trade (Felix 1998). A pity is that exchange rate movements reflect changes in speculative portfolio positions, rather than changes in patterns of trade. The greater the uncertainty regarding future exchange rates, the larger the investment globally, just as the Keynes study of liquidity preference and investment predicted. As a result, trade and real investment spending in open economics have become the tail, wagged by the international speculative exchange rate dog. Hence, instead of producing the utopian promises of greater stability, more rapid economic growth and full employment, claimed by classical economists, liberalization of capital flow regulations has been associated with exchange rate instability, slower global economic growth and higher global unemployment. Liberalization drove the final nail into the coffin of the post-war golden age of economic development.

The *Financial Times of London* and *The Economist* acknowledged that the system was a failure and was sold to the public and the politicians under false advertising claims (The Economist 1990). But Keynes' aphorism (Keynes 1936, p. 158) that 'worldly wisdom teaches that it is better for reputation to fail conventionally, than succeed unconventionally', again seems to rule the day. 'There is no national leader willing to challenge conventional economic analysis and call for a complete and thorough overhaul of international payments system that is far worse than the one we abandoned in 1973. Instead, there are calls for plumbing patches on the current payment systems in terms of marginal transactions tax here and/or a marginally larger lender of last resort there, and/or marginally higher capital adequacy ratios for banks as part of the package for more transparency and even inconsistent calls for Keynesian spending in Japan, while lauding fiscal budget surplus in the US and reducing government deficits in the European Union', Davidson (2000) put it so well.

6 Global Finance: Theory and Reality

The case for financial 'globalization' is based on the 'allocative efficiency' case, with the first fundamental theorem of welfare economics that (competitive markets yield Pareto Efficient equilibria), with the Efficient Market (EM) hypothesis that financial markets use information 'efficiently'. Recently, Paul Krugman argued that the Treasury view, which held that a fiscal deficit crowded out private investment and which Kahn had criticized in his famous 1931 article, makes good sense in normal times, but he regards the interest rate is being determined by monitory policy. Moreover, Krugman argues for a Keynesian fiscal stimulus in the current situation, because the advanced countries are in a liquidity 'trap' where there is no expectation of a rise in the interest rate in the foreseeable future and no question of any crowding out.

As Juliet Schor put it, 'Within the context of "global neo-classicism", financial regulations are said to disturb competitive markets and cause them to deviate from Pareto efficiency' (Schor 1992). A corollary of the case suggests that financial liberalization, the adoption of flexible exchange rate and flexible labour markets help countries to operate at or near their 'natural rate of unemployment'. Hence, financial liberalization delivers a combination of theoretical advantages and policy goals that promote specific macroeconomic targets and the efficiency of the financial sector (Grabel 1995).[1]

As Garten observed, the global economy stepped back too quickly from the brink in 1999. The crisis receded and no reforms were launched. We should underline here that the greater the uncertainty regarding future exchange rates, the lesser the investment globally, just as Keynes' analysis of liquidity preference (LP) and investment predicted (Garten 1999; Keynes 1936, p. 17). Joseph Stiglitz in his foreword to the third edition of Polanyi's (1944) classic book put the case so well, 'The advocates of neo-liberal Washington consensus emphasize that it is government interventions that are the source of the problem; the key to transformation is "getting prices right"

[1] For a critique of the fundamental theorem, see Dasgupta (1986). Also see Arestis and Demetriades (1997) and Eatwell and Taylor (2000).

and getting the government out of the economy through privatization and liberalization . . . We tell developing countries about the importance of democracy but then . . . they are told the iron laws of economics give you little or no choice, and since you (through your democratic political process), are likely to mess things up. You must cede key economic decisions, say concerning macro-economic policy, to an independent central bank, almost always dominated by representatives of the financial community and to ensure that you act in the interests of the financial community; you are told to focus exclusively on inflation—never mind jobs or growth and to make sure that you just do that you are told to impose on the central bank rules, such as expanding the money supply at a constant rate, and when one rule fails to work, as had been hoped, another rule is brought out, such as inflation targeting' (Davidson 2002; Eatwell 1996; Stiglitz 2001). We can say that Keynes' argument for the inefficiency of financial markets in an uncertain economic reality provides basic reasons for the apparent divide between theory and the economic reality, as moulded by financial globalization. Moreover, there is a possibility that international financial markets may involve an inherent element of embedded power structures and conflicts (Strange 1998).

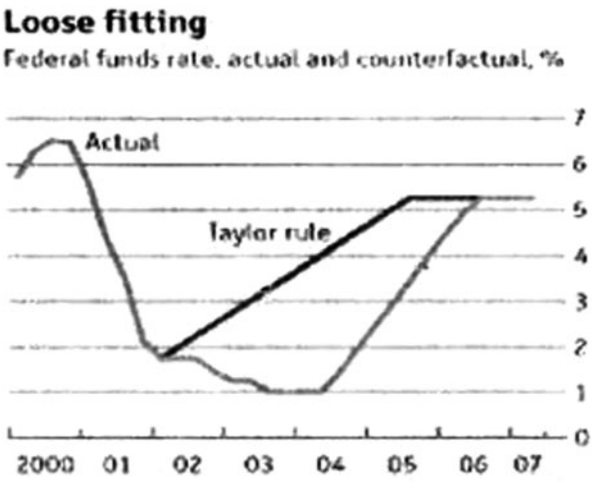
Keynes also underlined that the institutional changes that liberalized financial markets and globalized financial motives, incentives, interests and 'laissez faire' policies are likely to lead to economic and financial instability (Krishner 1999). One of the crucial issues here, as Keynes' LPT points out, is that decision makers do not really know and cannot really know the future outcome of current financial decisions; the future is uncertain. However, if decisions are taken on the basis of perfect information and/or of an unsound conceptual frame, volatility in financial markets or crisis may well be a natural result (Stiglitz 1998).

7 Causes of the Financial Crisis

The most popular explanation focuses on mistakes in monetary policy and failures of financial regulation. Also, it has been suggested that global imbalances had little or nothing to do with it, as though the existing payment imbalances would have been sustained, had there been no failure of US and UK monetary policies or financial regulations (Hutton 2009). Of course, the tightening of credit standards and the failure of cost of credit to households and businesses to fall, despite a sharp loosening of monetary policy, have led to a common view that monetary policy has not been effective during the recent financial crisis. Paul Krugman put this view in his New York Times Column that 'we are already well into the realm of what I call Depression Economics. By that, I mean state of affairs like that of the 1930s, in which the usual tools of monetary policy, above all, the Federal Reserve's ability to pump up the economy by cutting interest rates—have lost all traction'(Krugman 2008).

Mishkin also observed that since August 2007, the Federal Reserve has eased monetary policy aggressively in the face of the worst financial crisis that the USA has experienced since the Great Depression, lowering the federal funds rate target

Fig. 8.1 Chart from *The Economist*, 18 Oct 2007



from 5.25 % in September 2007 to 2.25 % in December 2008; despite substantial decline in the federal funds rate and interest rate on Treasury securities, the cost of credit to both the households and businesses has generally risen since September 2007; interest rates on riskier debt instruments have risen sharply. Baa corporate bond rates have risen by more than 200 basis points (2 percentage points) since September 2007, while interest rates on junk bonds have risen by more than 1000 basis points. Banks and other financial mediaries have also sharply tightened credit standards for both the houses and businesses (Mishkin 2009).

The above-mentioned view was recently expressed by some participants in the Federal Open Market Committee (FOMC) as the minutes from the October 28–29, 2008 meeting indicate, 'Some members were concerned that the effectiveness of cuts in the target Federal fund rates may have been diminished by the financial dislocations, suggesting that further policy action might have limited efficacy in promoting a recovery in economic growth' (Board of Governors of the Federal Reserve System 2008).

This is a comforting belief, because it implies there is nothing wrong with the larger system, we just have to learn not to make the same policy mistakes next time.

7.1 Loose Fitting Monetary Policy

The Economist in October 2007 published Fig. 8.1 as a simple way to illustrate the story of monetary excesses. This examines Federal Reserve policy decisions in terms of Federal Funds interest rate—from 2000 to 2006. This was called the Taylor rule because it is a smoothed version of interest rate one gets by plugging actual inflation and GDP into a policy rule that Taylor proposed in 1992 (Taylor 2009; Krishner 1999; Hutton 2009).

Taylor observes that the Fed cut interest rates when the rule required them to be raised, which generated a housing boom and rising levels of mortgage debts, which ended in a bust. Figure 8.1 shows that the actual interest rate decisions fell well below what historical experience would suggest policy should be. It, thus, provides an empirical measure that monetary policy was too easy during this period. This was an unusually big deviation from the Taylor rule.

Thus, the bust was made worse in September 2008, when the Treasury announced the 'Troubled Asset-Relief Program' entailing massive Government outlays, with no clear rationale for their use and no effective oversight. This sparked panic and the panic became self-generating. Everything else is just compliations. John Taylor rejects out of hand the idea that global imbalances were somehow involved. In the same tune, Alan Blinder argues that it was largely a series of avoidable—yes avoidable—human errors. Recognizing and understanding these errors will help us fix the system so that it does not malfunction so badly again and we can do so without ending capitalism as we know it (Blinder 2009a, b). Blinder's series of errors include Taylor's main cause as well as the failure to regulate over-the-counter derivatives, the decision of the Securities and Exchange Commission in 2004 to let security firms raise their leverage sharply (to an average of around 33 units of liabilities to one of assets from an earlier average of around 12 to 1), the failure to restrain the sub-prime mortgages surge, plus two more (Wade 2009b).

If global imbalances are too important to ignore, then the current policy responses nationally and internationally focused too narrowly on the 'financial system' and not enough on the 'imbalances' and what lies behind them, including polarization in national and international income distributions. Much more change will be needed to achieve a stable expanding world economy than correcting mistakes in monetary policy and financial regulations. Robert Wade is so worried that 'the global crisis will probably will become worse by the last quarter of 2010. For several more years, economic growth will remain low and unemployment high, the international monetary system will become more disrupted and the inter-state system will become more dislocated, as governments try to export their unemployment' (Wade 2009a).

8 Leverage Dynamics

Leijonhufvud put it so well, 'the system wide leverage has proven an unstable magnitude in the present regulatory regime. When leverage is increasing all around, with all the parties buying on credit, and all also find themselves making a profit. This reinforces the process. The risk is increasing, but securitization and default swaps obscure the facts. Competition makes it all but impossible to opt out of the process even for those decision-makers who perceive the rising risks. Whoever does not run with the herd, will show sub-par returns (as long as the going is good). Hence the system as a whole gets constantly more fragile until a Minsky movement arises and the process goes into reverse' (Leijonhufvud 2009b).

On the other hand, Ricardo Caballero argues that conventional wisdom is that both the bubble and the risk concentration were the result of mistakes in regulatory policy . . . and expansionary monetary policy during the boom period of the bubble and failure to reign in the practices of unscrupulous lenders. They further argue that while correct in some dimensions, this story misses two key structural factors behind the securitization process that supported the real estate boom and the corresponding leverages— (1) Over the past decade, the USA has experienced large and sustained capital inflows from foreigners seeking US assets to store value (Caballero et al. 2008) and (2) especially after the National Association of Securities Dealers Automated Quotations (NASDAQ)/Tech bubble and bust, excess world savings have looked predominantly for safe debt investments. This should not be surprising because a large amount of capital flow to the USA has been from foreign central banks and governments who are not expert investors and merely looking for a store of value (Krishnamurthy and Vissing-Jorgensen 2008).

The neoclassical theory tells us that the 'endogenous' market outcomes depend only on whether or not prices and wages are sticky (due to potential price and wage inflexibilities) and related market failure. And as far as whether money and finance can affect long-run growth, Lucas' proposition that only 'real' forces can truly affect employment and production became the only game in town (Lucas 1981). Moreover, Davidson quips, 'if anything went wrong, policy makers could suggest that they could not be blamed for, after all, the market knows best, as Nobel Prize winners Fridman, Lucas, Merton and Scholes continually assure us. The resulting new international world of finance made the "exchange rate" an object of speculation' (Davidson 2000). But Wade is firm, 'If global imbalances are too important to ignore, then the current policy responses nationally and internationally are focused too narrowly on the financial system and not enough on the imbalances and what lies behind them, including polarization in national and international income distribution. Much more change will be needed to achieve a stably expanding world economy than correcting mistakes in monetary policy and financial regulations' (Wade 2009b).

Leijonhufvud further says, 'the American banks apparently increased their leverage substantially in the years preceding the crash. The large investment banks had leverage ratios in the high 20s or low 30s. Hedge funds and some European banks may have been even more highly leveraged. At leverage ratios in this range, a loss in asset values of a couple of percentage points will suffice to make a bank insolvent. 'The estimates of financial sector losses in the present crisis, lie anywhere between a large number and unthinkably large number', told the Bank of England Director for financial stability. This statement would apply just as well to the USA and several other countries. What it means is that many banks are, in fact, insolvent. Moreover, it also explains the sudden desire among financial institutions to get rid of mark-to-market accounting, now that it works against them. The maturity mismatch between their liabilities and assets then poses an immediate danger, to the extent that they do not finance their position with insured deposits, they must constantly roll over their short-term debts, in a market where several institutions know themselves to be technically insolvent (Halden 2009).

Leijonhufvud is so worried, ‘When more or less the entire financial sector is in this situation, capital injections of the requisite magnitude in practice can come only from governments. When the sector as a whole tries to deleverage by reducing liabilities, a variety of destabilizing processes are set in motion. If many institutions try to sell the same classes of assets, the prices fall to the point, where the sale proceeds will not retire enough debt to improve leverage ratios. They may actually deteriorate. Naturally, the banks will then hold off, selling as long as possible and the markets “freeze”. The result is that a large volume of hard-to-value assets carried by highly leveraged institutions is looming over the markets. Once some banks are forced to sell, the decline in asset prices raises the capital requirements on other banks and this in turn, is reinforced when the rating agencies downgrade the assets in question. This is a highly unstable situation for the sector as a whole’ (Leijonhufvud 2009b). Moreover, the slow and gradual way for a bank to claw its way back from the brink of insolvency is to use the cash flow from the customer’s debt service to pay down their own debt. This cuts off credit to the non-bank sector. The priority that the banks have been forced to give to deleveraging explains the unavailability of ordinary trade credit in the USA for the past several quarters.

Hence, financial crises can occur for more reasons than irresponsible government deficits, moral hazard due to government guarantees, bad luck, mob psychology, crony capitalism, misguided policy or other exogenous factors, interfering with the otherwise perfectly efficient allocation of resources by financial markets.

Eichengreen et al. pity, indeed, that ‘Looking back for a quarter century, James Tobin has been almost the only voice, with significant visibility in the economic profession, warning that “free” international financial markets with flexible exchange rates can be extremely volatile and can have a devastating impact on specific industries and whole economies’ (Eichengreen et al. 1995; see also Tobin 1974).

Davidson says further, ‘Of course, significant exchange rate movements affect the international competitive position of domestic vis-à-vis foreign industries and therefore tend to depress the inducement to invest in large products, with irreversible sunk costs. In an uncertain (non-ergodic) world, where the future cannot be reliably predicted from past and present price signals, volatile exchange rates undermine entrepreneurs’ confidence in their ability to appraise the potential profitability of large investment projects.’

Even Greenspan acknowledged in his famous testimony in October 2008 to Congress that he and his free market ideology were in a state of shocked disbelief. ‘And that his real business cycle type thinking was behind his conviction that in financial markets, there are no major market failures and that the incentive of share holders to maximize their value would lead them to control the behaviour of managers and traders. Hence, he had entirely missed the possibility that financial deregulation could unleash such destructive forces on the economy. Greenspan also acknowledged that the current crisis shows that the basic premise of the traditional risk management theory is wrong and that financial markets indeed can be inherently unstable, especially due to their increasing complexity’ (Greenspan 2009).

We conclude with Palma’s terse comments, ‘Of course, Greenspan style virtual “wealth creation” meant that between 1982 and 2007, households’ net worth in the

USA increased by US$42 trillion, or by a factor of 3. Basically each had its net worth increased, on average, by US$400,000 (or by 12 times the average income of the bottom 90 %. Conversely, since the bubble blew up, the average net worth of households has declined by US$130,000 or by an amount larger than the average household debt' (Palma 2009, p. 855).

Acknowledgments For unending discussions for several days, our grateful thanks are due to our beloved Chairrman—Mr. Apoorva Verma, CMD—Firdaus Siddiqui, Director of DBS School—Dr. Ramapati Dubey and old colleagues in the University of Lucknow, India.

References

Arestis P, Demetriades P (1997) Financial development and economic growth; assessing the evidence. Econ J 107(May):783–799
Argitis G, Pitelis C (2006) Global finance, income distribution and capital accumulation, Contrib Polit Econ 25(1):63–81
Argitis G, Pitelis C (2008) Global finance and systemic instability. Contrib Polit Econ 27(1):1–11
Bellofiore R, Ferri P (2001) Introduction: 'Things fall apart, the centre cannot hold'. In: Bellofiore R, Ferri P (eds) Financial Keynesianism and market instability: the economic legacy of H. Minsky, vol 1. Edward Elgar, Cheltenham
Bernstein P (1996) Against the gods: the remarkable story of risk. Wiley, New York
Bernstein P (1998) Stock market risk in a post-Keynesian world. J Post Keynesian Econ 21(1):15–24
Blinder A (2009a) 6 Bad moves that led the US into crisis. International Herald Tribune, 26 Jan 2009
Blinder A (2009b) Talking about monetary policy: the virtues (and vices) of central bank communication. Mar 2009, BIS Working Papers No. 274. Bank for International Settlements, Basel
Board of Governors of the Federal Reserve System (2008)
Caballero RJ, Farhi E, Gourinchas P-O (2008) An equilibrium model of global imbalances and low interest rates. Am Econ Rev 98(1):358–393
Dasgupta P (1986) Positive freedom, markets and the welfare. Oxford Rev Econ Pol 2(2):25–36
Davidson P (2002) Financial markets, money and the real world. Edward Elgar, Cheltenham
Davidson P (2007) John Maynard Keynes. Palgrave, London
de Antoni E (2009) Minsky's financial instability hypothesis: the not-too-Keynesian optimism of a financial Cassandra. In: Zambelli S (ed) Comparable conservative and behavioural economic dynamics: essays in honour of Kumaraswamy (Vela) Velupillai. Routledge, Oxford
Dyer G (2009) China's dollar dilemma. Financial Times, 23 Feb 2009
Eatwell J (1996) International financial liberalization: the impact on world development. ODS Paper Series, No. 12. UNDP, New York
Eatwell J, Taylor L (2000) Global finance at risk. New Press, New York
Eichengreen B, Tobin J, Wyplosz C (1995) The case for sand in the wheels of international finance. Econ J 105:162–172
Felix D (1998) On drawing policy lessons from recent Latin American currency crisis. J Post Keynesian Econ 20(2):191–221
Frenkel R, Rapetti M (2001) A developing country view of the current global crisis: what should not be forgotten and what should be done. Camb J Econ 33(2):685–702
Garten J (1999) A Crisis without a reform. New York Times, 18 Aug 1999
Grabel I (1995) Speculation-led development: a post-Keynesian interpretation of financial liberalization programme. Int Rev Appl Econ 9(2):127–149

Greenspan A (2009) We need to better cushion against risks. Financial Times, 26 Mar 2009
Halden AG (2009) Why banks failed the stress test. Speech at Marcus Evans Conference on Stress Testing, 9 Feb 2009
Hutton W (2009) High stakes, low finance. Guardian, 2 May 2009
Keynes JM (1930/1972) The great slump of 1930. Nation and Athenaeum. In: Essays in persuasion. The collected writings of J.M. Keynes, Vol ix. Macmillan, London
Keynes JM (1936) The general theory of employment interest and money. Harcourt Brace, New York
Kindleberger C (1986) Manias, panics and crashes: a history of financial crises. Wiley, Chichester
Kindleberger CP (1978) Manias, panics and crashes: a history of financial crises. John Wiley and Sons, Hoboken
Krishnamurthy A, Vissing-Jorgensen A (2008) The aggregate demand for treasury debt. Working papers, North Western University
Krishner J (1999) Keynes, capital mobility and the crisis of embedded liberalism. Rev Int Polit Econ 6(3):313–37
Krugman P (2008) Depression economics returns. New York Times, 14 Nov 2008
Krugman P (2009) Fighting off depression. International Herald Tribune, 6 Jan 2009
Leijonhufvud A (2009a) Out of the corridor: Keynes and the crisis. Camb J Econ 33(4):741–757
Leijonhufvud A (2009b) Two systemic problems. CEPR Policy Insight, No. 29. Centre for Economic Policy Research, London
Lucas RE Jr (1981) Tobin and monetarism: a review article. J Econ Perspect 19(2):558–567
Minsky H (1977) Theory of systemic fragility. In: Altman E, Sametz A (eds) Financial crises, institutions and markets in a fragile environment. John Wiley and Sons, New York
Minsky H (1986) Stabilizing an unstable economy. Yale University Press, New Haven
Mishkin FS (2009) Is monetary policy effective during financial crises? Am Econ Rev 99(2): 573–577
Palma JG (2009) The revenge of the market on the rentiers: why neo-liberal reports of the end of history turned out to be premature. Camb J Econ 33(4):829–869
Pauly LW (2008) Financial crisis management in Europe and beyond. Contrib Polit Econ 27(1): 73–89
Reinhart C, Rogoff K (2008) Banking crises: an equal opportunity menace, Dec 2008, NBER Working Paper No. 14587. National Bureau of Economic Research, Cambridge
Schnabl G, Hoffman A (2008) Monetary policy, vagabonding liquidity and bursting bubbles in new and emerging markets—an overinvestment view. World Econ 31(9):1226–1252
Schor J (1992) Introduction. In: Baluri T, Schor J (eds) Financial openness and national autonomy. Clarendon Press, Oxford
Stiglitz J (1998) The private uses of public interests, incentives and institutions. J Econ Perspect 12:3–22
Stiglitz J (2001) Foreward. In: Polanyi K (ed) The great transformation: the political and economic origins of our time. Beacon Press, Boston
Stiglitz J (2003) The roaring nineties: a new history of the world's most prosperous decade. W.W. Norton, New York
Strange S (1998) Mad money. Manchester University Press, Manchester
Taylor JB (2009) The financial crisis and the policy response: an empirical analysis of what went wrong, Jan 2009, NBER Working Paper No. 14631. National Bureau of Economic Research, Cambridge
The Economist (6 Jan 1990)
Tobin J (1974) The new economics one decade older. The Janeway lectures on historical economics. Princeton University Press, Princeton
Tobin J (2000) Financial globalization. World Dev 28(6):1104
Wade R (2009a) From global imbalances to global reorganizations. Camb J Econ 33:539–562
Wade R (28 March 2009b) Steering out of the crisis. Econ Polit Wkly 44(13):39–46

Part II
Development

Chapter 9
Widespread Poverty Amidst High Economic Growth: Some Lessons from South Asia

G. K. Chadha

1 Some Features of South Asia

South Asia is a conglomerate of eight countries: Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka. It is a big and important region of the world. It has the seventh largest land area in the world, is the second most populous region (after East Asia and Pacific) and is tending to become the third largest global economic entity after the USA and China (Debroy 2008, p. 1; World Bank 2010b, p. 379). The Region has witnessed a fairly high economic growth in recent years but the level of real per capita income is still among the lowest in the world, primarily because of the sheer size of its population. No wonder, the Region is also a house of widespread poverty and inequalities; close to 45.0 % of the world's poor live here (World Bank 2010a, p. 92). Understandably, therefore, the Millennium Development Goals (MDGs), most ostensibly those focused on rural poverty alleviation, are especially focused on this region.

South Asia enjoys unique cultural, religious, political, linguistic and ethnic diversity that attracts worldwide attention of public analysts, academics and researchers, all over the world. Yet again, although the eight member states of the Region, through varying combinations and strategic partnerships, have been fostering many geopolitical initiatives, economic fora and trade exchanges, cultural get-togethers, developing 'friendly understandings', etc., to promote regional integration yet, the Region is not free of political strife, mutual distrust and occasional clashes; intra-regional trade is still of a fairly low order. Finally, many countries of the Region are nearly constantly in the news for a multitude of reasons: territorial disputes and political/military rivalry between countries, varying form and intensity of political rhetoric, diverse types of domestic agitations, law and order disturbances, religious conflicts and so on. All these have their own cumulatively threatening effect on regional stability which, in turn, affects the pace and pattern of growth, for individual countries as well as for the Region as a whole.

G. K. Chadha (✉)
South Asian University, Akbar Bhawan, Chanakyapuri,
New Delhi 110021, India
e-mail: gkchadha@southasianuniversity.org

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_9, © Springer India 2014

South Asia's share of world population has been increasing steadily from 21.0 % in 1990 to 23.0 % in 2008 and is projected to increase further to 24.0 % by 2015. This is in contrast to the shrinking share of population of high-income economies from 18.0 % to 16.0 % and to 15.0 % during the same period (World Bank 2010a, p. 64). Likewise, South Asia's share in Asia's population was 40.0 % in 2010 and is likely to be the same in 2020 and 2030 (UN-HABITAT 2011, p. 193). Again, its share of world labour force has been increasing steadily from 18.0 % in 1990 to 20.0 % in 2008, against a drop from 18.0 % to 16.0 % in the case of the high-income economies (World Bank 2010a, p. 68). The labour force participation rate in South Asia, 85.0 % in 1990 and 82.0 % in 2008, has been the highest among the regions while the same for female workers being 35.0 % in both 1990 and 2008 has been the second lowest (higher than only Middle East and North Africa) in the world. While it is the low level of development, especially the widespread rural economic distress, which heavily drives men to become working-hands including self-employed non-formal entrepreneurs, it is the inhibiting sociocultural milieu that keeps women away from the workforce. South Asia's low growth rate of labour productivity (gross domestic product (GDP) per person employed) of 3.1 % during 1990–1992 and 5.5 % during 2003–2005, contrasted to 6.5 % and 7.8 %, respectively, for East Asia and Pacific, lends some support to our argument (World Bank 2010a, p. 76).

The Region has its own intra-country similarities as well as heterogeneities that have been shaping the pace and pattern of its economic development, on one hand, and injecting country-specific content to its international integration, on the other. First, the population size differs starkly among the eight countries. For example, the region has a country as big as India, with its most domineering 74.0 % share in its population and around 75.0 % in GDP as also a country as small as Maldives whose share of population is a trifle 0.02 %. The eight countries clearly fall into four categories, domineering India, big Bangladesh and Pakistan, small Afghanistan, Nepal and Sri Lanka and tiny human habitats Bhutan and Maldives. Perhaps, the acute unevenness in size has stood, inter alia, in the way of fostering intra-regional economic, social, cultural and political cooperation, on an enduring basis, although there has been no dearth of intergovernmental initiatives, goodwill exchanges and sociocultural get-togethers, from time to time, most ostensibly under the South Asian Association for Regional Cooperation (SAARC) auspices. 'Trust deficit' still looms large in the Region.

Thanks to the fast-changing global economic environment under which the South Asian Region as a whole, much more than individual countries by themselves, has started foreseeing tremendous growth potential vis-à-vis the rest of the world or other regional economic groups. The blessings of regional economic cooperation are being re-assessed, in varying form and content, by intergovernmental committees, public analysts and academic researchers, and most enlightened initiatives are being put forward to foster regional consciousness among the South Asian countries. Perhaps, never before has the list of common problems, as also the potential for reaping rich dividends through the emerging common opportunities and scale advantages, stood out as clearly as in recent years. *The most common menace is a high degree of poverty, especially rural poverty, in spite of high growth rate in recent years,* *witnessed practically by each country of the Region. Highly uneven distribution of gains of economic growth between rural and urban areas and among different strata of society is another ground reality of the whole Region. A weak human capital base, surrogated, for example, by the level of education, is another Achilles' heel of the Region, with none-too-happy implications for the Region's preparedness for future growth and competitiveness in the world economic arena.* The present chapter goes into some of those problems, largely to plead that not only the recent growth momentum witnessed by the Region must be sustained, and if possible, accelerated, but common visions, understandings and strategies must also be developed to overcome the common economic and human infirmities. First, we better look at the Region's recent growth profile.

Economic growth profile of a region or a country can be sketched out in many different ways, depending upon the prime interest of the researcher which, in turn, dictates the associated aspects of growth that need to be gone into. In the present chapter, we choose to unfold the temporal profile of annual growth rate of GDP of the South Asian countries as also to see the recent structural changes occurring in these economies.

2 Growth and Structural Transformation

Table 9.1 gives a 17-year-long GDP growth profile for individual South Asian countries as also for the three subregions of Asia, including South Asia. The table throws up many interesting features and contrasts. First, on the whole, it is rather redeeming to see that South Asia has put up a fairly impressive economic growth profile for the preceding 17 odd years. Although, on a year-to-year basis, China and, by implication, the whole of East Asia, excelled every other country or country conglomerate, including the most vibrant Indian economy in South Asia, for most years during 1995 and 2011, the robust growth performance of South Asia stands out not only in Asia but also in the rest of the world. For example, the average annual growth rate of GDP during 2000–2008 was 7.4 % in South Asia, 5.2 % in sub-Saharan Africa, 4.7 % in the Middle East and North Africa, 3.9 % in Latin America and the Caribbean and a mere 2.3 % in the developed world. Only East Asia and the Pacific did better than South Asia with a growth rate of 9.1 % (World Bank 2010b, p. 379).

Second, within the South Asian Region, keeping aside sporadic setbacks of varying magnitude, the rate of growth of GDP has been fairly satisfying for most of the eight countries. Most markedly, the growth profile during the past decade or so has been satisfying in particular. For example, in Afghanistan and Bangladesh, for 9 out of 12 years during 2000–2011, the GDP growth was 6.0 % and higher; in Bhutan, it has been higher than 6.0 % for each of the 12 years; for India, it has been 6.0 % and higher for 10 years; for Sri Lanka, this level of growth performance was discernible for 8 years; for Maldives, it was so for 6 out of 12 years, and so on. It is only in Nepal, and Pakistan, that the GDP growth profile reflected a little less pleasing picture; Nepal because of low growth rates and Pakistan because of a mingle of low

Table 9.1 Growth profile of South Asian countries and other developing Asian economies. (Source: UN-ESCAP 2008; ADB 2009; UN-ESCAP 2011)

Year	Afghanistan	Bangladesh	Bhutan	India	Maldives	Nepal	Pakistan	Sri Lanka	South Asia[a]	Southeast Asia	East Asia[b]	China
1995	5.0	4.9	6.8	7.3	7.1	3.5	5.1	5.5	6.5	8.3	9.7	10.9
1996	9.0	4.6	6.4	7.8	8.8	5.3	6.6	3.8	7.1	7.6	8.5	10.0
1997	10.1	5.4	5.9	4.8	11.5	5.3	1.7	6.4	5.0	4.6	7.7	9.3
1998	11.9	5.2	6.1	6.5	9.3	2.9	3.5	4.7	4.9	−6.9	3.4	7.8
1999	−5.9	4.9	7.9	6.1	7.8	4.5	4.2	4.3	3.3	4.0	7.5	7.6
2000	−33.6	5.9	7.6	4.4	4.4	6.1	3.9	6.1	5.2	6.4	8.2	8.4
2001	−9.4	5.3	7.2	5.8	3.3	5.6	1.8	−1.4	2.3	2.1	5.6	8.3
2002	39.5	4.4	10.0	3.8	6.1	0.1	3.1	4.0	5.2	5.0	7.5	9.1
2003	14.3	5.3	7.6	8.5	9.2	4.0	4.7	5.9	7.2	5.3	7.3	10.0
2004	9.4	6.3	7.5	7.5	11.3	4.7	7.5	5.4	7.4	6.5	8.4	10.1
2005	14.5	6.0	6.1	9.4	−4.6	3.1	9.0	6.2	8.2	5.6	8.1	11.3
2006	8.2	6.6	6.4	9.7	18.0	4.1	5.8	7.7	9.0	6.1	9.4	12.7
2007	12.1	6.4	14.1	9.0	7.2	2.7	6.8	6.8	7.6	6.6	10.5	14.2
2008	3.4	6.2	11.5	6.7	5.8	5.3	4.1	6.0	4.7	4.2	6.5	9.6
2009	15.7	5.9	6.0	8.0	−2.3	4.0	1.2	3.5	3.9	1.0	4.4	9.1
2010	8.9	5.8	6.8	8.6	4.8	3.5	4.1	8.0	7.5	8.1	7.1	10.3
2011	6.8	6.4	7.2	8.7	4.0	4.0	2.8	8.0	6.8	5.5	–	9.5

[a] Growth rates for the period 1995–2005 relate to South and Southwest Asia. In other words, these include Iran and Turkey in addition to the eight South Asian countries given in the table

[b] Growth rate for the period 1995–2005 relate to East and Northeast Asia while these for 2006–2011 relate to East Asia

and high and highly fluctuating growth rates. For example, jumping up from 4.7 % in 2003 to 7.5 % in 2004 and 9.0 % in 2005, the GDP growth rate in Pakistan slipped down to 5.8 % in 2006, 4.1 % in 2008, 1.2 % in 2009 and so on. In short, Nepal and Pakistan are slightly weak partners in the club of high-growth South Asian countries.

Third, leaving aside the devastating effects of internal and external disturbances faced by specific countries, for specific years, on GDP growth, none of the eight countries of South Asia has allowed its growth process to be halted or stultified. The most telling example is Afghanistan. Except for 3 consecutive years (1999–2001), out of 17 years (1995–2011), GDP growth rate has been reigning pretty high, sometimes unbelievably high (e.g. 39.5 % in 2002, 14.5 % in 2005, 15.7 % in 2009, etc.). In any case, as a long-term phenomenon, negative growth of GDP has been nearly conspicuous by its absence in the South Asian countries in general, except for 3 years (1999–2001) in Afghanistan, 2 years (2005 and 2009) in Maldives and 1 year (2001) in Sri Lanka. Again, as a general phenomenon, the frequency of cyclical up- and downswings has been declining fairly substantially in the recent past. This indeed is a qualitative improvement in the recent GDP growth profile of the South Asian countries, compared with their own earlier performance.

Fourth, yet one more pleasing feature for the South Asian Region as a whole comes through comparing its GDP growth profile with that of Southeast Asia and East Asia. In general, for all the three regions, the rate of growth of GDP has not been deviating too much from the trend line; a relatively more steady growth regime comes out, even in the midst of occasional up- and downswings. But then, it is particularly fascinating to compare South Asia's growth performance for the 3 crisis years that shattered the economies of East Asia earlier during 1997–1999, and recently, again, under the impact of the world recession since 2008. Clearly, with its growth rate having precipitously tumbled down from 4.6 in 1997 to −6.9 in 1998, the Southeast Asian Regime lost its complete verve under the stress of financial crisis, whereas South Asia kept its growth momentum, albeit, at a slightly lower (5.0 %) level. Again, the more recent world recession pushed down the Southeast Asian growth rate from 4.2 % in 2008 to 1.0 % in 2009, while for South Asia, it turned out only to be a small drop from 4.7 % to 3.9 %. That, in sharp contrast to the situation in the developed world, most of the Asian countries, in general, and most of the South Asian countries, in particular, could neutralize the adverse effects of global recession through their own expanding domestic markets is, by now, a well accepted reality. India and China are the two bold examples from South Asia and East Asia, respectively.

To put the record straight, without trivializing the impressive growth performance of the South Asian Region, it needs, nevertheless, to be underlined that East Asia, in general, and China, in particular, have consistently been ahead of South Asia, both in terms of the year-to-year level of GDP growth rate and in terms of cyclical deviations along the long-run trends. The only South Asian country that comes close to East Asia or China, in terms of both the level of GDP growth rate as well as its stability, is India. But then, much remains still to be learnt from Chinese growth experience by India, and each of the other seven South Asian countries.

Besides high economic growth, the Region has also witnessed, in recent years, a fairly substantial degree of structural transformation. Table 9.2 clearly shows that the

Table 9.2 Structural transformation of economies in SAARC (*South Asian Association for Regional Cooperation)* and some other countries/regions: 1990–2008. (Source: World Bank 2010a; World Bank 2010b)

Country/region	Percent share of gross domestic product in								Percentage of agricultural workers (2004–2008)				Agricultural productivity per worker (US$ 2,000 prices)	
	Agriculture		Industry		Manufacturing		Services		Male		Female			
	1990	2007	1990	2007	1990	2007	1990	2007	1990–1992	2004–2008	1990–1992	2004–2008	1990–1992	2003–2005
Afghanistan	–	37	–	25	–	15	–	38	–	–	–	–	–	148
Bangladesh	30	19	22	29	13	18	48	52	54	42	85	68	254	338
India	31	18	28	29	17	17	41	53	–	64	–	77	324	392
Nepal	52	34	16	17	6	8	32	49	75	–	91	–	191	207
Pakistan	26	20	25	27	17	19	49	53	45	36	69	72	594	696
Sri Lanka	26	13	26	29	19	15	48	58	–	28	–	37	679	702
South Asia	31	18	27	29	17	17	43	53	–	–	–	–	335	406
East Asia and Pacific	25	12	40	47	30	33	35	41	–	–	–	–	295	438
Sub-Saharan Africa	20	12	34	33	17	15	47	55	–	–	–	–	263	279
Middle East and North Africa	16	11	34	43	15	12	50	46	–	–	–	–	1,583	2,204
Latin America and the Caribbean	9	7	36	33	22	18	55	60	21	20	13	9	2,125	3,055
Europe and Central Asia	13	7	36	32	22	18	51	61	–	16	–	16	1,749	2,076
Developed Countries	2	1	30	26	20	17	68	73	6	4	5	2	15,906	25,500
World	6	3	33	28	22	18	61	69	–	–	–	–	731	908

share of GDP originating in agriculture has been declining, by a varying proportion, in each country, as well as each group of countries. In South Asia, it has declined from 31.0 % in 1990 to 18.0 % in 2008. Nepal and Afghanistan's dependence on agriculture continues to be fairly high, even in 2008. Compared with the East Asian economies, and more so the Latin American economies, and still more so, the advanced industrialized economies, South Asia is still saddled with an 'agrarian tag'. The 'agrarian tag' is far more manifest in terms of the percentage of workers, most markedly, the female workers, employed in agriculture. There is, thus, a glaring hiatus between agriculture's share in GDP and its share in employment, and it poses a common problem to the South Asian economies. In other words, while the process of income generation has been steadily moving away from agriculture, the shift of workers from agriculture to non-agricultural sectors has been far more sluggish.

Perhaps, the prerequisites for staging the switchover are missing, in varying form and content, in most of the South Asian countries. Inadequacy or lack of human capital is the most convincing explanation for a big proportion of workers 'staying on in agriculture' in many of the South Asian economies while no more than 2.0–3.0 % of workers manning agriculture in the UK or Germany clearly indicates that human capital endowment of the populace there is far more stronger for moving out of agriculture (Chadha 2007, p. 191). The formidable differences in per worker agricultural productivity between the South Asian Region and the Western industrialized world, or even the Latin American and the Caribbean countries, on the one hand, and among individual countries within South Asia itself, suggest, in their own right, that human capital plays a crucial role in determining the levels of agricultural productivity (Chadha 2007, p. 192). That such differences are bound to be far more glaring in other sectors of the economy needs hardly to be emphasized. It is in this context that South Asia's stagnancy in the share of manufacturing in GDP, 17 % in 1990 stubbornly carried over to 2008, is a matter of serious common concern to the whole Region. It is a pity that even in the fast growing, and dominant, South Asian economy of India, the share of manufacturing in GDP remained frozen at 16.0–17.0 % all the way between 1990 and 2008, while, in 2008, it was as high as 34.0 % in China, 35.0 % in Thailand, 33.0 % in East Asia as a whole and 28.0 % in Malaysia (World Bank 2010a, pp. 230–232).

3 Looking Beyond Growth Statistics

By any objective reckoning, South Asia's record of economic growth, for the past about two decades, has been fairly creditable. The Region is now behind only East Asia, most markedly China, in terms of GDP growth momentum. In general, income levels have been increasing, by a varying degree, for all segments of South Asia's population which, in turn, has given rise to a burgeoning middle class, in many of the South Asian countries. That the consumption basket of households, both in rural and urban areas, is steadily moving away from food to non-food items, especially towards education, health, entertainment, etc., is a happy development.

For some countries, most notably India, industrial activities have proliferated into newer areas, reflecting an expanding cosmos of manufacturing technologies, market reaches and international collaborations. Scale upgradation is constantly at work; giant enterprises are emerging whose global presence is now catching up, *albeit* on a limited scale (Arun 2011, p. 16). A steadily increasing self-reliance, most discernibly, for industrial inputs and semi-processed products, is clearly at work. For some countries and some branches of manufacturing and services, most notably in India, the corporate sector is expanding and, with that, the club of globally recognized business tycoons and multinational companies, has been expanding, albeit slowly, in recent years. The expanding global visions and business ambitions, to-and-fro international movement of industrial capital and technology collaborations, etc. are pushing many economies of the Region into the global investment, production and trade vortices (Srinivasan 2011, p. 58–60, 71–73; Tendulkar and Bhavani 2007, pp. 125–138). Globalization is growing by the day. Many more developments, both domestic and international, can be added. All these developments naturally lead many international development analysts to foresee a global economic role for the South Asian Region, in the foreseeable future. The Region has reasons to feel proud of itself.

Besides high economic growth, the South Asian countries have also witnessed some degree of structural transformation of their economies. Most of these economies have now moved far away from being agrarian economies; non-farm sectors have been expanding, in varying form and proportion. Rural industrialization is steadily expanding; construction, trade and banking services are also catching up fast in the rural areas. Again, rural–urban connectivity is improving, by a varying degree, in country after country; and with that, rural-to-urban migration is also growing apace which, in turn, is bound to affect the pace and pattern of urban industrialization in the days ahead.

The preceding two to three paragraphs portray a cheering picture of the South Asian Region. But then, there are numerous studies to show that high economic growth in the Region has not been an unmixed blessing. A mingle of *some* cheers but *many more* despairs is the reality of the South Asian Region (Nachane 2011, p. 3–22). For example, the ongoing sectoral problems, especially those relating to agriculture (i.e. slumbering crop yields; slow, uncertain and fluctuating output growth; deforestation and depleting under-ground water resources; poverty, indebtedness and suicides among farmers, etc.) and industry (i.e. labour market rigidities and sluggish employment growth; industrial relations and wage structures, technological and marketing infirmities of the most dominant and fast proliferating tiny informal enterprises; infrastructure bottlenecks for the increasingly urban-centric industrial expansion, etc.), are well analysed, especially for India (Nagaraj 2011, p. 75–77; Ramaswamy 2011, p. 88–90; D'Souza and Bhattacherjee 2011, p. 111–112; Tendulkar and Bhavani 2007, p. 138–148). Likewise, trade-related issues, especially export sophistication and technology intensity of exports, are now engaging increased attention than has ever been the case. Environmental challenges of high growth, especially issues related to energy insecurity, depletion and pollution of water resources, land degradation and poaching, increasing solid waste and its unscientific management in urban commercial–industrial centres, deforestation and climate change, etc. pose another genre of problems, endangering the sustainability of growth itself (Sharma 2011, p. 156–159; Soz 2004, p. 231–234; Kumari 2004, p. 269–284). There is a strong opinion that the negative socio-economic fallouts of growth have eclipsed the positive gains of high growth and structural transformation of the South Asian Region. The menace of widespread poverty, increasing inequalities in the distribution of income and production assets, the unabating rural–urban socio-economic gaps, slow employment growth and deteriorating quality of employment, etc. are the major areas of concern (Sobhan 2010, pp. 14–44; Ghani 2010, pp. 20–60; Srinivasan 2011, pp. 68–70; D'Souza and Bhattacharya 2011, pp. 106–114; Motiram and Vakulabharanam 2011, pp. 59–67). The increasing incidence of corruption in public life, quite substantially occasioned by the nature and content of recent economic growth, e.g. increasing presence of private sector investment in most sectors of the economy, especially in infrastructure, very high premium *now* attached to ownership of, control over and access to diverse categories of resources such as minerals, land, information and communications technology (ICT) marketing rights, etc., makes a daily headline in some South Asian countries, most unfailingly in India (Roy 2011, p. 10).

On a different plain, the incidence of malnutrition in South Asia, most notoriously, in terms of the number of undernourished children, continues to be the highest in the world. Social progress, especially in terms of mass-based improvement in education and health, has been extremely tardy; gender inequalities in all aspects of socio-economic existence are still of very high order; rural–urban economic gaps continue to grow, most ostensibly because most of the high value-adding non-farm activities have been located in urban areas; rural to urban migration, hugely triggered by rural distress and social exploitation, still being an unbridled phenomenon, in most countries of South Asia, is threatening to choke urban economic development, on the one hand, and to whittle the tempo of growth achieved in the recent past.

We have, thus, numerous unflattering and unpalatable issues about the socio-economic realities prevailing in South Asia. The most baffling reality is the continuing prevalence of widespread poverty. Throwing bare the South Asian poverty profile immediately dispels the 'unqualified assertion' that high level of growth, by itself, makes a decisive dent on poverty; the decisive impact comes through increasing involvement of people in the growth process, and for that, the pattern and composition of growth matter far more than the mere level of growth.

3.1 Headcount Poverty in South Asia

South Asia undoubtedly reflects a highly depressing scenario on the poverty front. Table 9.3 shows the post-1981 poverty profile for the major world Regions. A few disheartening facts about South Asia come up loud and clear. First, in 1981, the Region had the second highest proportion of the world's poor living in it; East Asia

Table 9.3 Poverty estimates at international poverty lines (IPL) for major regions: 1981–2005. (Source: World Bank 2010a)

Region	IPL type	1981	1984	1987	1990	1993	1996	1999	2002	2005
Region's poor as a percentage of world's poor										
South Asia	IPL-I	28.9	30.2	33.0	31.8	31.0	35.8	34.7	38.5	43.4
	IPL-II	31.4	31.8	33.3	33.5	33.6	36.0	35.9	38.8	42.6
East Asia and Pacific	IPL-I	56.5	52.2	47.7	48.0	47.0	37.5	37.4	31.6	23.0
	IPL-II	50.3	48.7	46.8	46.0	44.6	39.5	38.4	34.1	28.4
Sub-Saharan Africa	IPL-I	11.0	13.3	15.0	16.3	17.6	21.5	22.6	24.3	28.3
	IPL-II	11.6	12.5	13.3	14.2	15.0	16.8	17.6	19.2	21.7
Middle East and North Africa	IPL-I	0.7	0.7	0.7	0.6	0.6	0.7	0.7	0.6	0.8
	IPL-II	1.8	1.7	1.8	1.6	1.7	1.9	1.8	1.8	2.0
Latin America and Caribbean	IPL-I	2.5	3.3	3.3	2.8	2.6	3.2	3.2	3.6	3.3
	IPL-II	3.5	4.2	3.8	3.5	3.4	3.8	3.9	4.1	3.7
Europe and Central Asia	IPL-I	0.4	0.3	0.3	0.5	1.2	1.3	1.4	1.4	1.2
	IPL-II	1.4	1.1	1.0	1.2	1.7	2.0	2.4	2.0	1.6
World	IPL-I	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	IPL-II	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 9.3 (continued)

Region	IPL type	1981	1984	1987	1990	1993	1996	1999	2002	2005
Proportion of the poor within each region										
South Asia	IPL-I	59.4	55.6	54.2	51.7	46.9	47.1	44.1	43.8	40.3
	IPL-II	86.5	84.8	83.9	82.7	79.7	79.9	77.2	77.1	73.9
East Asia and Pacific	IPL-I	77.7	65.5	54.2	54.7	50.8	36.0	35.5	27.6	16.8
	IPL-II	92.6	88.5	81.6	79.8	75.8	64.1	61.8	51.8	38.7
Sub-Saharan Africa	IPL-I	53.4	55.8	54.5	57.6	56.9	58.8	58.4	55.0	50.9
	IPL-II	73.8	75.5	74.0	76.1	75.9	77.9	77.6	75.6	72.9
Middle East and North Africa	IPL-I	7.9	6.1	5.7	4.3	4.1	4.1	4.2	3.6	3.6
	IPL-II	26.7	23.1	22.7	19.7	19.8	20.2	19.0	17.6	16.9
Latin America and Caribbean	IPL-I	12.9	15.3	13.7	11.3	10.1	10.9	10.9	10.7	8.2
	IPL-II	24.6	28.1	24.9	21.9	20.7	22.0	21.8	21.6	17.1
Europe and Central Asia	IPL-I	1.7	1.3	1.1	2.0	4.3	4.6	5.1	4.6	3.7
	IPL-II	8.3	6.5	6.6	6.9	10.3	11.9	14.3	12.0	8.9
World	IPL-I	51.9	46.7	41.9	41.7	39.2	34.5	33.7	30.5	25.2
	IPL-II	69.4	67.7	64.3	63.4	61.6	58.3	57.1	53.3	47.0

IPL-I: International poverty line-I is less than 2005 ppp US$ 1.25 a day
IPL-II: International poverty line-II is less than 2005 ppp US$ 2.00 a day

and Pacific had nearly one half of the world's poor against about 30.0 % of them in South Asia. Actually, the story of poverty nearly completely ends with three Regions only: East Asia and Pacific, South Asia and sub-Saharan Africa. In subsequent years, the situation improved dramatically for East Asia and Pacific while it worsened, quite sizably, for South Asia (and sub-Saharan Africa). For example, while East Asia and Pacific's share of the world's poor, in terms of purchasing power parity (PPP) US$ 1.25 a day norm (international poverty line-I (IPL-I)), witnessed a dramatic decline from 56.5 % in 1981 to 23.0 % in 2005, that of South Asia increased sizably from 28.9 % to 43.4 % during the same period. The situation seems to have gone particularly awry since the mid-1990s which, incidentally, marked the beginning of high economic growth regime in most of South Asia.

Second, there is hardly any difference when we look at South Asia's share of the world's poor, in terms of another international poverty yardstick (PPP US$ 2.00 a day). This tends to suggest that the more virulent type of poverty is also as much unequally distributed among different world regions, as the one based on milder yardstick (PPP US$ 1.25 a day). In plain terms, South Asia carries the heaviest burden of world-level poverty, irrespective of the yardstick used for measuring the same.

Third, South Asia reflects, once again, an acutely depressing scenario from the point of view of the poor among its own population. In 1981, according to IPL-I (PPP US$ 1.25 a day), 59.4 % of South Asia's population was poor. By any objective yardstick of a 'decent human existence', this looked menacingly high and yet, looking at the proportions for East Asia and Pacific (77.7 %) and sub-Saharan Africa (53.4 %), the anguish for South Asia would subside, *albeit* deceptively. But then, the post-1981 history of poverty decline differed hugely between South Asia and East Asia and the Pacific Regions. The poverty started declining in both but it did far more sharply in East Asia and the Pacific so that, towards the close of the 1980s, the proportion of the poor was the same (54.2 %) in both. From then onwards, the story of East Asia and the Pacific's victory over poverty looks highly authentic, and satisfying, while the one for South Asia reveals an extremely sluggish pace of poverty reduction. It cannot escape our notice that while in East Asia and the Pacific, the poverty declined from 54.2 % in 1987 to as low as 16.8 % even in 2005, in South Asia, it could not decline below 40.3 % in 2005 from the same level of 54.2 % in 1987. The only other Region which seems to have been competing with South Asia in 'slower pace of poverty reduction' is sub-Saharan Africa which, consequentially, continues to have no less than half of its population living below the poverty line. Clearly, in contemporary times, the two Regions which have a sizeable proportion of their population living below US$ 1.25 a day are sub-Saharan Africa and South Africa. That South Asia could not rid itself of much of its poverty, in spite of its highly impressive recent economic growth profile, is what differentiates it from the sub-Saharan situation, not at all credited with a high growth profile. In plain terms, the South Asian experience defies the existence of any automatic, or preassured, relationship between economic growth and poverty reduction. Many questions about the base from where the growth is measured, the pattern of economic growth and the institutional mechanisms under which the benefits of growth are so distributed need, therefore, to be answered.

On a broader plane, the relative neglect of agriculture, combined with a hugely urban-centric growth of industrialization and tertiary sector activities, on one hand, and the gross neglect of human capital formation, especially for the rural populations, on the other, have been held responsible, inter alia, for the unabating menace of poverty in South Asia (Sobhan 2010, pp. 4–7; Ghani and Mittal 2010, pp. 323–337). As shown in Table 9.2, a shrinking share of GDP originating in agriculture against an extremely sluggish movement of workforce from low-productivity agriculture to high-productivity non-agricultural sectors, on one hand, and very low level of agricultural productivity per worker in most of the South Asian countries, on the other, give some clue about low per capita earnings, and the widespread phenomenon of 'working but poor' in South Asia.

India typically portrays the reality. In 1993–1994, the share of agriculture, industry and services in GDP was 30.0 %, 19.4 % and 50.6 %, respectively, against 64.5 %, 14.3 % and 21.2 % share in employment (Gokarn 2011, p. 1271). The inter-sector differentials in productivity are, thus, built into the pattern of growth followed by India, and, by a varying degree, other South Asian countries. In more recent years, the asymmetry between a sector's share in GDP and that in employment grew wider, most ostensibly, because the expanding non-farm sectors, largely guided by competitive urges and productivity considerations, offered employment only on a selective basis. For example, in 2004–2005, the share of GDP originating in agriculture, industry and services changed, respectively, to 19.0 %, 20.2 % and 60.8 %, respectively, against 57.0 %, 18.2 % and 24.8 % share in employment; for 2007–2008, the figures were 16.8 %, 20.6 % and 62.6 %, respectively on the income side, and 55.9 %, 18.7 % and 25.4 %, respectively, on the employment side (Gokarn 2011, p. 1271).[1] Accordingly, the productivity differential between industry and agriculture was 4.1:1.0 in 1999–2000 and 4.2:1.00 in 2004–2005, while that between services and agriculture was 4.8:1.0 in 1999–2000 and 5.5:1.00 in 2004–2005 (Gokarn 2011, p. 1271).

3.2 Reaching MDGs

To further authenticate the slow pace of poverty reduction during the past two decades, most expressly since the mid-1990s when MDGs were set afoot, Table 9.4 clearly shows that South Asia has more disappointments than plaudits to claim in terms of meeting MDGs. For our discussion on poverty reduction in South Asia, all the MDGs need not be brought in. We choose to concentrate on seven of them, one directly dealing with poverty reduction (US$ 1.25 per day norm; see Table 9.3), and six other MDGs indirectly related to, or, reflecting on poverty, in one form or the other. A few features need to be underlined.

[1] The 2009–2010 CSO estimates for income and National Sample Survey (NSS) estimates for employment put the figures at 15.0 %, 28.0 % and 57.0 % on the income side, and 53.0 %, 22.0 % and 25.0 %, on the employment side, respectively. These figures essentially reinforce our contention of increasing per worker or per capita income gaps between agricultural and non-agricultural sectors (Mehrotra et al. 2012, p. 2).

Table 9.4 On- and off-track performance for the poverty-related MDGs (Millennium Development Goals) for South Asia and other sub-regions of Asia as on 20.02.2012 (Source: UN-ESCAP 2011)

Country/region	Description of MDGs related to poverty						
	Poverty alleviation	Safe drinking water	Basic sanitation	Under weight children	Under-5 mortality	Infant mortality	Maternal mortality
Afghanistan	–	On track	Slow	On track	Slow	Slow	Slow
Bangladesh	Slow	Slow	Slow	On track	On track	On track	Slow
Bhutan	–	Slow	Slow	Slow	On track	Slow	Early achiever
India	Slow	Early achiever	Slow	Slow	Slow	Slow	Slow
Maldives	Early achiever	Slow	Early achiever	On track	Early achiever	Early achiever	Slow
Nepal	Slow	Early achiever	Slow	Slow	On track	Slow	Slow
Pakistan	Early achiever	Slow	Slow	Slow	Slow	Slow	Slow
Sri Lanka	Early achiever	Early achiever	Early achiever	No progress	Slow	Slow	Slow
South Asia	Slow	Slow	Slow	Slow	Slow	Slow	Slow
Southeast Asia	Early achiever	On track	On track	On track	Slow	Slow	Slow
North and Central Asia	Slow	Slow	Slow	–	Slow	Slow	Slow
Asia Pacific	Early achiever	Early achiever	Slow	Slow	Slow	Slow	Slow

The progress towards achieving a millennium goal is put into one of the following four categories: Early achiever—if the country has already achieved the 2015 target; On track—if the MDG is expected to be achieved by 2015; Off track: Slow—if the country is expected to achieve the MDG but later than 2015; and Off-track: Regressing/No progress—if the country is slipping backwards or stagnating (UN-ESCAP 2012, p. 7)
MDG Millennium development goals

First, it is fairly distressing to see that for each of the chosen seven MDGs, the South Asian Region will not be able to meet the target by 2015. That this position has persisted over the past many years, or that the South Asian countries have failed to improve their performance, especially in recent years, in spite of international pressures and country commitment, internally to itself and externally to different global fora or development agencies, is a sad commentary, on the policy makers and development administrators of the Region. In comparison, Southeast Asia has a far more creditable record of performance. For example, in poverty alleviation, Southeast Asia has already achieved the 2015 target, and for safe drinking water, basic sanitation and underweight children, it is well poised to reach the same by 2015. Inasmuch as Southeast Asia, like South Asia, has also been a region to register high economic growth in recent years (in fact, for a much longer period) but, in terms of growth impact on poverty, the former stands out distinctly superior to the latter. The South Asian policy makers and public authorities have, thus, to do a lot more of strategic reorientation that cares for, and ensures, wider distributional benefits of growth, rather than chasing high growth targets, per se, whose benefits are hugely skewed in favour of some people or regions.

Second, the inter-country variations in target achievements notwithstanding, it is quite clear that each of the eight South Asian countries has more despairs and less cheers. For example, India, the biggest and one of the fastest growing countries of the Region, seems to be totally oblivious to the crucial importance of basic sanitation for the masses, or health needs of children and women. India's steadily rising graph of GDP growth in the face of failure to provide safe drinking water and basic sanitation facilities to its masses, or substantially reduce the under-5 mortality, infant mortality and maternal mortality, is a typical case of social neglect in the midst of economic growth. A similar story unfolds itself for Pakistan, Nepal and Bangladesh. Maldives is the only visible exception with some bright performance spots, although, for a different set of reasons, safe drinking water is likely to elude the country for some more years.

Third, an overview of Table 9.4 clearly gives the impression that Asian countries in general, and South Asian countries in particular, have differing approaches to poverty alleviation. That a direct attack on poverty, through expanding employment and earning opportunities (*a la* US$ 1.25 per day earning/consumption expenditure), being an inescapable prerequisite for each country, has registered limited success for close to two decades, is evident from the fact that a big majority of Asian countries would not be able to lift their masses out of poverty even in 2015. That the pledge to substantially reduce poverty, say, by one half between 1995 and 2015, would remain unredeemed, is a direct proof of the growth strategies that have bypassed masses. But then, people's expectation to get basic health and sanitation facilities when 'economies grow noisily', are not yet fulfilled; perhaps, promise for this to happen in the near future is not forthcoming either.

3.3 Poverty Mirrored Through Inflation

Table 9.5 clearly shows that in Asia, South Asia in relation to other subregions, has been a region of high inflation, and accordingly, the poor in South Asia have been suffering relatively more in terms of corrosion of the purchasing power of their earnings. For example, in South Asia, the average rate of inflation has been 17.2 % during 1999–2001, 6.9 % during 2004–2006 and 10.0 % during 2009–2011. In contrast, the three figures for Southeast Asia, a subregion with better growth performance, sustained over a much longer period, are 5.1 %, 5.7 % and 3.7 %, respectively. Likewise, in East and Northeast Asia, the three averages have been 0.2 %, 1.1 % and 1.1 %, respectively. More interestingly, inflation has been far better, and much more effectively, managed in China where it remained a low of 0.1 % during 1999–2001, 2.4 % during 2004–2006, and 2.4 % during 2009–2011, against 4.2 %, 5.0 % and as high as 10.3 %, in China, during the same three sub-periods. On the whole, it is abundantly clear that not only that the proportion of people living under poverty has been the highest in South Asia, but the poor have also been suffering more grievously under the weight of higher rate of inflation. The double-edged disadvantage of the poor in South Asia is, thus, a hugely shocking outcome of the type of growth strategy pursued; 'the working but poor' hypothesis seems to hold sway in this Region and this, by itself, is a bold pointer of the proliferation of informal and low productivity activities.

For many of the South Asian countries, the rate of inflation has been ever increasing, by varying magnitude, during the past decade or so. For example, the rate of inflation, averaged over 1999–2001, was 4.2 % in India which increased to 5.0 % during 2004–2006, and further on to 10.3 % during 2009–2011; the corresponding figures for Nepal were 5.7 %, 5.5 % and 10.6 %; for Pakistan, it increased from 4.6 % during 1999–2001 to 7.3 % during 2004–2006 and as high as 16.0 % in 2009–2011. Clearly, while the poor in South Asia as a whole have been suffering more grievously due to rising rate of inflation, in some of these, the poor have been hit harder than their counterparts in other South Asian countries. Most notably, Pakistan, India, Nepal and Bangladesh have not been able to insulate their poor from the ravages of price inflation. Corrosion in the purchasing power of the poor, widely accepted to be more debilitating and dehumanizing for the poor than for the rest of the population, has, thus, been a hard reality of the recent growth history of South Asia. While the chain of circumstances, and recent policy strategies, leading to high economic growth in South Asia are worth probing, it is equally essential to enquire why the poor were not helped, and left high and dry to live under the backbreaking effects of high inflation rates. That for most part of the Region, the history of poverty alleviation programmes is riddled with operational infirmities and corrupt practices is what the poor themselves have learnt to live with.

Table 9.5 Poverty in South Asia and other subregions of Asia mirrored through inflation. (Source: UN-ESCAP 2011)

Country/region	Annual rate of inflation											
	1999	2000	2001	Average 1999–2001	2004	2005	2006	Average 2004–2006	2009	2010	2011	Average 2009–2011
Afghanistan	–	–	–	–	13.2	12.3	5.1	10.2	−8.3	8.2	9.5	3.1
Bangladesh	7.2	2.8	1.9	4.0	5.8	6.5	7.2	6.5	6.7	7.3	7.2	7.1
Bhutan	6.8	4.0	3.4	4.7	4.5	5.3	5.0	4.9	3.0	6.1	7.5	5.5
India	4.7	4.0	3.8	4.2	3.8	4.4	6.7	5.0	12.4	11.0	7.4	10.3
Maldives	3.0	−1.2	0.7	0.8	6.4	3.3	3.5	4.4	4.0	6.0	7.2	5.7
Nepal	11.4	3.4	2.4	5.7	4.0	4.5	8.0	5.5	13.2	10.7	8.0	10.6
Pakistan	5.7	3.6	4.4	4.6	4.6	9.3	7.9	7.3	20.8	11.7	15.5	16
Sri Lanka	4.7	6.2	14.2	8.4	9.0	11.0	10.0	10	3.4	5.9	7.5	5.6
South Asia[a]	19.8	16.1	15.9	17.2	6.0	6.5	8.2	6.9	11.0	10.3	8.6	10.0
Iran	20.1	12.6	11.4	14.7	15.2	12.1	13.6	13.6	10.8	12.0	17.0	13.3
Turkey	64.9	54.9	54.4	58.0	8.6	8.2	9.6	8.8	6.3	8.6	6.0	7.0
Southeast Asia	7.9	2.3	5.0	5.1	4.2	6.1	6.7	5.7	2.3	4.0	4.8	3.7
East and Northeast Asia	−0.6	−0.1	0.1	0.2	1.6	0.8	0.9	1.1	−0.2	1.2	2.4	1.1
China	−1.4	0.4	0.7	0.1	3.9	1.8	1.5	2.4	−0.7	3.3	4.5	2.4
Pacific	1.4	4.4	4.2	3.3	2.3	2.7	3.5	2.8	1.9	2.7	3.4	2.7

Rates of inflation in this table refer to changes in the consumer price index (CPI) and reflect changes in the cost of acquiring a fixed basket of goods and services by an average consumer. Data and estimates for countries relate to fiscal years defined as follows: 2009 refers to the fiscal year spanning 21 March 2009 to 20 March 2010 in the Islamic Republic of Iran; 1 April 2009 to 31 March 2010 in India; 1 July 2008 to 30 June 2009 in Bangladesh and Pakistan; and 16 July 2008 to 15 July 2009 in Nepal. Data on India refer to the industrial workers index; data on Nepal are for urban consumers; and data on Sri Lanka are for Colombo

[a] Includes Islamic Republic of Iran and Turkey which together constitute Southwest Asia

Table 9.6 Proportion of urban population living in slum areas. (Source: UN-HABITAT 2010; UN-Human Settlements Programme 2011)

Country/region	1990	1995	2000	2005	2007
South Asia	–	–	–	–	–
Bangladesh	87.3	84.7	77.8	70.8	70.8
India	54.9	48.2	41.5	34.8	32.1
Nepal	70.6	67.3	64.0	60.7	59.4
Pakistan	51.0	49.8	48.7	47.5	47.0
Sri Lanka	–	–	–	–	10.5
East Asia and the Pacific	–	–	–	–	–
Indonesia	50.8	42.6	34.4	26.3	23.0
Philippines	54.3	50.8	47.2	43.7	42.3
Vietnam	60.5	54.6	48.8	41.3	38.3
China	43.6	40.5	37.3	32.9	31.0
Sub-Saharan Africa[a]	95.5	95.5	88.6	81.8	79.1
Middle East and North Africa[b]	50.2	39.2	28.1	17.1	17.1
Latin America and the Caribbean[c]	36.7	34.1	31.5	29.0	28.0
Europe and Central Asia[d]	23.4	20.7	17.9	15.5	14.1
Developed countries	–	–	–	–	–

Computed from country household data using the four components of slum life (i.e. availability of water, sanitation standards, housing and per person living space)

[a] Proxied by Ethiopia

[b] Proxied by Egypt

[c] Proxied by Brazil

[d] Proxied by Turkey

3.4 Poverty Mirrored Through Urban Slums

Table 9.6 reflects the menace of poverty in terms of the proportion of urban population living in slums. It is now a well-accepted fact that the poor, the deprived and the disadvantaged in the rural areas move to urban areas in the hope of getting better employment and earning opportunities. For most of them, urban slums are the natural living destinations and low-paid informal-sector jobs the earning avenues. To a large extent, therefore, the army of the urban poor consists of the rural poor, and a big proportion of the people living in urban slums are originally the migrants from the rural areas themselves. In plain terms, the proportion of the urban population living in slums and unauthorized colonies is a powerful surrogate of a country's poverty, perhaps, in its most virulent form. South Asia's record is not very happy to look at.

That, as late as 2007, no fewer than 71.0 % of urban population in Bangladesh, 59.0 % in Nepal, 47.0 % in Pakistan and nearly one third in India were living in slums, does not cover South Asia with glory. That the proportion is also high in many other developing economies (e.g. 79.0 % in Ethiopia, 42.3 % in Philippines, 38.0 % in Vietnam, etc.) does not trivialize the seriousness of South Asia's situation inasmuch as the workforce needed to sustain urban economic growth *a la* urban industrialization and proliferating service sector activities has to come, largely, from the expanding army of urban slum dwellers, and one does not have to labour hard to stress that the quality of urban workforce would not be anywhere near the standards operating in the highly urbanized and industrialized economies. Of course, much depends on the development perspectives of country governments in general, and urban development agencies in each big city.

That people in urban slums, most notoriously in many of the South Asian countries, more often than not, live under subhuman conditions, often devoid of basic civic amenities such as minimum per capita living space, safe drinking water, electricity, toilet facilities, medical care, etc., is authenticated by numerous studies. The living in these slums is internally riddled with diverse human stresses, on the one hand, and is externally subjected to severe social disapproval, as well as official abuse and ridicule, on the other. The most dehumanizing aspect is that children born and brought up in these slums get poor quality education (if they do at all), get unskilled/ low-paid jobs, and remain largely devoid of decent human qualities. The inter-generational change in occupation largely eludes them, most markedly the women folk in these settlements. Many people living in urban slums may not be poor in the conventional sense of being able to earn a predefined level of per capita income, yet, it does express poverty, perhaps more virulently, in varying other forms, which not only makes their present living a nauseating experience but also bears harsh realities for their future as well. One or two examples would add weight to our contention.

Open defecation and open urination, a national shame for any country, is a very common practice in urban slums of India. It is a shame not just from the aesthetic, human dignity and cleanliness angles but from the health angle as well. Where do these refuses end up ultimately? There is a view that rapidly modernizing India is drowning in its own excreta … And it is this ocean of excreta that our rural (and urban slum) children are being raised in. What are the health consequences of such a situation? Of the 555-million preschool children in developing countries, 32 % have stunted growth and 20 % are underweight. These two conditions together cause the death of one in every five children before they turn 5 years of age. For those alive, the long-term consequences are severe poor performance in school, dropping out, intellectual deficits and, therefore, lower economic productivity as adults (Balasubramanian 2012, p. 18).

It deserves to be underlined, *albeit* in passing, that each country of South Asia, and elsewhere in the world, has succeeded in reducing, by varying degree, the incidence of slum dwelling among its urban population, during the preceding 2 decades. It is not a trivial achievement, especially because the South Asian Region has had the distinction of growing fast during most part of these 2 decades. However, it also needs to be interpreted that fast economic growth in South Asia has not been responsible, at least in relative terms, for increasing the menace of slum dwelling. The absolute expansion of slum population is, however, a different matter.

3.5 Poverty Through Global Kaleidoscope

Table 9.7 lends some additional evidence on poverty through the huge asymmetry between South Asia's share in world population and its share in world GDP. Around

Table 9.7 Population/income growth and per capita income levels for major world regions

Region	Share of world population	Average annual growth rate of population	Average annual growth rate of GDP	Per capita gross national income (ppp terms)		Share of gross national income (ppp terms)	
	2010	2000–2008	2000–2008	2002[b]	2008[b]	2002	2008
South Asia	25.0	1.6[a] (2.1)	7.4	2,460	2,734	7.1	6.1
East Asia and the Pacific	28.6	0.8 (1.3)	9.1	4,280	5,398	16.6	15.0
Sub-Saharan Africa	11.8	2.5 (2.8)	5.2	1,700	1,991	2.4	2.3
Middle East and North Africa	5.0	1.9 (2.4)	4.7	5,670	7,308	3.5	3.4
Latin America and the Caribbean	8.4	1.2 (1.7)	3.9	6,950	10,309	7.4	8.4
Europe and Central Asia	5.9	0.1 (0.3)	6.3	6,900	12,219	6.9	7.8
Developed countries	15.3	0.7 (0.7)	2.3	28,480	37,141	55.9	57.0
World	100.0	1.2 (1.6)	3.2	7,820	10,357	100.0	100.0

In Column 3, figures in parentheses show growth rate for 1990–1995
GDP gross domestic product
[a] Countries with high population growth rate are India: 1.4, Bangladesh: 1.6, Pakistan: 2.3, Nepal: 2.0
[b] The ratio between developed countries and South Asia was 11.6:1.00 in 2002 and 13.6:1.00 in 2008 while the one between world as a whole and South Asia was 3.2:1.00 in 2002 and 3.8:1.00 in 2008

2010, the Region had a 25.0 % share in world population against about 6.0 % share in world GDP; except for sub-Saharan Africa, in none of the other world regions, the gap was as yawning as in South Asia. As has been the historical legacy of the world economic landscape, the developed world continues to command a lion's share of world GDP (57.0 % in 2008) against a fairly small share (15.0 % in 2010) in world population. Hardly surprising, therefore, that, in spite of its creditable growth performance during the recent past (7.4 % during 2000–2008), the level of per capita GDP in South Asia was (PPP) US$ 2,734 only against US$ 5,398 in East Asia and Pacific, US$ 7,308 in the Middle East and North Africa, US$ 10,309 in Latin America and the Caribbean and as high as US$ 37,141 in the developed world; it was only sub-Saharan Africa that trailed at the bottom, with a figure of US$ 1,991. Yawning gaps were in evidence in 2002 as well. To some extent, a high rate of growth of population, say, in Pakistan, Bangladesh, Nepal, etc., in South Asia, seems to be responsible for pulling down the level of per capita GDP but, in the ultimate analysis, it is the sheer size of South Asia's huge population that sets the balance.

4 Moving Out of the Tunnel

South Asia has, thus, a formidable economic agenda to pursue for conquering over poverty. There is no scope, whatsoever, for complacency on the growth front; it has to grow faster in the coming times. Then, the pattern of growth has also to change; agriculture has to be accorded its due importance, most ostensibly, because the poverty alleviation effects of higher agricultural growth are proven to be far more pronounced, and enduring, than those emanating from non-farm growth that is typically selective in offering employment and is hugely disparate in wage rates and per person earning capabilities. There are many studies, for example, in India, to show that 'aggregate GDP growth matters in poverty alleviation but agricultural growth matters more despite a sharp reduction in its contribution to GDP'; more pointedly, 'the elasticity of the headcount poverty ratio, measured at the cut-off point of US$ 1.25 (2005, ppp), with respect to agricultural value added is nearly twice as high as that of GDP (both on a per capita basis)' (Gaiha and Kulkarni 2012, p. 12). Taking cognizance of South Asia's economic realities, the most effective medium-term strategy of making a marked dent into poverty, especially rural poverty, is to ensure a high, and sustained, growth of agriculture. How best to do it?

4.1 Technology Upgradation for Agricultural Growth

What needs to be done to improve agricultural growth is well known: technology, institutions and markets need to be harmonized, and for that to happen, the rich knowledge, experience and the needed institutional expediencies already available within the Region can be readily drawn upon. For paucity of time and space, we dwell upon agricultural technology alone because, in our view, technology upgradation is the most basic need of the Region for meeting the future challenges that South Asian agriculture is likely to be saddled with; this is, however, not to suggest that many other policy areas such as agrarian and institutional reforms, marketing and price support policies, extension and input delivery systems, non-farm growth strategies, rural education, etc. are less important, and deserve less attention. What are the ground realities on technology front?

4.2 Public Expenditure on Agricultural Research and Development

The best way for looking at a country's or a region's status of agricultural technology is to look at the pace and pattern of public and private investment in agricultural research and development (R&D). Let us begin our story with public investment. Table 9.8 shows considerable north–south asymmetries, or more expressly, the technological weakness of agriculture in the developing countries, in more ways than one.

Table 9.8 Global public agricultural research intensity ratios: 1981–2000. (Source: Alston et al. 2006)

Country group	Expenditure as a percentage of agriculture GDP			Expenditure per capita (2000 International Dollars)			Expenditure per economically active member of agricultural population (2000 International Dollars)		
	1981	1991	2000	1981	1991	2000	1981	1991	2000
Developing Countries	0.52	0.50	0.53	2.1	2.3	2.7	7.0	8.3	10.2
Sub-Saharan Africa	0.84	0.79	0.72	3.1	2.7	2.3	11.2	10.5	8.2
China	0.41	0.35	0.40	1.0	1.5	2.5	2.5	3.5	6.2
Asia and Pacific	0.36	0.38	0.41	1.3	1.7	2.4	3.8	5.2	7.6
Latin America and Caribbean	0.88	0.96	1.16	5.5	6.6	5.9	45.1	50.5	60.7
Middle East and North Africa	0.61	0.54	0.66	3.2	3.6	3.7	19.2	27.3	30.2
Developed countries	1.41	2.38	2.36	10.9	13.0	11.9	316.5	528.3	691.6
Total	0.79	0.86	0.80	3.8	4.2	4.1	15.1	17.2	18.1

First, between 1981 and 2000, public expenditure on agricultural R&D as a proportion of agricultural GDP has remained nearly the same in the developing countries, while it has improved impressively in the developed world. In other words, in the developing world as a whole, a rather negligible proportion of what is being produced in agriculture is ploughed back into R&D activities through public investment, and accordingly, with every passing decade, the developed world is leaving the developing world much behind in terms of this measure of investment. *Interestingly, Asia has been doing much worse than most other Regions a la Latin America, sub-Saharan Africa and Middle East and North Africa. In all likelihood, South Asia could not have been doing any better than Asia as a whole.*

Second, per capita public expenditure on agricultural R&D has not witnessed any sizable increase in the developing world, either during the 1980s or during the 1990s, while in the developed world, it increased from US$ 11 in 1981 to US$ 13 in 1991 and stood at US$ 12 in 2000. From the point of view of this research intensity ratio too, the developing countries continue to be way behind their developed counterparts. *Here also, Latin America, Middle East and North Africa, and, to a slightly lesser extent, sub-Saharan Africa, have been ahead of Asia* (and China), most plausibly because of the sheer size of the Chinese and the Indian economies. *That the situation has been no better in South Asia is, again, to be taken for granted.*

Third, research expenditure per economically active member of agricultural population shows the north–south gaps far more tellingly. In 1981, public expenditure for every active member of agricultural population was US$ 316.5 in the developed countries against US$ 7.0 in the developing countries; in 1991, it was US$ 528.3

Table 9.9 Agricultural R&D (research and development) investment expenditure in the developed and developing countries: 2000. (Source: Alston et al. 2006)

Country group	Expenditure (million international US$)			Expenditure per capita (million international US$)			Expenditure/economically active member of agricultural population Expenditure per capita (million international US$)		
	Public	Private	Total	Public	Private	Total	Public	Private	Total
Developing countries	12,909 (91.6)	1,180 (8.4)	14,089 (100.0)	2.70	0.25	2.95	10.25	0.88	11.13
Developed countries	10,191 (44.8)	12,577 (55.2)	22,767 (100.0)	11.90	14.68	26.58	691.60	853.26	1,544.86
Total	23,100 (62.7)	13,756 (37.3)	36,856 (100.0)	4.10	2.44	6.54	18.10	10.78	28.88

The figures in parentheses are the respective shares of public and private investments

against US$ 8.3, and in 2000, US$ 691.6 against US$ 10.2. Putting the asymmetry differently, the ratio between the developed and the developing countries was a high of 45.0:1.0 in 1981, went up to 64.0:1.0 in 1991 and further on to 68.0:1.0 in 2000. These figures are good enough to show how public investment expenditure is woefully inadequate in the developing countries from the point of view of the productivity, income and welfare of the people actively engaged in agriculture and allied activities.

That the situation has been particularly despairing in Asia (including South Asia) becomes evident if we look at the ratio between Asia and the developed world. It was 83.0:1.0 in 1981, 102.0:1.0 in 1990 and 91.0:1.0 in 2000. That Asia has been figuring poorly, even among the developing countries, becomes evident from the ratio between, say, Latin America and Asia being 12.0:1.0 in 1981, 14.4:1.0 in 1991 and 8.0:1.0 in 2000. Most unexpectedly, South Asia (surrogated here by Asia) has not been doing too well in comparison to sub-Saharan Africa either.

4.3 Total Expenditure on Agricultural R&D

Table 9.9 carries the public sector R&D investment story further. It shows private and public sector components of investment expenditure on agricultural R&D, for the developed and the developing countries, as also for some Regions within the latter group. It throws bare many discomforting facts for the developing world in general, and South Asia (surrogated by Asia), in particular.

First, among the developing economies, agricultural R&D is largely a public sector concern; in 2000, no less than 92 % of total R&D expenditure came through government spending. On the other hand, in the developed world, private corporate sector plays a substantial role; as much as 55.0 % of total expenditure on agricultural

R&D is contributed by the private sector. The worldwide agricultural R&D scenario, thus, clearly points towards strategic dichotomies between the developed world where, to a considerable extent, market determines the pace and pattern of research, and the developing world where the state assumes an overwhelming responsibility for generation and dissemination of research output. Clearly, the international market for agricultural technology, and all its associated components/ingredients would put the developing world to a huge competitive disadvantage. Markets do not always operate for the needy and the disadvantaged.

Second, the worldwide total (public plus private) expenditure on agricultural R&D is hugely unevenly distributed between the developed and the developing countries. For example, the developing world has only a 38 % share of total investment expenditure, while, on the basis of its share of rural or agricultural population, or livelihood stakes, or the incidence of rural poverty, a much higher share should accrue to it. In relative terms, the developing world is, thus, acutely underfunded. An aspect of crucial significance in these days of globalization and privatization is that private investment in agricultural R&D is overwhelmingly located in the developed world; the developing countries have a meagre 9 % share in total private R&D expenditure. This fact alone has the potential of creating a big dichotomy between the needs of millions of smallholders, and other marginalized sections in rural areas of the developing countries, and the business interests, profit motive and 'getting the prices right attitude' of a few global or sub-global level corporate firms. Perhaps, it is pertinent to keep in mind that the developing countries' small (9 %) share of the global-level private expenditure is most unequally distributed among individual countries; only such of the developing countries as have a minimum acceptable framework of intellectual property rights in place and a well-oiled rural connectivity network will count; India qualifies to be one such country.

Third, the developing countries pale into sheer insignificance if we measure R&D investment expenditure on per capita basis. In respect of private expenditure, the developing world stays pathetically behind the developed world; for every person, the developed countries are spending nearly 60 times as much as the developing countries are doing. Largely because of such private expenditure gaps, the per capita gaps at the level of total expenditure also look rather frightening; the ratio is 9:1.

Finally, the most telling differences between the two groups of countries are discernible when we measure them in terms of expenditure per economically active member of agricultural population. On an average, for every economically active member of agricultural population, public expenditure on agricultural R&D by the developed countries is 67 times of that by the developing countries, private expenditure is shockingly 970 times as much higher and total expenditure is no less than 140 times as much higher.

The foregoing analysis unquestionably prompts us to conclude that agricultural R&D for less developed countries in general, and South Asia (here surrogated by Asia) is at a crossroads. The recent years have witnessed changing policy contexts, fundamental shifts in the scientific basis for agricultural R&D, and shifting funding patterns for agricultural research in rich countries. The agenda in richer countries is shifting away from areas like yield improvement in major food crops to other crop

Table 9.10 A synoptic view of access to education by principal regions. (Source: UNDP 2010)

Region	Enrolment ratio (2001–2009)				
	Primary level[a]		Secondary level[b]		Tertiary level[c]
	Gross	Net	Gross	Net	Gross
Developed countries					
OECD	101.7	95.6	101.1	91.8	71.4
Non-OECD	108.4	95.6	93.6	86.7	43.0
Developing countries					
Arab States	96.4	80.9	68.8	60.4	22.7
East Asia and the Pacific (China)	112.2 (112.1)	93.3	72.8 (74.0)	62.6	20.9 (22.1)
Europe and Central Asia	98.5	92.3	89.3	82.1	54.2
Latin America and the Caribbean	116.5	94.4	89.8	72.5	36.7
South Asia (India)	108.2 (113.1)	86.9 (89.8)	53.5 (57.0)	42.0	12.8 (13.5)
Sub-Saharan Africa	101.8	73.6	34.4	29.5	5.5
World	106.9	86.1	66.4	60.2	25.7

OECD Organisation for Economic Co-operation and Development

[a] Percentage of primary school age population

[b] Percentage of secondary school age population

[c] Percentage of tertiary school age population

characteristics and even to non-agricultural production frontiers such as nutrition and environment. These changes imply a need to rethink, on one hand, national policies in less developed countries and reconsider, on the other, multinational approaches in order to determine what types of activities to conduct through global research networks and how to organize and finance them. There is sufficient ground as well as tremendous scope for the bigger countries of South Asia to cooperate amongst themselves, as also to foster links with international R&D agencies.

In the long run, based upon improvements in human capital base of the Region, and numerous other policy interventions, a gradual shift of rural people out of agriculture would be the final assault on poverty (Ozden and Sewadeh 2010, p. 312–313). That a large ground is yet to be covered, inter alia, on the front of human capital formation for making such shifts possible and economically meaningful, is cryptically conveyed by Table 9.10. South Asia is a lame duck even in this matter. The Region is particularly vulnerable in tertiary education; except for sub-Saharan Africa, it is much behind most other developing regions, and far too behind the developed world. Again, except for sub-Saharan Africa, it is not too good for secondary level of schooling either.

References

Alston JM et al (2006) Introduction and overview. In Pardey PG et al (eds) Agricultural R&D in the developing world: too little, too late? IFPRI, Washington, pp. 21–22

Arun TK (2011) Why didn't the dog bark? The Economic Times, 28 July, 2011

Asian Development Bank (2009) Asian development outlook 2009 update: broadening openness for a resilient Asia. ADB, Mandaluyong
Balasubramanian D (2012) The science behind total sanitation campaign. The Hindu, 21 July, 2012
Chadha GK (2007) Human capital base of SAARC economies: does it assure international competitiveness? In Saradgi IA et al (eds) SAARC: the road ahead, foundation for peace and sustainable development
Debroy V (2008) India's attempts at regional integration with South Asia and East Asia, Paper Read at 2008 EABER-NCAER Conference on micro-economic foundations of economic policy performance in Asia, The Imperial Hotel, New Delhi, 3–4 April 2008
D'Souza E, Bhattacharya D (2011) Employment and industrial relations in India: informality and inequality. In Nachane DM (ed) India Development Report 2011, Oxford University Press, New Delhi, 2011
Gaiha R, Kulkarni V (2012) Politics, paralysis on poverty. The Economic Times, 21 July, 2012
Ghani E, Mittal S (2010) Stimulating agricultural growth. In Ghani E (ed) The poor half billion in South Asia: what is holding back lagging regions, Oxford University Press, New Delhi
Gokarn S (2011) Economic reforms for sustainable growth. RBI Monthly Bulletin, August, 2011
Kumari J (2004) South Asia's degraded landscapes: grave challenges to environmental sustainability. In Soz S et al (eds) Towards South Asian economic integration and unity, Foundation for peace and sustainable development, New Delhi, 2004
Mehrotra S et al (2012) Creating employment in the 12th five year plan. IAMR Occasional Paper No. 3/2012, Planning Commission, Govt. of India, May 2012
Motiram S, Vakulabharanam V (2011) Poverty and inequality in the age of economic liberalization. In Nachane DM (ed) India Development Report 2011, Oxford University Press, New Delhi
Nachane DM (2011) Overview—two decades of structural reforms in India: a balance sheet. In Nachane DM (ed) India Development Report 2011, Oxford University Press, New Delhi, 2011
Nagaraj R (2011) Industrial performance, 1991–2008: a review. In Nachane DM (ed) India Development Report 2011, Oxford University Press, 2011
Ozden C, Sewadeh M (2010) How important is migration? In Ghani E (ed) The poor half billion in South Asia: what is holding back lagging regions. Oxford University Press, New Delhi
Ramaswamy KV (2011) Regional disparities in manufacturing growth in India. In Nachane DM (ed) India Development Report 2011, Oxford University Press, New Delhi
Roy A (2011) I'd rather not be Anna. The Hindu, 22 August, 2011
Sharma VK (2011) Environmental challenges of development strategies in India. In Nachane DM (ed) India Development Report 2011. Oxford University Press, New Delhi
Sobhan R (2010) Challenging the injustice of poverty: agendas for inclusive development in South Asia. Sage Publications India Pvt Ltd, New Delhi
Soz S (2004) Human development and environmental degradation in South Asia. In Soz S, Sahni SK, Srivastava RN (eds) Towards South Asian economic integration and unity. Foundation for peace and sustainable development. New Delhi
Srinivasan TN (2011) Growth, sustainability, and India's economic reforms. Oxford University Press, New Delhi
Tendulkar SD, Bhavani TA (2007) Understanding reforms: Post-1991 India, Oxford University Press, New Delhi
UNDP (2010) Human Development Report 2010—The real wealth of nations: pathways to human development. 20th Anniversary Edition. UNDP, 2010, p. 196
UN-ESCAP (2008) Economic and social survey of Asia and the Pacific 2008: sustaining growth and sharing prosperity. UN-ESCAP, February 2008, p. 158
UN-ESCAP (2011) Economic and social survey of Asia and the Pacific 2011: sustaining dynamism and inclusive development. UN-ESCAP, April 2011, p. 51–104, 202, 205
UN-ESCAP (2012) Accelerating equitable achievement of the MDGs: closing gaps in health and nutrition outcomes. Asian Development Bank–UNESCAP, Bangkok, February 2012
UN-HABITAT (2011) Global report on human settlements. Cities and climate change. UN–Human Settlements Programme, 2011, p. 207–209

World Bank (2009a) World development report 2009—reshaping economic geography, World Bank, 2009

World Bank (2009b) World development indicators. World Bank, April 2009a

World Bank (2010a) World development indicators. World Bank, April 2010a, pp. 230–232

World Bank (2010b) World development report 2010: development and climate change. World Bank, 2010b, pp. 379, 384, 385

Chapter 10
Development Dividend of Peace: Experience of South Asia

Sucha Singh Gill

1 Introduction

South Asia comprises countries which largely fall in the Indian subcontinent. This region has common geography, history and culture. The countries in this region have a common resource base and common problems. Though there is large diversity, there are several common issues between neighbouring countries. India is the largest country of the region in terms of both geographic area and population and has a common border with all countries of this region except Afghanistan. In the post-Second World War era, which is also known as post-colonial phase, the countries of this region have faced conflicts with each other and also within themselves, thereby affecting the course and trajectory of their development. These conflicts not only shaped their history but also the pace and nature of socio-economic development. Pakistan, India and Bangladesh were part of the British India Empire, which experienced partition in 1947 into India and Pakistan. Subsequently, in 1971, East Pakistan was separated from Pakistan, and Bangladesh, as a country, came into existence after a bitter conflict within Pakistan and a war with India. India and Pakistan have experienced four wars (1948, 1965, 1971 and 1999). Though other countries have had some clash of interests with their neighbours, yet they have been free from armed conflicts. Afghanistan is one of the countries which has had a prolonged period of conflicts involving global powers, earlier the Soviet Union and currently the USA The countries of this region joined hands with each other to create the South Asia Association of Regional Cooperation (SAARC) in 1985. In April 1993, the SAARC countries agreed to set up the South Asia Preferential Trading Agreement (SAPTA) by 1997. Afghanistan joined SAARC in 2005; and since 2006, these countries have launched South Asian Free Trade Area (SAFTA). Thus, these countries have decided to join each other within the framework of regional economic integration. There are several hurdles in the way of economic integration to realize the full potentials of the

S. S. Gill (✉)
Centre for Research in Rural and Industrial Development (CRRID),
Sector 19-A, Madhya Marg, Chandigarh 160019, India
e-mail: dgcrrid@yahoo.in, gsuchasingh@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_10, © Springer India 2014

regional trading bloc and to convert it into a common economic bloc, yet there are considerable achievements in this direction. The countries in this region have committed themselves to strengthen regional cooperation and are making efforts, though at different paces, to move closer. In spite of the internal and external pressures, the process of regional integration is making progress, although at a slow pace. The international process of globalization under the World Trade Organization (WTO) has created a favourable climate as the countries in the region have joined it. At the same time, this region along with China has enabled their economies to grow very fast. The dynamism of development has been pushing forward the process of regional economic integration. The recent phenomenon of impending climate change has further created an environment of joint efforts among countries to meet the challenges of climate change, food security, malnutrition, illiteracy, unemployment and ill health. The present chapter is an attempt to understand the long-term dynamism of development in relation to peace and security in South Asia. This is organized into five sections. Section 1 introduces the theme, and Sect. 2 relates the process of economic development with peace and security. Section 3 attempts at the historical experience of this region over the long period of colonial domination. Section 4 makes an attempt to analyse the post-colonial experience of peace and development among SAARC countries. The last section provides a summary and attempts at the framework for peace, security and development in this region.

2 Economic Development and Security

There is a close relation between peace and security and economic development in a country or group of countries. The absence of peace and security affects the process of economic development negatively. This is the reason that Adam Smith thought that defence is more important than opulence. In his opinion, opulence or prosperity of a nation cannot be saved without defence. Smith (1776, p. 570) wrote:

> That wealth, at the same time, which always follows the improvements of agriculture and manufactures and which, in reality, is no more than the accumulated produce of those improvements, provokes the invasion of all their neighbours . . . a wealthy nation is of all nations the most likely to be attacked; unless the state takes some new measures for public defence, the natural habits of the people render them altogether incapable of defending themselves.

Smith believes that division of labour in the economy leads to specialization. This makes individuals promote their private self-interest by confining themselves to a particular trade. They have no interest and time left for the protection of the country. This task has to be performed by the state by maintaining and employing a certain number of citizens in the constant practice of military exercises. The profession of a soldier has to be specialized and separate as well as distinct from all others. Military exercises have to be the principal occupation of soldiers of a standing army. The maintenance or payment to the army is made by the state from its own funds. The state justifies the taxes on other trades for maintenance of an army to protect the country from external invasion and internal violence. The task of defending the society from violence and injustice of other independent societies gradually grows more expensive as the society advances in civilization. The changes in war technology and a variety of weapons, which have come into existence, have enhanced the expense of armies both in peace and in war. Along with defending the country from external violence, the protection of property within the country is an equally important task of the state. The protection of life and property from external invasion and internal violence is the main task of the state in Smith's framework of economic progress or wealth of nations. The successful protection of life and property in a country defines prevalence of peace and security. The absence of threat to peace and the existence of security exercise positive effect on variables/factors affecting economic progress.

There is an often-repeated argument for improving productivity of agriculture, which emanates from improving the security of tenants engaged in agricultural production. If a tenant has long-term tenure of a land, he may find it in his own interest to lay out part of the capital in the further improvement of the land/farm because he expects to recover it, with a large profit, before the expiration of his lease. But if the land ownership right is given to the cultivator, he develops long-term interest in the improvement of productiveness of the land. Thus, the security of tenure for cultivators and security of property in trade and manufactures provide incentives for capital formation to increase productivity, introduce new technology or new practices in production. The existence of peace at a particular location creates confidence/security in the mind of investors to make investments at such places. Thus, there is high and positive relationship between peace and security and investment at specific locations. Where there is breach of peace and threat to life and property, capital flight takes place from such locations. The emergence of disturbances and insecurity also drive away skilled and trained workforce to secure places. The investment of capital takes place in the hope of better rewards/return in the future. The absence of peace generates insecurity of future returns and, consequently, drives away capital as well as talented men and women. Capital investments as well as entrepreneurs are very sensitive to peace and security. The recent growth theory (Lucas 1988, pp. 3–42; Romer 1990, pp. S71–S102) considers human capital as the main driver of growth. It is through human capital that the law of diminishing return can be thwarted. Human capital formation opens up the possibility of a sustainable and a higher level of development. The progress of human capital demands universal and quality schooling followed by high and growing enrolment in higher education. This demands a higher proportion of public spending on education and health and also high priority of education and health in family budgets/household expenditure. To achieve goals in human capital formation, peace and tranquillity in the country is the necessary condition. In conditions of violence and insecurity, the governments give high priority to expenditure on achievement of peace and maintenance of law and order rather than to promotion of education and public health. Similarly, at household level in the areas infested with violence, the strategy adopted is to protect mainly family members physically and save property/physical capital. Many a time, violence leads to the shifting of family members and businesses to safer places to avoid loss of life and property. The lack of peace and security, thus, leads to relative neglect of human as well as physical capital formation. The absence of peace and security in a country not only proves detrimental for economic progress through destruction of property and killing of human resources in violence, but it also has a long-term impact on the psychology of people, negatively affecting the savings and investment for the creation of physical capital. The incentives for such processes become weak, and the lack of peace and security also lead to capital flight to secure places. Similarly, the trained/skilled workforce as well as entrepreneurs also leave the places mired by violence to safer locations. The prevalence of peace and security becomes one of the necessary conditions for economic and social development.

The absence of security especially in a situation of conflict imposes a cost on the society. The persons engaged in managing security or those who join the fighting forces can no longer work productively. Conflicts always result in the withdrawal of some persons engaged in productive activities to be deployed in the task of conflict management and restoration of peace. This causes loss of production in the economy. Even if productively engaged persons are not drafted towards conflict management, the heightened efforts in this direction definitely involve a shift of some productive resources towards that. This includes rail, road and air vehicles for transport of men and material for war efforts or activities related to war. The diversion of key inputs like oil and energy takes a toll on some productive activities. Besides, conflicts and wars lead to destruction of schools, hospitals, power stations, bridges, roads and railway lines, airports, etc. and reduce productive capacity of the economy. There is a complex relationship between war efforts and economic development through macro fundamentals to micro behaviour of economic actors and individual behaviour of households and production units. The war efforts impose a series of constraints such as reduction of export earnings leading to reduced import capacity for essential materials. At the same time, destruction of schools and hospitals or their closure leads to adverse changes in individual entitlements to education and health care. While the constraints of materials lead to reduction in employment, the adverse changes in entitlements lead to reduced capacity in earnings and wages. The latter could adversely affect the whole generation of people. The cross-country estimates suggest that economies in conflict on an average grow 1–2 % more slowly than peacetime economies. Agriculture is hit hard as people are forced to shift in the course of conflict and exports are negatively affected by a general fall in production, a shift towards domestic markets and disruptions of international markets (Chatterjee 2011, pp. 105–106). This also leads to slow growth/decline in government revenue, diversion of the public expenditure towards war efforts and reduced spending on social sectors. Where destruction of physical capital is involved, there is decline in new investment especially foreign and private investment. All these factors put a heavy cost of war on the economy and its pace of development. The actual cost depends on the spread of the conflict zone in the country and time period of such conflicts. The larger the area covered and longer the time period involved, the higher will be the cost and greater will be the adverse effect on the economy. The countries, which experience a long spell of peace and security grow fast and become economically strong.

With economic development, modern nations acquire the capacity to employ a trained and disciplined army and acquire and produce sophisticated and high-technology weapons, which are products of specialized modern industries based on science and technology. The application of science and technology to manufacturing became possible after the Industrial Revolution in Europe. The development of modern industrial enterprises produced a variety of products. On the one side, production of manufactured output increased manifold, which made modern economies synonymous with industrial economies. Increased scale of production and enhanced efficiency of manufacturing increased the capacity of industrialized nations to spend more on regular armies to protect themselves from external invasions and internal strife. At the same time, the Industrial Revolution contributed massively on invention of modern weapons ranging from ground war force to air and space war. Kennedy (1989), on the basis of nearly 500 years' history of empires in the world, writes that the relative strength of nations have varied in the world affairs because of uneven rate of growth among them and of the technological and organizational breakthroughs, which bring a greater advantage to one society than to another. The society with greater advantage gets higher economic capacity than others including superiority in warfare technology and weapons. Better economic performance of a nation enables its government to share the prosperity with the sections lagging behind. This enables the government to gain more legitimacy within the country to restore peace and security. The discussion in this section can be summed up by saying that establishment of peace and security in the country creates conditions for economic development, which, in turn, develop the capacity of the economy to afford better organization of defence against external aggression and internal strife. Thus, without proper defence, prosperity cannot be protected, and without economic prosperity, proper defence of the country cannot be organized in modern societies.

3 Colonial Dominance and Underdevelopment

South Asia had one of the world's oldest civilization in the Indus Valley. This region experienced considerable economic prosperity compared to other regions of the world in the ancient and medieval periods. In fact, economic prosperity based on settled agriculture and advanced handicrafts-based industrial production attracted a large number of invaders earlier from central and West Asia and later from western Europe, and also a large number of merchants and traders. In fact, the absence of effective defence contributed to the decline of prosperity of the region. The potentials of further economic progress could not be realized due to plundering by the invaders and ultimately subjugation of the whole region by the British in the eighteenth and nineteenth centuries. Though the Industrial Revolution began earlier in Europe, the major changes in manufacturing occurred around 1780. This coincided with the era of colonialism in South Asia. By 1757, the East India Company, after the Battle of Plassey, had acquired the right to rule in eastern India. After this, its rule expanded to the rest of the subcontinent and by 1849, when Punjab was annexed, the British rule in India was complete in the subcontinent. The Industrial Revolution brought massive increase in productivity in industries by replacing animate power with inanimate power based on steam engine using coal, a new source

Table 10.1 Relative shares of different countries and regions in total world manufacturing output (in percentage; triennial average). (Source: Bairoch 1982, Table 9 (p. 294) and Table 12 (p. 302))

Country/Region	Years					
	1750	1800	1860	1900	1938	1953
A. Developed countries	27.0	32.3	63.4	89.0	92.8	93.5
B. Europe	23.2	28.1	53.2	62.0	53.6	42.1
Belgium	0.3	0.5	1.4	1.7	1.1	0.8
France	4.0	4.2	7.9	6.8	4.4	3.2
Germany	2.9	3.5	4.9	13.2	12.7	5.9
Italy	2.4	2.5	2.5	2.5	2.8	2.3
Russia	5.0	5.6	7.0	8.8	9.0	10.7
Spain	1.2	1.5	1.8	1.6	0.8	0.7
Sweden	0.3	0.3	0.3	0.9	1.2	0.9
UK	1.9	4.9	19.9	18.5	10.7	8.4
USA	0.1	0.8	7.2	23.6	31.4	44.7
C. Third world countries	73.0	67.7	36.6	11.0	7.2	6.5
China	32.87	33.3	19.7	6.2	3.1	3.5
India (including Pakistan)	24.5	19.7	8.6	1.7	2.4	1.7

of energy. The use of steam engine in transport increased the speed and capacity to transport men and materials several fold. The continuous growth of the economy and knowledge created a regular flow of investment and innovations producing unprecedented dynamism of the economy. Although the Industrial Revolution spread to western Europe and North America, it did not reach South Asia. Rather, the Indian subcontinent experienced deindustrialization during the nineteenth century (Gadgil 1971, pp. 33–46). The Industrial Revolution not only strengthened the economies of European countries by accelerating their pace of development but also increased their military strength. The Industrial Revolution produced new technology, which made an impact on raising the strength of military and naval forces. It raised firing power and increased the mobility and speed of movement of troops through railways and motorized ships. This enabled them to bring countries in Asia, Africa and Latin America under their political control. The colonies were plundered by the major industrial powers to boost capital formation leading to consolidation of their economic development and underdevelopment of the colonies. This led to a shift in economic balance in the world in favour of the industrial economies and against the non-industrialized countries (Amin 1974, pp. 169–299; Frank 1975, pp. 1–115).

In a major work, Bairoch (1982, pp. 269–333) brought out that due to the uneven impact of the Industrial Revolution, there was a decline in the Chinese, Indian and other non-European economies and a rise in European economies (including North America). The data (Table 10.1) show that the share of Europe as a whole in world manufacturing output was 23.2 % in 1750. In this year, the share of China and India in world manufacturing output stood at 32.8 % and 24.5 %, respectively. The share of the Third World as a whole in world manufacturing output was 73.0 %. No doubt, the world economy at that time was non-industrial or agrarian in nature, characterized by poverty, low productivity and low output per head. In that sense, there was not much difference in the living standards and levels of development among different

Table 10.2 Per capita levels of industrialization (UK in 1900 = 100) between 1750 and 1953. (Source: Bairoch 1982, Table 9 (p. 294) and Table 12 (p. 302))

Country/Region	Years					
	1750	1800	1860	1900	1938	1953
A. Developed countries	8	8	16	35	81	135
B. Europe	8	8	17	33	94	107
Belgium	9	10	28	56	89	117
France	9	9	20	39	73	95
Germany	8	8	15	52	128	144
Italy	8	8	10	17	44	61
Russia	6	6	8	15	20	38
Spain	7	7	11	19	23	31
Sweden	7	8	15	41	135	163
UK	10	16	64	100	157	210
USA	4	9	21	69	167	354
C. Third world countries	7	6	4	2	3	4
China	8	6	4	3	4	5
India	7	6	3	1	4	6
World	7	6	7	14	31	48

countries in the world. The uneven impact of the Industrial Revolution accompanied by colonialism led to a relative and an absolute decline of China, India and non-European economies and a steep rise of European countries and the USA. The share of Europe, as a whole, rose slowly in the eighteenth century but drastically in the nineteenth century. This share rose from 23.2 % of the world manufacturing output in 1750 to 28.1 % in 1800, 34.2 % in 1830, 53.2 % in 1860, 61.3 % in 1880 and 62.0 % in 1900. The share of UK alone rose from 1.9 % in 1750 to 22.9 % in 1880 and fell to 18.5 % in 1900. Similarly, the share of other European countries such as France, Germany, Italy, Russia and the Habsburg Empire also rose substantially. But there was a dramatic rise in the share of the USA from 0.1 % of the world manufacturing output in 1750 to 7.2 % in 1860, 14.7 % in 1880 and 23.6 % in 1900. Correspondingly, the share of the Third World as a whole dropped from 73.0 % of the world manufacturing output in 1750 to 36.6 % in 1860 to 20.9 % in 1880 and 11.0 % in 1900. The share of China in the world manufacturing output fell from 32.8 % in 1750 to 6.2 % in 1900 and that of India from 24.5 to 1.7 % in 1900.

The industrialized countries gained not only in terms of their share in the world manufacturing output but also in terms of per capita levels of industrialization. Taking the UK's level of 1900 as 100, this has been worked out by Bairoch (1982). As given in Table 10.2, data show that the level of per capita industrialization in Europe as a whole rose from 8 in 1750 to 33 in 1900. UK's rise was from 10 in 1750 to 100 in 1900, much faster than in other European countries. But all of them gained at various levels from double to more than four times. The raise experienced by the USA was more than 17 times. The per capita level of industrialization of the USA increased from 4 in 1750 to 69 in 1900. The Third World as a whole lost in a big way, 7 in 1750 to 2 in 1900. China's level fell from 8 in 1750 to 3 in 1900. The loss of India was much more dramatic from 7 in 1750 to just 1 in 1900. The relative decline of China and India is explained by massive increase in industrial productivity in the Western countries arising from the Industrial Revolution. In the UK, the mechanization of spinning increased productivity in that sector alone by a factor of 300–400. The same was the case with other European countries and the USA when they followed the British path of industrialization. Since China and India did not follow this path, they lagged behind the industrialized countries of Europe and the USA. Besides, India and China being under colonial domination opened their markets to cheap manufactured goods from the Western industrialized countries destroying their local handicrafts-based industries. This was especially the case in India. In the decline of the Third World, colonial wars, succeeded by occupation and plunder, played a major role. The success of the industrialized countries in the colonial wars was mainly contributed by the strength of their economies and superior war technology produced by the Industrial Revolution. In the year 1800, Europeans occupied or controlled 35 % of the world's area of land. This increased to 67 % in 1878 and 84 % by 1914 (Kennedy 1989, p. 150). The occupation of territories of the third world countries by industrialized countries led to the colonial pattern of development. These countries were developed as a source of raw material and market for cheap manufactured output of factories. The handicrafts of the colonies could not face this competition and were destroyed. Besides economic competition, the colonial governments used several other methods to discourage local handicrafts. This includes dress codes in the offices and courts using Western robes, withdrawal of patronage given by the local rulers and other measures discouraging the use of local products and promotion of foreign-made products.

It is evident that the Industrial Revolution changed the balance of the economy in favour of Western countries, and the modern technology produced by the industrial process made them more powerful to conquer the Third World and convert them into their colonies. This caused the decline of the economies of the third world countries and put them in the subordinate position.

The Britishers, after the Battle of Plassey, increased their presence in the subcontinent by associating with the Nawabs and princes under the pretext of providing them security. In fact, with the decline of the Mughal Empire, a large number of Nawabs and princes had become distrustful of each other. They feared attacks from plunderers and intriguers. The growing feeling of insecurity of the local rulers was used by the Britishers to expand the area of influence and the rule of the East India Company. Superior technology and better organizational ability enabled them to occupy the whole of the subcontinent by the mid-nineteenth century. This enabled them to use India as a source of raw material and market for manufactured goods of modern industry. This resulted in deindustrialization and a virtual period of stagnation. This increased the pressure on land-generating poverty, hunger and famines in extreme form. In fact, the Britishers created, during their 200 years rule in India, an economy colonial in nature. Bipan Chandra (1979, p. 26) has aptly called it a 'colonial mode of production and as a distinct historical stage'. He described: 'It is a well structured whole, a distinct social formation (system) or sub formation (sub system) in which the basic control of the economy and society is in the hands of a foreign capitalist class which functions in the colony (or semi colony) through a dependent and subservient economic, social, political and intellectual structure whose forms can vary with the changing conditions of the historical development of capitalism as a world-wide system'. This stage could not be overcome without waging an active struggle against it to introduce a new stage/system which creates possibilities of modern development. The end to colonialism came in the Indian subcontinent in 1947 after a prolonged independence movement/struggle.

4 Post-colonial Experience of South Asia

The end of colonialism in the subcontinent in 1947 was accompanied by partition leading to the creation of two independent countries of India and Pakistan. The relations between countries of South Asia are largely influenced by continuing animosity between India and Pakistan (Karim 2009, p. 2). In spite of the conflicts between these two countries, they embarked on the path of social and economic development in the post-colonial period. Given the specificities of these countries and post-colonial variations in the structure of governance, these countries have opted for a democratic political system and a capitalist system of economy and market-based decision making mediated through state intervention. After passing through the initial phase of protectionism (1950–1980), all of them have accepted liberalization and globalization under the WTO. The long-term process of capital accumulation has picked up and considerable improvements in indicators of social and economic development have happened in the post-colonial period. The life expectancy was 31.8 years in 1941 in India (including Pakistan and Bangladesh); it has increased to more than 60 years by 2001 in all the major countries of the subcontinent. The literacy rate was 13.9 % in 1941, which has become more than 79 % in the South Asian countries (India attaining 74.04 %) by 2011. Similarly, there are considerable gains in per capita income also. The per capita income in the region was around or less than US$ 100 in 1950 has gone up to US$ 1,990 for Sri Lanka, US$ 1,180 for India, US$ 1,020 for Pakistan, US$ 590 for Bangladesh, US$ 390 for Afghanistan and US$ 440 for Nepal in 2009. The share of India (including Pakistan and Bangladesh) in world industrial gross domestic product (GDP) which stood at 1.7 % in 1953 (Bairoch 1982, p. 302) has increased to 2.67 % in 2009 (The World Bank 2011, pp. 345–355). Although the development process has revived, not all countries in South Asia are at the same level of development. Countries like Afghanistan and Nepal have a per capita income which is lower than the average per capita income (US$ 503) of low-income countries, while the per capita income of Sri Lanka is close to four times and of India and Pakistan more than double that level.

The development process in the post-colonial phase has strengthened and consolidated independence of these countries. At the same time, it has added to the aspiration of the people. Although external threats have visibly declined (except for Afghanistan and Pakistan), internal conflicts have become a common occurrence in the countries of South Asia. In the post-Cold War era, which is accompanied by a strong wave of globalization, there is a paradigm shift in the concept of security. It has been

Table 10.3 Population below poverty line in South Asian countries (in percentage). (Source: The World Bank 2011, pp. 308–309)

Country	Year of survey	Below national poverty line	Year of survey	Below international poverty line US$ 1.25 a day
Afghanistan	2007	42.0	NA	NA
Bangladesh	2008	40.0	2005	49.6
India	1999–2000	28.6	2004–2005	41.8
Nepal	2003–2004	30.9	2003–2004	55.1
Pakistan	1998–1999	32.6	2004–2005	22.6
Sri Lanka	2002	22.7	2002	14.0

NA not available

broadened to include political, economic, societal and environmental concerns. The necessity of protection of human life has generated the concept of human security. The United Nations Development Programme's (UNDP's) concept of human security encompasses not only the achievement of minimal levels of material needs but also the absence of severe threats, economic and political in nature. This concept includes several types of securities for a human and dignified life: job security, income security, health security, environment security and security from crime (Chatterjee 2011, pp. 100–102). Going by these securities, the whole of South Asia seems to suffer from large varieties of lack of security among a large section of population. The working of a socio-economic system, in spite of a renewed, fast dynamism of progress, has not been able to provide these basic securities. The existence of absolute poverty is a common concern of the South Asian countries. The estimates of population below national poverty lines bring out that between 22.7 and 42.0 % of the population in the South Asian countries is living below the poverty line. If the international lowest poverty line (as US$ 1.25 per day consumption) is accepted, then 14.0–55.1 % of the population in these countries lives below the poverty line (Table 10.3); most of the people in these countries are engaged in employment in the unorganized sector. This estimate for India is 92 % (NCEUS 2007, p. 1). The story is the same in other countries. This population suffers from job insecurity, and the irregular nature of employment makes them vulnerable. A vast majority among them is deprived of security against sickness or assured adequate income in old age. Due to lack of job security, poor quality of employment and low income a large section of the population, especially children, suffers from malnutrition. As given in Table 10.4, the percentage of such children varies between 21.1 and 43.5 in the South Asian countries. Malnutrition leads to stunted growth among children and makes them vulnerable to various kinds of health ailments. Consequently, incidences of poor health lead to high mortality rates and low expectancy of life at birth. Given the constitutional provisions, it was expected that the process of governance will lead to progressive reduction in social and economic inequalities. But the statutory government agencies could not safeguard the entitlements to marginalized groups to ensure growth with social justice. This has been recently expressed as the inclusive growth in India. The Eleventh Five Year Plan (Planning Commission 2007, p. 2) has articulated this issue by stating that the fast economic growth in the country in the post-liberalized era has not properly benefited

Table 10.4 Malnutrition among children below the age of 5 in South Asian countries (2000–2008) (in percentage). (Source: The World Bank 2011, pp. 310–311)

Country	Percentage of malnutrition of children
Afghanistan	42.0
Bangladesh	40.0
India	28.6
Nepal	30.9
Pakistan	32.6
Sri Lanka	22.7

the scheduled castes (SCs), scheduled tribes (STs), women and minorities. This plan's document is subtitled as 'Inclusive Growth'. One of the major objectives has been to introduce some correctives at the policy level to ensure that benefits of fast economic growth are shared with the weaker sections of the society. This has been important in view of the tension and the violence continuing for many years in extremist-affected areas in India. The Report of the Expert Group to Planning Commission (Planning Commission 2008) titled as *Development Challenges in the Extremist Affected Areas* has brought out that 'The Government of India has estimated that the Naxalite movement is now active in about 125 districts spread over 12 states' (pp. 2–3). In these areas, the causes of discontent among people have led to the rise and spread of the Naxalite movement which is now almost four decades old. The Expert Group looked into the causes of people's discontent leading to the Naxalite movement and came to the conclusion that 'the causes are varied depending on the characteristics of an area; social, economic and cultural background; a history of not working out solution to lingering structural problems and ineffective application of ameliorative steps undertaken since independence and more so since mid sixties of the last century'. It was further brought out that the failure, inadequacy or injustice of the state's mechanism and institutions to deal with issues of life and livelihood of the people created space for Naxalite activities. The data given in Table 10.5 show a high rate of poverty, mortality and malnutrition among the SC and ST population compared to the other castes. SC and ST population had low asset base, less self-employment (cultivators) and high percentage of casual wage labour. They have a high rate of unemployment, low non-agricultural wage rate and lower rate of literacy. The number of registered cases of discrimination and atrocities against SC and ST populations has been very large. The economically exploited and socially oppressed SC and ST populations have resorted to support the Naxalite movement to protect their livelihood (land and common property), oppose displacement from their land and habitats under special economic zones (SEZs) and mining contracts in forest areas. Thus, this movement has consolidated the resentment against distortions (inequalities), displacement, discrimination and social oppression against the poor and weaker sections of the society. This is against both the constitutional provisions and declared objectives of planning in the country. Similarly, in other South Asian countries there are sections of the population which face growing inequality, deprivation, displacement and discrimination based on ethnic factors. As their grievances are not resolved by statutory bodies and they do not get justice within the present structure of governance, these social groups express their resentment in a variety of ways causing threat to security of the existing order of

Table 10.5 Caste and ethnic group inequality in India (2000). (Source: Planning Commission 2008, p. 4)

	SC	ST	OC
A. Poverty			
1. Poverty percentage of poor (rural)	36	46	21
2. Percentage of poor (urban)	38	35	21
B. Mortality and under nutrition			
1. Infant mortality per 1,000 live birth (2005–2006)	51	44	36
2. Under 5 mortality per lakh population (2005–2006)	88	96	59
3. Percentage of children with anaemia	78	79	72
4. Percentage of underweight children	21	26	14
C. Access of land and capital			
1. Value of asset per household (Rs.) in 1992	49,159	52,660	134,500
2. Percentage of self-employed cultivators	16	48	41
3. Percentage of wage labour (rural)	61	49	25
4. Percentage of casual labour (urban)	26	27	07
Literacy			
Literacy rate 2001 (rural)	51	45	63
Literacy rate 2001 (urban)	68	69	82
Discrimination and atrocity			
Total number of cases of discrimination atrocity (1992–2001)	285,871	47,225	–

SC scheduled caste, *ST* scheduled tribe, *OC* other castes

power structure. In other words, the major threat to peace and security is from internal sources rather than the external sources. It becomes imperative for the policymakers and practitioners of development programmes that development is achieved in such a way that it becomes inclusive in nature. People in general are able to benefit from them and they develop their stake in it. Once people have a stake in the development, they not only ensure its continuance but also work for it by ensuring peace. This is the area where the entire South Asia has a convergence of interest. This is the region with high potentials for growth but has the highest incidence of poverty, malnutrition, ill health, illiteracy and other related problems including impending threat to food security from global warming and climate change (Nelson et al. 2010). The issues related to climate change, food security and problems of poverty, hunger, malnutrition, ill health and illiteracy need to be handled collectively. The scarce physical resources, limited human and physical capital require careful and efficient handling for the benefit of the people of this region. This would demand avoidance of conflicts and wars between countries and within themselves to avoid their adverse effect on development processes.

5 Summing Up

The world is changing with time. The regions and countries, which are most prosperous today, were not so 500 years back. The USA and Europe were not in the dominant position in the world in 1500 AD. It was China and India, which occupied a prominent position in the share of the world industrial output. A combination of factors operating in peace and war times altered the relative position of the countries in the world. The uneven rate and pattern of growth rates among countries altered their status and position in the world. The Industrial Revolution which originated in Great Britain made her a dominant power in the world in the eighteenth century and stayed at that level till the First World War. The spread of the Industrial Revolution to other countries and the absence of it in China and India reduced them to insignificance especially after they came to be colonialized directly and indirectly in the mid-eighteenth century and after. They were reduced to a marginalized position on the eve of their independence. During the twentieth century, the USA and the Soviet Union rose to the position of 'super powers' after the Second World War. The disintegration of the Soviet Union in 1991 led to its demise as a super power and to the decline in its successor state of the Russian Federation. Following the global meltdown and its aftermath, the US economy is experiencing a slow growth rate and decline in its share in the world income. In the twenty-first century, China and India are fast emerging in the manufacturing sector and have already emerged as the second and fourth largest economies, respectively in terms of purchasing power parity income. The strength of China and India is primarily based on modernization of their economies on the basis of application of modern technology in various fields of the economy. Their stress on human capital formation has given them an added advantage. The stability of power structure, prevalence of peace for a large period of post-independence era accompanied by careful policy planning have put them on a strong footing. The rise of these economies and their non-involvement in major army conflicts with neighbours and other countries in the world have strengthened them economically and made them major powers in the world. The rise in productive capacity has enabled them to spend a major proportion of their income (35 % in India and 45 % in China) on gross capital formation raising their growth rates to be highest in the world economy. They are also able to spend a large amount to maintain large armies in peacetime and modernize their arms for armies and fleets. As a result of economic strength, political stability and military power, China and India do not face major external threat to their prosperity.

In a nuclear and globalized world, the concept of security is changing. It is getting concentrated on human security, which has many dimensions such as job security, income security, health security, food security, environmental security and security from crime. These securities, in fact, are related to removing insecurities that bother human lives. They are expected to make human lives secure and enable human beings to live with dignity. These securities are much more connected with the internal working of the socio-economic system. In this context, the issue of peace and security in South Asia is very serious as the insecurities concerned with jobs and income affect the vast majority especially those, who are working as casual labourers both in the rural and urban areas. The proportion of poor and vulnerable population is very large (in India 76.8 %) in South Asia. They quickly fall prey to other insecurities such as those related to food, health, environment and crime. These insecurities make a large section of the population sensitive to environmental changes and policy measures affecting their livelihood, terms and conditions of employment, wage rate, public distribution system for food grains and social security measures. They are also prone to mobilization processes, in view of growing aspirations created by the media. The process of emerging inequalities and internal divisions based on class, caste, ethnicity, regionalism, race, language or religion breed alienation among some sections of the population. The persistence of alienation generates civil strife which threatens peace and increases insecurity until tackled appropriately. Conflict within a country weakens it and reduces the capacity to resist external pressure. In South Asia, Pakistan and Afghanistan are the best illustrations (Ali 2003, pp. 316–328; Cohen 2004, pp. 267–299; Akbar 2011, pp. 288–313). Afghanistan has paid the price in the form of destruction and continued low-income level with little prospects for growth. Pakistan had once a higher per capita income, now trails behind India with continuous low rate of growth in income. Thus, in the contemporary situation, the handling of internal divisions and conflict becomes of paramount importance.

The post-Cold War era has been dominated by dynamism of capitalist globalization based on market forces. This has generated a fast and an unrestricted flow of capital, finance, commodities and technology across the countries. This has been supported by giant multinational corporations and Bretton institutions viz. the World Bank and International Monetary Fund (IMF). A new institution of the WTO has considerably strengthened this process. Globalization is accompanied by information technology revolution which has resulted in fast exchange of information, data and films, electronically transmitted at a high speed but with low and affordable cost. Thus, expansion of trade and commerce is accompanied by exchange of ideas and cultural values. The dominant aspect in ideas and cultural exchanges are based on capitalist value system. This has linked educated communities globally, creating high levels of aspirations among the educated young population in developing countries, especially in South Asia. For peace and security in South Asia, the idea of I. K. Gujral, India's former Prime Minister is quite valid. Mr. Gujral (2007, p. 6) wrote:

> There is increasing tendency in the world today to view security issues entirely in military terms and seek military answers to all human conflicts. No part of the world is exempt from this folly but it seems particularly prevalent in South Asia. Such a view is tragically wrong because any elementary survey would show that the dynamic of commerce and culture has had a much greater impact than military force. Indeed, the most bitter division in human history was decided without a shot being fired. The cold war, which saw the mightiest armies in human history arrayed against each other, ended not with a bang but a whimper. The Berlin Wall was brought down not by tanks and soldiers but by the invisible pull of commerce and culture. Conversely, we now see the examples of Iraq and Afghanistan, where conventional armies have failed to defeat what are essentially tribal militias.

The freedom struggle in India could force the British to withdraw from the country not by organized army but by mass resistance, which was largely non-violent. In recent times, Egypt and other Middle East countries have seen mass resistance producing some successful results by removing the ruler established for many decades. Thus, ensuring peace and security in a society, the internal cohesion, is of utmost importance. This depends on a feeling of sense of belonging, removal of alienation and provision of human security. This depends on shaping the development process, which is inclusive in nature. Such a development process is described as growth with social justice. This enables all sections of the society to share gains of development and avoid cost of development through displacement on a few groups/sections.

The political leadership in the region must grasp the complexity of the situation, understand others' compulsions and accommodate each other in the spirit of statesmen to harness the creativity of the people and natural resource to ensure peace and prosperity in the region. In the era of nuclear weapons, countries cannot defeat enemies without inviting self-destruction. The enlightened self-interest demands that South Asian countries join hands in constructive cooperation to turn this region peaceful and secure to live. This can create the best opportunity for human creativity to ensure sustainable development. The framework of SAARC established in 1985 can be extremely useful to meet the challenges of climate change and solve internal problems of poverty, hunger, disease, illiteracy and inequalities based on gender, race, religion, region, language, caste, etc. The region can learn positive lessons from the experience and evolution of the European Union (EU) after the Second World War. The countries of the region must move from cooperation in trade to cooperation in development by pooling their resources for solving collective problems of the region.

References

Akbar MJ (2011) Tinderbox: the past and future of Pakistan. Harper Collins, New Delhi
Ali T (2003) The clash of fundamentalisms: crusades, jihads and modernity. Rupa Co., Delhi
Amin S (1974) Accumulation on world scale, vol I. Monthly Review Press, New York
Bairoch P (1982) International industrialization levels from 1750 to 1980. J Eur Econ Hist 11(2): 269–333
Chandra B (1979) Nationalism and colonialism in modern India. Orient Longman, New Delhi
Chatterjee B (2011) Peace, security and development. Man Dev XXXIII(1):100–102, 105–106
Cohen SP (2004) The idea of Pakistan. Oxford University Press, New Delhi
Frank AG (1975) On capitalist under development. Oxford University Press, Bombay
Gadgil DR (1971) Industrial evolution of India in recent times 1860–1939. Oxford University Press, Bombay
Gujral IK (2007) Peace and development. In: Banerjee S (ed) Peace and development: Haksar memorial volume iv. CRRID, Chandigarh
Karim T (2009) Rethinking regional cooperation in South Asia: changing paradigm from regional framework in integrated sub-regional development. In: Gaur N, Laxmi V (eds) South and Central Asia: issues of peace and development. CRRID, Chandigarh
Kennedy P (1989) The rise and fall of great powers: economic and military conflicts from 1500-2000. Vintage Books, New York
Lucas RE (1988) On the mechanics of development. J Monet Econ 22(1):3–42
NCEUS (2007) Report on conditions of work and promotion of livelihoods in unorganized sector. Government of India, New Delhi
Nelson GC et al (2010) Food security, farming and climatic change to 2050. International Food Policy Research Institute, Washington DC
Planning Commission (2007) Eleventh five year plan 2007–12, vol I. Government of India, New Delhi
Planning Commission (2008) Development challenges in extremist affected areas: report of an expert group. Government of India, New Delhi

Romer PM (1990) Endogenous technical change. J Polit Econ 98(5):S71–S102

Smith A (1776) An inquiry into the nature and causes of wealth of nations. Book V, e-edition. www.gutenberg.org/files/3300/3300-h/3300-h.htm. Accessed 28 Feb 2009

The World Bank (2011) World development report 2011: conflict, security and development. The World Bank, Washington DC

Chapter 11
Well-Being in Human Development Framework: Constituents and Aggregation

P. K. Chaubey

1 Introduction

Welfare economics ignored for its long capability dimension of development and development economics did not pay sufficient attention to the softer side of human progress. This has been made somewhat better in the recent decades. Although certain terms will be explained further in the chapter, we can assert the contrast between opulence of people and their well-being in the following terms: Well-being is over the functioning space while opulence is on the commodity space.

It appears, Sen has succeeded in impressing upon, the fact that being well[1] is not the same as being well-off. If someone is fabulously rich in terms of his possessions but poor in health he could scarcely be called being well while he is exceedingly well-off. Therefore, a developmental approach which follows the path making people well-off is likely to be different than the one that tries for the well-being of the people. Moving from concerns of being well-off to that of being well is a tedious journey. Let us follow it.

The primary specification of a person's well-being is in terms of a functioning vector. It can be converted into a scalar measure of well-being only through a real-valued 'valuation function' $v(.)$ which maps functioning vectors into numerical representations. This valuation may not even be complete. However, the most difficult part would be to judge well-being level of social states, which involves (1) comparisons over time or across space and (2) preferences of different people. Let us start from the well-being of an individual and then move to that of the society of which the individuals are constituents.

[1] Of late, it is the word well-being rather than welfare that is used in the economics literature, particularly in the part which is concerned with development.

P. K. Chaubey (✉)
Indian Institute of Public Administration (IIPA), New Delhi, 110002, India
e-mail: pkchaubey@yahoo.com, chaubey.pk@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_11,

2 Well-Being of an Individual

Let X be the set of commodity vectors x over which the individual i has command. It is his set of entitlements whether he earns them by sweat or property or if societal norms oblige him to have them through societal institutions including State. He can choose any one of these vectors. The situation can be likened with the income line in the consumer theory, which shows a number of commodity bundles that the consumer can buy with his income. This is set X. But he chooses only one of them for whatever reasons. This is x. Thus, $x \varepsilon X$. Commodity command X is a means of well-being.

Let F be the set of personal utilization functions $f(.)$ each one of which is a characteristic set of personal features of the being of the individual i that he can utilize to achieve a set of functionings from x chosen from the set of his entitlements X. Each f denotes a set of capabilities. This $f(.)$ has, however, to be conceived in the terms of a matrix if a vector has to be converted into another vector.

A functioning is an achievement of the individual through a combined choice of $f(.)$ and x. Let ϕ be a vector of such achieved functionings—beings and doings[2]. This is obviously obtained from a choice of $f(.)$ and x. Therefore, $\phi = f(x)$.

It may be noted that Sen has mediated the matter of getting ϕ through (commodity-) characteristic vector y from commodity vector x. This conversion function, usually a matrix, has noting to do with the individual i, while all other vectors and matrices are associated with the individual i. Bringing y into picture is not strictly necessary for it is x, which has to be selected in the exercise though selection of x depends definitely on the characteristics of the commodities in consideration.

For a given commodity vector $\bar{x}$, functioning vector feasible for the individual i has, naturally, to be given by

$$p(x) = [\phi/\phi = f(\bar{x}), \textit{for some } f(.)\varepsilon F]$$

Obviously, p is a set of ϕ varying in accordance with choice about $f(.)$. For the set of commodity vectors X, therefore, the functioning vector feasible for the individual i would be given by

$$Q(x) = [\phi/\phi = f(x), \textit{ for some } f(.)\varepsilon F \textit{ and for some } x\varepsilon X],$$

where there is a scope for varying $f(.)$ and x. Obviously, Q is a set of ϕ varying in accordance with choice about $f(.)$ and x both. $Q(x)$ represents the freedom that the individual i has in terms of choice of functioning through choice of using personal utilization functions $f(.)$ and of selecting commodity bundle x. Naturally, the individual i is selecting $f(.)$ (for example, hearing) from the set of his personal features

[2] Beings and doings are two different things. Beings are stocks while doings are flows. Functionings can be therefore of two varieties. Being literate is a stock feature while swimming across a river is a flow. Although discussion of time dimension is missing in the literature, in aggregative exercise, annual data is used.

F and x (for example, a ghazal by Jagjit Singh) from the set of his entitlements X, which is his command over commodities.

Well-being can then be understood as a valuation of ϕ, indicating the kind of being he is achieving (Sen 1987, p. 8). In case ϕ's can be ranked completely, a scalar value can be attached to it. If values v can be attached to ϕ, which means $v = v(\phi)$, then the values of well-being the individual can attain will be given by the set

$$V = [v/v = v(\phi), \textit{ for some } v(\phi) \textit{ in } Q]$$

Human development is then concerned about expanding the limits of choice reflected in X (command over outside environment) and F (ingrained capabilities inside) both.[3]

3 Functionings

There is, however, a need to focus on functionings. A functioning is an achievement of a person: what he manages to do or to be. It reflects, as it were, a part of the state of that person (Sen 1987, p. 7, lines 5–6). A functioning is neither a commodity, nor having it, nor having utility from it. A functioning refers to a doing or a being (Sen 1987, p. 8). Functioning vector thus reflects a set of doings and beings (p. 23)

Longevity and literacy in particular, have received a good deal of attention in the development literature. This relates functionings to non-market observations of personal state. It is natural to make extensive use of non-market data, direct observations of conditions of persons, to understand the functioning that they achieve. It can be argued that this type of investigation can be sensibly extended to cover other conditions of persons which economics literature has been rather reluctant to examine, in particular, morbidity and undernourishment, which relate to some important functionings that frequently fail badly in poor developing countries (Sen 1987, p. 30).

Functionings involving longevity, nourishment, basic health, avoiding epidemics, and being literate may have less variation. There are others, like playing golf or acting in films, with great deal of variation even in rich countries. Functionings involving literary, cultural and intellectual pursuits and vacationing and traveling have a great deal of variation (Sen 1987, pp. 30–31)

The conversion of commodities (characteristics) into personal achievement of functionings demands a variety of factors—personal and social. For example, in the case of nutritional achievements, it depends on factors such as (1) metabolic rates, (2) body size, (3) age, (4) sex, (5) activity levels (6) medical (health) conditions and (7) nutritional knowledge, and also on (8) climatic conditions and (9) access to medical services. The first seven can be taken as personal features—the beings as part of the persona. The last two can be taken as part of environment/entitlement

[3] Human development literature in most cases/places discusses human development as if it is exclusively a matter of capabilities. Capabilities to read, to swim, to play guitar, to enjoy classical songs and to appreciate paintings are fine but these capabilities may make me frustrated if the concomitant commodities are absent.

(both of which enabled from outside). These make one choose f from F, which is only partly a matter of choice but there is scope for choice. One cannot choose age, sex and height but he or his society can improve medical conditions and nutritional knowledge. Strictly speaking, such traits as are outside the choice of the individual are not part of F.

4 Capability

Since human development has been defined in terms of extending capabilities and enhancing functionings, it would do good to discuss the idea of capability in some detail. We may recall that Q has been called by Sen as the capabilities of the individual i, given certain parameters. To be more precise, Q is the capability set—a set of functioning n-tuples, each representing an alternative vector of functionings ϕ. It is, thus, but a combination of functionings which the individual can achieve. It does not reflect the personal features of the individual, which are normally understood as one's beings[4].

The totality of all the alternative functioning vectors the person can choose from, given by the contingent circumstances, is Q. Q reflects the person's choice set, i.e. the various alternative functioning bundles he or she can achieve through choice of $f(.)$ and x (Sen 1987, p. 18, lines 14–19). But Q obviously depends not only on capabilities, as ordinarily understood, but also on circumstances.

If a capability to do is understood as a personal feature, a being, then it is better to regard f as a capability vector, not ϕ, which operates on x to produce ϕ. Sen refers to bike as commodity, transportation as its characteristic, and bicycling as a functioning. Functioning of bicycling will depend on the fact whether the individual is able-bodied (and skilled in cycling) or crippled (or unskilled). Functioning ϕ thus depends on capability f and commodity x.

Capability that is acquired is of more relevance. Some of the capabilities can be acquired only through mediation of commodities like cycling or swimming. They are learnt. If naturally deficient, some of capabilities can be enhanced through commodities like seeing through glasses, hearing through aid, speaking through synthesizer, and so on.

5 Aggregation of Functionings

Although the whole capability approach concentrates on the individual, numerical presentation has exclusively been made only at an aggregate level. Aggregates may be small groups, communities, and nations. Aggregation of sets of potential functioning n–tuples (capabilities) of the individual constituting a community has not

[4] By slip, Sen (1987, p. 10, line 6) has once referred to f as the vector of functionings, which it is not correct as per the definitions (Sen 1987, p. 7) given by him.

been attempted. Aggregation of actual achieved functionings (achievement in human development literature) has however been attempted. Let us designate this set of n achieved functionings of p individuals as A and represent it in the following two equivalent manners:

$$A = \begin{bmatrix} \begin{Bmatrix} a_{11} \\ a_{21} \\ : \\ a_{j1} \\ : \\ a_{n1} \end{Bmatrix}_1 \begin{Bmatrix} a_{12} \\ a_{22} \\ : \\ a_{j2} \\ : \\ a_{n2} \end{Bmatrix}_2 \ldots\ldots \begin{Bmatrix} a_{1k} \\ a_{2k} \\ a_{jk} \\ : \\ a_{nk} \end{Bmatrix}_k \ldots\ldots\ldots \begin{Bmatrix} a_{1p} \\ a_{2p} \\ a_{jp} \\ : \\ a_{np} \end{Bmatrix}_p \end{bmatrix} = \begin{bmatrix} \{a_{11}\ a_{12}..a_{1k}..a_{1p}\}_1 \\ \{a_{21}\ a_{22}..a_{2k}..a_{2p}\}_2 \\ \{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\} \\ \{a_{jq} a_{j2}..a_{jk}..a_{jp}\}_l \\ \{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\} \\ \{a_{n1}\ a_{n2}..a_{nk}..a_{np}\}_n \end{bmatrix}$$

All these elements have to be aggregated. But how do we do it? There could be two ways. One could be that we attempt aggregation of each individual's vectors into their respective well-beings and then individuals' well-beings into society's well-being. This purports to converting each column vector into a scalar and thus first creating a row vector of p-size and then converting this column vector into a scalar number. The other could be that we attempt to aggregate a single being/doing of all p individuals (like literacy or life expectancy) and then these aggregated n social states of beings/doings into society's well-being. This purports to converting each row vector into a scalar and thus first creating a column vector of n-size and then converting this column vector into a scalar number.

The Human Development Reports have attempted the second course. For example, health-beings of all individuals were first converted into life-expectancy of the society and literacy capabilities of all individuals were first converted into literacy level of the society. The same could be said about enrolment ratios or schooling years but not so clearly about per capita GDP. Then, these societal components of functionings (beings and doings) have been aggregated into Human Development index in a linear/geometric fashion (UNDP 1990, 2010).

References

Sen A (1987) Commodities and capabilities. Oxford University Press, Delhi

United Nations Development Program (1990) Human development report. Oxford University Press, Delhi

United Nations Development Program (2010) Human development report. Oxford University Press, Delhi

Chapter 12
Human Capital Accumulation, Environmental Quality, Taxation and Endogenous Growth

Bidisha Chakraborty and Manash Ranjan Gupta

1 Introduction

There exists a substantial theoretical literature focusing on the interaction between the economic growth and the environmental degradation.[1] In these models, the degradation of environmental quality either lowers the utility of the consumer or lowers the productivity of the factors. Most of these models are built in a one-sector Ramsey–Solow framework. Environmental degradation is viewed as the social by-product of the use of modernized machineries in the production sector because the operation of these modenized machines requires the use of pollution-enhancing raw materials like oil, coal, etc. Some authors like Mohtadi (1996), Bretschger and Smulders (2007), Perez and Ruiz (2007), Hettich (1998), etc. assume a direct relation between the level of environmental pollution and the stock of physical capital when entire physical capital stock is fully utilized.[2] Other authors like Grimaud and Tournemaine (2007), Hettich (1998), Grimaud (1999), etc. assume the level of environmental pollution to be a function of the level of output of the aggregate production sector.

[1] See, for example, Mohtadi (1996), Dinda (2005), Gradus and Smulders (1993), Hettich (1998), Rosendahl (1996), Perez and Ruiz (2007), Endress et al. (2005), Grimaud (1999), Ricci (2007), Grimaud and Tournemaine (2007), etc.

[2] If capital accumulation means replacement of old machines by more eco-friendly machines, then environmental pollution should vary negatively with capital accumulation.

B. Chakraborty (✉)
Department of Economics, Jadavpur University, Kolkata,
West Bengal 700032, India
e-mail: bidisha.chakraborty@gmail.com

M. R. Gupta
Economic Research Unit, Indian Statistical Institute, Kolkata,
West Bengal 700108, India
e-mail: manas_t@isical.ac.in

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_12,

There exists another set of theoretical literature focusing on the role of human capital accumulation on economic growth.[3] The literature starts with the Lucas (1988) model, and this model has been extended and reanalysed by various authors in different directions. The rate of labour augmenting technical progress, i.e. the rate of human capital accumulation is endogenous to the analysis; and the productivity parameter of the human capital accumulation technology is an important determinant of the rate of growth. Some of the works focusing on the interaction between economic growth and environmental pollution are based on the Lucas (1988) framework. In Hettich (1998), environmental pollution negatively affects the welfare of the household, and, in Rosendahl (1996), environmental quality produces a positive effect on the productivity of capital. Ricci (2007) conducts a survey of the literature. However, in none of these existing works, except of Gradus and Smulders (1993), environmental quality affects the learning ability of the individuals.

When human capital accumulation is the engine of economic growth, the learning ability of the individual becomes an important determinant of the rate of human capital accumulation. Environmental pollution produces negative effects on the health of the individual, and this lowers the ability to learn. Noise pollution disturbs the academic environment. Margulis (1992) finds significant empirical correlation between levels of lead in air and in blood. Next, he shows that children with higher blood lead levels have a lower cognitive development and require supplemental education. Kauppi (2006) shows that methyl mercury, whose exposure to human comes from fish consumption, may lower the learning ability of children. Air pollution also causes problems related to eye sight and functioning of the brain. Gradus and Smulders (1993) consider this negative effect of environmental pollution in an otherwise identical Lucas (1988) model. However, they do not analyse the effects of various fiscal policies and the transitional dynamic properties of that model.

Human capital accumulation also has a positive effect on the upgradation of the environmental quality. Education makes people aware of the environmental problems and of the importance of protecting environment, and educated people can protect the environment in a scientific way. This positive effect of human capital accumulation on the environmental quality is ignored not only by Gradus and Smulders (1993) but also in the other theoretical models like that of Mohtadi (1996), Dinda (2005), Hettich (1998), Rosendahl (1996), Ricci (2007),[4] etc. However, there are empirical supports in favour of this positive relationship. Torras and Boyce (1998) regress environmental pollution on income, literacy rate, Gini coefficient of income inequality, etc. and find that the literacy rate has a significant negative effect on pollution, particularly in low-income countries. Petrosillo et.al. (2007) find that the attitudes of the tourists, who visit marine-protected areas, are highly dependent on their education level. Clarke and Maantay (2006) find that the participation rate of the people in the recycling

[3] See, for example, Lucas (1988), Rebelo (1991), Ben-Gad (2003), Caballe and Santos (1993), Ortigueira (1998), Faig (1995), Mino (1996), Greiner and Semmler (2002), Alonso-Carrera and Freire-Seren (2004), Chamley (1993), etc.

[4] Some authors, e.g. Grimaud (1999), Goulder and Mathai (2000), Hart (2004), study the issue of environment in R&D-driven growth model where innovations help to improve the environment.

programme conducted in New York City and its neighbourhood is highly dependent on the education level of the participators.

In this chapter, we consider a modified version of the Lucas (1988) model with two special features. (i) Environmental quality positively affects the marginal return to education; and (ii) environmental quality varies positively with the stock of human capital and negatively with the stock of physical capital whose full utilization is ensured by the perfect flexibility of factor prices. We analyse the effect of taxation on the steady-state equilibrium rate of growth of the economy. The interesting results obtained in this chapter are as follows. Firstly, the steady-state equilibrium rate of growth, in this model, varies positively with the proportional tax rate imposed on output or on capital income when tax revenue is spent as lump sum payment. However, this rate of growth is independent of the tax rate imposed on labour income. In Lucas (1988), this rate of growth is independent of the tax rates imposed either on output or on capital income. In Rebelo (1991), the rate of growth varies inversely with the tax rate imposed on output or on capital income. Secondly, there exists unique saddle path converging to the unique steady-state equilibrium point. Thirdly, the positive effect of output taxation on the steady-state equilibrium rate of growth is strengthened when tax revenue is spent as abatement expenditure. Fourthly, the optimum output tax rate, which is obtained by maximizing the balanced rate of growth, varies proportionately with competitive output share of human capital when tax revenue is spent as educational subsidy.

The rest of the chapter is organized as follows. Section 2 presents the basic model and contains the analysis of the effect of output taxation on the steady-state equilibrium rate of growth when tax revenue is distributed as lump sum payment. Section 3 presents the analysis related to the transitional dynamic properties of the model. Section 4 contains the analysis related to the effects of factor income taxation. In Sect. 5, we reanalyse the basic model when tax revenue is spent on abatement activity. Section 6 contains the analysis with tax revenue financing the educational subsidy. In Sect. 7, we study the relationship between growth and welfare. Concluding remarks are made in Sect. 8.

2 The Model

The model presented in this chapter is an extension of the Lucas (1988) model. The government imposes a proportional tax on the output and the tax revenue is distributed among the individuals as lump sum payment. The dynamic optimization problem of the representative individual is to maximize

$$\int_0^\infty U(C)e^{-\rho t}dt$$

subject to the production function given by

$$Y = A(aH)^\gamma K^{1-\gamma}, \tag{12.1}$$

with $A > 0$ and $0 < \gamma < 1$; the dynamic budget constraint given by

$$\dot{K} = (1-\tau)Y - C + P, \tag{12.2}$$

with $0 \leq \tau \leq 1$; the human capital accumulation function given by

$$\dot{H} = m(E)^{\delta}(1-a)H, \tag{12.3}$$

with $\delta > 0$; and the environmental stock accumulation function given by

$$E = E_0 K^{-\beta} H^{\beta}, \tag{12.4}$$

with $E_{0,}\ \beta > 0$. Here, A is the technology parameter; K is the stock of physical capital; H is the stock of human capital; and τ is the proportional output tax rate. E is the environmental quality; P is the lump sum income transfer resulting from the distribution of tax revenue; and C is the level of consumption of the representative household. Y is the level of output and a is the fraction of labour time allocated to production. m is the productivity parameter in the human capital accumulation function; u(.) is the utility function; ρ is the rate of discount; and γ is the elasticity of output with respect to human capital. Equations (12.3) and (12.4) make the present model different from Lucas (1988). Equation (12.4) with $\beta > 0$ implies that the environmental quality varies positively with the stock of human capital and negatively with the stock of physical capital. Equation (12.3) with $\beta > 0$ implies that the positive external effect of environmental quality is present in the human capital accumulation function. If $\delta = 0$ or $\beta = 0$, then we come back to the original Lucas (1988) model. The representative individual solves this optimization problem with respect to the control variables C and a. K and H are two state variables. However, the individual cannot internalize the externality.

We assume $U(C) = lnC$ for the sake of simplicity. We also impose a restriction on the parameters given by

$$\gamma > \frac{\beta\delta}{1-\beta\delta}.$$

The current value Hamiltonian is given by

$$Z = lnC + \lambda_K[A(1-\tau)(aH)^{\gamma}K^{1-\gamma} - C + P] + \lambda_H[m(E)^{\delta}(1-a)H].$$

Here, λ_K and λ_H are the costate variables of K and H.

The first-order optimality conditions are given by the following:

$$\frac{\partial Z}{\partial C} = \frac{1}{C} - \lambda_K = 0, \tag{12.5}$$

and

$$\frac{\partial Z}{\partial \alpha} = \lambda_K A(1-\tau)\gamma K^{1-\gamma}a^{\gamma-1}H^{\gamma} - \lambda_H m(E)^{\delta}H = 0. \tag{12.6}$$

Time behaviour of the costate variables along the optimum growth path should satisfy the following:

$$\dot{\lambda}_K = \rho\lambda_K - \lambda_K A(1-\tau)(1-\gamma)(aH)^{\gamma} K^{-\gamma}, \tag{12.7}$$

and

$$\dot{\lambda}_H = \rho\lambda_H - \lambda_K A(1-\tau)\gamma(a)^{\gamma} H^{\gamma-1} K^{1-\gamma} - \lambda_H m(E)^{\delta}(1-a). \tag{12.8}$$

Transversality conditions are given by the followings:

$$\lim_{t\to\infty} e^{-\rho t}\lambda_K(t)K(t) = \lim_{t\to\infty} e^{-\rho t}\lambda_H(t)H(t) = 0.$$

The budget of the government is balanced, and, hence,

$$P = \tau Y.$$

Hence, at the aggregate level, Eq. (12.2) is modified as follows:

$$\dot{K} = Y - C. \tag{12.9}$$

2.1 Steady-State Equilibrium

Along the steady-state equilibrium growth path, $\frac{\dot{K}}{K} = \frac{\dot{H}}{H} = \frac{\dot{C}}{C}$ and $\frac{\dot{a}}{a} = 0$. Using Eqs. (12.5) and (12.7), we have

$$\frac{\dot{C}}{C} = A(1-\tau)(1-\gamma)a^{\gamma}(\frac{H}{K})^{\gamma} - \rho. \tag{12.10}$$

From Eq. (12.6), we have

$$\frac{\lambda_K}{\lambda_H} = \frac{m(E)^{\delta} H}{A(1-\tau)\gamma K^{1-\gamma} a^{\gamma-1} H^{\gamma}}. \tag{12.11}$$

Differentiating both sides of Eq. (12.11) with respect to time and then using the steady-state equilibrium condition, we have

$$(\frac{H}{K})^{\gamma-\beta\delta} = \frac{m(E_0)^{\delta}}{A(1-\tau)a^{\gamma}(1-\gamma)}. \tag{12.12}$$

Along the steady-state equilibrium growth path, $\frac{\dot{C}}{C} = \frac{\dot{H}}{H}$. So, using Eqs. (12.3), (12.4), (12.10) and (12.12), we have

$$a = [\frac{\rho^{\gamma-\beta\delta}(A(1-\tau)(1-\gamma))^{\beta\delta}}{m^{\gamma}(E_0)^{\delta\gamma}}]^{\frac{1}{(\gamma-\beta\delta-\gamma\beta\delta)}}. \tag{12.13}$$

We assume $\gamma > \beta\delta + \gamma\beta\delta$. Thus, we find a negative relationship between optimum a and τ in the steady-state growth equilibrium. Note that $a < 1$; and this is guaranteed if

$$\frac{\rho^{\gamma-\beta\delta}(A(1-\tau)(1-\gamma))^{\beta\delta}}{m^{\gamma}(E_0)^{\delta\gamma}} < 1. \tag{12.14}$$

Hence, using Eqs. (12.10), (12.12) and (12.13), the balanced growth rate of the economy, denoted by g, is obtained as follows:

$$g = [A(1-\tau)(1-\gamma)a^{\gamma}]^{-\frac{\beta\delta}{\gamma-\beta\delta}}[m(E_0)^{\delta}]^{\frac{\gamma}{\gamma-\beta\delta}} - \rho. \tag{12.15}$$

Equation (12.13) shows that a is positively related to the discount rate, ρ, production technology parameter, A, and the output elasticity coefficient with respect to physical capital, $(1 - \gamma)$. This also shows that a is negatively related to the tax rate, τ, productivity parameter of the human capital accumulation function, m, and the initial environmental quality, $E_{0.}$

Substituting $A(1-\tau)(1-\gamma)$ from Eq. (12.13) to Eq. (12.15), we have

$$g = \rho(\frac{1}{a} - 1).$$

So the growth rate, g, varies negatively with a. As a varies negatively with the tax rate, τ, the growth rate varies positively with the tax rate, τ. So we have the following proposition.

Proposition 1 When human capital accumulation function in the Lucas (1988) model receives the negative external effect from environmental pollution, the steady-state equilibrium rate of growth of the economy varies positively with the tax rate on output.

If either $\delta = 0$ or, $\beta = 0$ then we go back to the Lucas (1988) model without external effect. In this case,

$$a = \frac{\rho}{m} \text{ and } g = m - \rho.$$

Hence, the tax rate on output cannot affect the growth rate in this case. In the model of Rebelo (1991), the increase in output tax rate reduces the rate of growth of the economy. However, in this present model, the rate of growth varies positively with the output tax rate.

As the tax rate on output is increased, the post-tax marginal productivity of physical capital is reduced. This reduces the net rate of return on physical capital, and, hence, the physical capital is accumulated at a lower rate. As a result, the rate of upgradation of environmental quality is increased. This produces positive external effect on human capital accumulation. So, in the new steady-state equilibrium, the rate of growth of human capital and the rate of growth of income are increased.

In the model of Rebelo (1991), an increase in the proportional output tax rate causes a decline in the rate of growth. So, the optimum tax rate is zero in the Rebelo

(1991) model. In Rebelo (1991), physical capital accumulation positively affects the human capital accumulation. So, the increase in the output tax rate reduces the rate of growth of human capital and the rate of growth of the economy in the steady-state equilibrium. In Mohtadi (1996), an increase in the output tax rate reduces the rate of growth of physical capital. Although there exists negative external effect of physical capital accumulation on the environmental quality in his model, the rate of growth of output is positively related to the rate of growth of physical capital accumulation. Hence, in his model, an increase in output tax rate reduces the rate of growth of output. So, our result contradicts the results obtained by Rebelo (1991) and Mohtadi (1996). This model points out a case where growth rate is positively related to the tax rate. This is so because physical capital accumulation has no positive effect on the human capital accumulation in this model. If we assume Rebelo (1991) type of human capital accumulation function where physical capital contributes positively to the human capital accumulation and consider the negative effect of physical capital accumulation on environmental quality, then we may not have a monotonic relationship between the growth rate and the tax rate. On the contrary, we may have an interior optimal tax rate which would maximize the balanced growth rate of the economy.

3 Transitional Dynamics

We now turn to analyse the transitional dynamic properties of the model around the steady-state equilibrium point. We derive the equations of motion which describe the dynamics of the system.

We define two new variables x and y such that $x = \frac{C}{K}$ and $y = \frac{H}{K}$. Using Eqs. (12.1), (12.9) and (12.10), we have

$$\frac{\dot{x}}{x} = Aa^{\gamma}y^{\gamma}[(1-\tau)(1-\gamma)-1]-\rho+x. \tag{12.16}$$

Using Eqs. (12.1), (12.3), (12.4) and (12.9), we have

$$\frac{\dot{y}}{y} = mE_0{}^{\delta}y^{\beta\delta}(1-a)-Aa^{\gamma}y^{\gamma}+x. \tag{12.17}$$

Differentiating both sides of the Eq. (12.11) with respect to time, *t*, and then using Eqs. (12.3), (12.7), (12.8) and (12.9), we have

$$\frac{\dot{a}}{a} = \frac{\{1-(1-\gamma+\beta\delta)(1-a)\}}{(1-\gamma)}m(E_0)^{\delta}y^{\beta\delta}+Aa^{\gamma}y^{\gamma}\{\frac{\beta\delta}{1-\gamma}+\tau\}-\frac{(1-\gamma+\beta\delta)}{1-\gamma}x. \tag{12.18}$$

The dynamics of the system is now described by the differential Eqs. (12.16)–(12.18). Their solutions describe the time path of the variables x, y and a. When either $\beta = 0$ or $\delta = 0$, and $\tau = 0$, these equations of motion become identical to those obtained in Benhabib and Perli (1994).

Along the steady-state equilibrium growth path, $\dot{x} = \dot{y} = \dot{a} = 0$. Their steady-state equilibrium values are denoted by x^*, y^* and a^*. From Eq. (12.16), we have

$$x^* = \rho - Aa^{*\gamma} y^{*\gamma}\{(1-\tau)(1-\gamma) - 1\}.$$

From Eqs. (12.17) and (12.18), we have

$$y^* = [\frac{\rho}{A(1-\tau)(1-\gamma)a^{*1+\gamma}}]^{\frac{1}{\gamma}};$$

and

$$a^* = [\frac{\rho^{\gamma-\beta\delta}\{A(1-\tau)(1-\gamma)\}^{\beta\delta}}{(mE_0{}^{\delta})^{\gamma}}]^{\frac{1}{\gamma-\beta\delta-\gamma\beta\delta}}.$$

So the steady-state equilibrium point is unique. If either $\beta = 0$ or $\delta = 0$, then

$$a^* = \frac{\rho}{m}.$$

This expression is similar to those obtained in Benhabib and Perli (1994).

We now turn to show that there exists a unique saddle path converging to the unique steady-state equilibrium point. Note that it is a system of three differential equations. Initial value of the variable, y, is historically given, and the values of other two variables x and a can be chosen by the controller. So if the roots are real then, in order to get the unique saddle path converging to the steady-state equilibrium point, we need exactly one latent root of the Jacobian matrix corresponding to the system of differential equations to be negative and the other two roots to be positive.

We can show that[5]

$$\text{Trace of } J = x^* + A(a^*y^*)^{\gamma}\gamma[\tau + \frac{\beta\delta}{1-\gamma} - 1]$$
$$+ m(E_0 y^{\beta})^{\delta}[(1-a^*)\beta\delta + \frac{(1-\gamma+\beta\delta)}{(1-\gamma)}a^*];$$

and

$$\text{Det of } J = m(E_0 y^{\beta})^{\delta} x^* \frac{\rho}{(1-\gamma)}[\beta\delta(1+\gamma) - \gamma].$$

Here, J is the 3×3 Jacobian matrix corresponding to the system of three differential equations.

Since, we have assumed $[\beta\delta(1+\gamma) - \gamma] < 0$, the determinant of J is always negative. Note that if there does not exist any external effect of aggregate human capital on production, then the determinant of the Jacobian matrix of corresponding differential equations in the Lucas (1988) model is always negative. The negative sign of the determinant of J implies that either all the three latent roots are negative

[5] Derivation in detail is shown in Appendix A.

or only one root is negative with the other two roots being positive. So, we have to look at the sign of the trace of J. Here, the trace of J is positive. So, all the roots cannot be negative. Hence, only one latent root would be negative and the other two roots would be positive.

Hence, in this case, there is a unique saddle path converging to the unique steady-state equilibrium point. So, we have the following proposition:

Proposition 2 There exists a unique saddle path converging to the unique steady-state equilibrium if $[\beta\delta(1+\gamma)-\gamma] < 0$.

Using Eqs. (12.4) and (12.17) and the definition of y, we have

$$\frac{\dot{E}}{E} = \beta[mE_0{}^{\delta} y^{\beta\delta}(1-a) - Aa^{\gamma} y^{\gamma} + x].$$

So, it is clear from this differential equation that, once we obtain time behaviour of y, x and a along the unique saddle path, we can then easily solve for the intertemporal transitional behaviour of the environmental quality, E. Since it is a 3×3 dynamic system, we cannot use the phase diagram to examine the transitional dynamics of environmental quality. However, it is clear that $\frac{\dot{E}}{E} > (<)0$ for $\frac{\dot{y}}{y} > (<)0$. So, it is the time behaviour of the capital intensity of production, $\frac{H}{K}$, which determines the time behaviour of the environmental quality, E.

4 Factor Income Taxation

We now consider taxation on factor income at different rates. Suppose that a tax at the rate of τ_K and a tax at the rate of τl are imposed on capital income and labour income, respectively. If $\tau_K = \tau l$, then it is equivalent to taxing output at that rate. The budget constraint of this individual, in this case, is given by:

$$\dot{K} = (1-\tau_K)rK + (1-\tau l)waH - C + P. \tag{12.19}$$

Here, r and w are rental rate on capital and wage rate, respectively. The dynamic optimization problem of the representative individual in this model is to maximize

$$\int_0^{\infty} U(C)e^{-\rho t}dt,$$

with

$$U(C) = lnC,$$

subject to the Eqs. (12.1), (12.3), (12.4) and (12.19).

The competitive equilibrium conditions of the profit-maximizing firm are given by

$$r = A(1-\gamma)(aH)^{\gamma}K^{-\gamma};$$

and

$$w = A\gamma(aH)^{\gamma-1}K^{1-\gamma}.$$

From the first-order optimality conditions,[6] we have

$$\frac{\dot{C}}{C} = (1-\tau_K)r - \rho; \tag{12.20}$$

and, in the steady-state equilibrium, we have

$$(1-\tau_K)r = m(E)^\delta. \tag{12.21}$$

Substituting $r = A(1-\gamma)(aH)^\gamma K^{-\gamma}$ in Eq. (12.21), we have

$$(\frac{H}{K})^{\gamma-\beta\delta} = \frac{m{E_0}^\delta}{(1-\tau_K)A(1-\gamma)a^\gamma}. \tag{12.22}$$

Equating the rate of growth of consumption to the rate of growth of human capital in the steady-state equilibrium, we have

$$m(E)^\delta a = \rho. \tag{12.23}$$

Substituting E and $\frac{H}{K}$ from Eqs. (12.4) and (12.22) in the Eq. (12.23), we have

$$a = [\frac{\rho\gamma^{-\beta\delta}(A(1-\tau_K)(1-\gamma))^{\beta\delta}}{m^\gamma(E_0)^{\delta\gamma}}]^{\frac{1}{(\gamma-\beta\delta-\gamma\beta\delta)}}. \tag{12.24}$$

Note that this equation is same as Eq. (12.18) with τ replaced by τ_K. a varies inversely with τ_K. Also note that a is independent of τl. Hence, $g = \rho(\frac{1}{a} - 1)$ also varies positively with τ_K and is independent of the change in τl.

Proposition 3 The balanced growth rate of the economy varies positively with the tax rate imposed on capital income and is invariant to the tax rate imposed on labour income.

If the tax rate imposed on physical capital income is increased, then the post-tax marginal productivity of capital is reduced. This lowers the rate of growth of consumption. So, the rate of growth of physical capital stock is reduced. As a result environmental quality is improved, and this, in turn, exerts positive external effect on human capital accumulation. Hence, the rate of growth of human capital is increased. In the steady-state growth equilibrium, this causes the rate of growth of output to rise. However, the change in the tax rate on labour income does not affect the marginal productivity of capital, and so it keeps the rate of growth unchanged.

---

[6] The optimality conditions are given in Appendix B.

5 Abatement Expenditure

The abatement expenditure is an important factor determining the quality of environment. In this section, we assume that abatement activity is undertaken by the government. Tax revenue is not distributed as lump sum payment. It is spent to meet the abatement expenditure denoted by S. Hence, $S = \tau Y$ and $P = 0$. Our modified environmental quality function is given by

$$E = E_0 K^{-\beta} H^{\beta-\theta} S^{\theta}, \tag{12.25}$$

with

$$\beta > \theta > 0.$$

The budget constraint of the household is given by

$$\dot{K} = (1-\tau)Y - C \tag{12.26}$$

which is same as Eq. (12.2) with $P = 0$. The dynamic optimization problem of the representative individual in this model is to maximize

$$\int_0^{\infty} U(C)e^{-\rho t}dt,$$

with

$$U(C) = lnC,$$

and subject to the Eqs. (12.1), (12.3), (12.25) and (12.26). S is treated as given in the optimization process because it is external to the individual in a competitive economy.

The optimality conditions remain same as obtained in Sect. 2, and, hence, these are represented again by Eqs. (12.5)–(12.8).

Here also, we obtain

$$\frac{\dot{x}}{x} = -A(1-\tau)\gamma a^{\gamma} y^{\gamma} - \rho + x. \tag{12.27}$$

Using Eqs. (12.1), (12.3), (12.25) and (12.26), we have

$$\frac{\dot{y}}{y} = mE_0{}^{\delta}(\tau A a^{\gamma})^{\theta\delta} y^{\delta(\beta+(\gamma-1)\theta)}(1-a) - A(1-\tau)a^{\gamma} y^{\gamma} + x. \tag{12.28}$$

Differentiating both sides of Eq. (12.11) with respect to time, t, and then using Eqs. (12.2), (12.3), (12.7) and (12.8), we have

$$\begin{aligned}\frac{\dot{a}}{a} = &\frac{\{1-\{(1-\gamma)(1-\theta\delta)+\delta\beta\}(1-a)\}}{\{(1-\gamma)(1-\theta\delta)\}} m(E_0)^{\delta}(\tau A a^{\gamma})^{\theta\delta} y^{\delta(\beta+(\gamma-1)\theta\delta))} \\ &+ A(1-\tau)a^{\gamma} y^{\gamma} \frac{\delta[\beta-\theta(1-\gamma)]}{\{1-\gamma(1-\theta\delta)\}} - \frac{\{(1-\gamma)(1-\theta\delta)+\delta\beta\}}{\{1-\gamma(1-\theta\delta)\}} x.\end{aligned} \tag{12.29}$$

In the steady-state equilibrium, from Eq. (12.27), we have

$$x^* = \rho + Aa^{*\gamma} y^{*\gamma}(1-\tau)\gamma.$$

From Eqs. (12.28) and (12.29), we have

$$y^* = [\frac{\rho}{(1-\gamma)(1-\tau)Aa^{*(\gamma+1)}}]^{\frac{1}{\gamma}};$$

and

$$a^* = [\frac{\rho^{\gamma-\beta\delta+\theta\delta(1-\gamma)}\{A(1-\gamma)(1-\gamma)\}^{(\beta\delta+\theta\delta(\gamma-1))}}{(mE_0{}^{\delta}A\tau^{\theta\delta)})}]^{\frac{1}{\gamma+\{\theta-(1+\gamma)\beta)\}\delta}}.$$

We have already assumed $\beta > \theta$, which implies $\beta > (1-\gamma)\theta$. Hence, a^* is negatively related to the tax rate and the growth rate, g, is positively related to the tax rate. Note that compared to Eq. (12.13), the negative effect of the change in the tax rate on a^* becomes stronger in this case. This is so because, firstly, physical capital accumulation has a direct negative effect on environment like that in the previous section. Moreover, abatement expenditure to physical capital ratio varies negatively with the stock of physical capital because abatement expenditure, being equal to tax revenue, is proportional to income and the average productivity of physical capital is diminishing. So, the negative effect of physical capital gets strengthened in this case.

Thus, we have the following proposition.

Proposition 4 Even if the environmental quality varies positively with the abatement expenditure, an increase in the tax rate on output raises the balanced growth rate of the economy.

6 Educational Subsidy

In this section, we assume that the tax revenue is spent to finance the educational subsidy only. The modified human capital accumulation function is given by

$$\dot{H} = mE^{\delta}(1-a)HG^{\phi}, \tag{12.30}$$

with $\phi > 0$. Here,

$$G = \frac{\tau Y}{H};$$

and G denotes the effectiveness of the educational subsidy that varies positively with the level of subsidy and inversely with the stock of human capital. If $\phi = 0$, we come back to Eq. (12.3) of the basic model in Sect. 2. Since G is proportional to Y and since Y varies positively with capital stock, K, the rate of human capital accumulation receives a positive external effect from physical capital accumulation.

The budget constraint of the household is given by Eq. (12.26). The representative individual maximizes

$$\int_0^\infty lnCe^{-\rho t}dt,$$

with respect to C and a subject to the Eqs. (12.1), (12.4), (12.26) and (12.30). G is treated as given in the maximization process.

We define again x and y such that $x = \frac{C}{K}$ and$y = \frac{H}{K}$. From the definitions of x and y and the optimality conditions,[7] we obtain following equations of motion:

$$\frac{\dot{x}}{x} = -A(1-\tau)\gamma a^\gamma y^\gamma - \rho + x, \tag{12.31}$$

$$\frac{\dot{y}}{y} = mE_0{}^\delta(\tau Aa^\gamma)^\phi y^{(\beta\delta+(\gamma-1)\phi)}(1-a) - A(1-\tau)a^\gamma y^\gamma + x, \tag{12.32}$$

and

$$\begin{aligned}\frac{\dot{a}}{a} = &\frac{\{1-\{(1-\gamma)(1-\phi)+\delta\beta\}(1-a)\}}{\{\phi\gamma+(1-\gamma)\}}m(E_0)^\delta(\tau Aa^\gamma)^\phi y^{(\beta\delta+(\gamma-1)\phi)} \\ &+ A(1-\tau)a^\gamma y^\gamma \frac{[\beta\delta-\phi(1-\gamma)]}{\{\phi\gamma+(1-\gamma)\}} - \frac{\{(1-\gamma)(1-\phi)+\delta\beta\}}{\{\phi\gamma+(1-\gamma)\}}x. \end{aligned} \tag{12.33}$$

The steady-state equilibrium values of the variables are given by

$$x^* = \rho + Aa^{*\gamma}y^{*\gamma}(1-\tau)\gamma;$$

$$y^* = [\frac{\rho}{(1-\gamma)(1-\tau)Aa^{*(\gamma+1)}}]^{\frac{1}{\gamma}};$$

and

$$a^* = [\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\tau)(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^\delta(A\tau)^\phi)^\gamma}]^{\frac{1}{\gamma+\phi-(1+\gamma)\beta\delta}}.$$

Here,

$$\frac{da^*}{d\tau} = -[\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^\delta A^\phi)^\gamma}]^{\frac{1}{\gamma+\phi-(1+\gamma)\beta\delta}}(1-\tau)^{\frac{\beta\delta(2+\gamma)+(\gamma-2)\phi-\gamma}{\gamma-(1+\gamma)\beta\delta+\phi}}$$

$$\tau^{\frac{(1+\gamma)\beta\delta-\phi-\gamma(1+\phi)}{\gamma-(1+\gamma)\beta\delta+\phi}}[\frac{(\beta\delta-\phi)\tau+\phi\gamma}{\gamma-(1+\gamma)\beta\delta+\phi}].$$

If $\beta\delta > (1-\gamma)\phi$, then a^* varies negatively with the tax rate. $\beta\delta > (1-\gamma)\phi$ implies that the negative effect of physical capital accumulation generated through environmental degradation outweighs its positive effect generated through subsidization to

[7] The optimality conditions are shown in Appendix C.

the human capital accumulation sector. However, if $\beta\delta < (1-\gamma)\phi$, then a^* is not monotonically related to the tax rate, τ. Here,

$$\frac{da^*}{d\tau} > (<)0 \quad \text{if } \tau > (<)\frac{\phi\gamma}{\phi - \beta\delta}.$$

So, the growth rate-maximizing (a^*minimizing)[8] tax rate is given by

$$\hat{\tau} = \frac{\phi\gamma}{\phi - \beta\delta} < 1.$$

This optimum tax rate, $\hat{\tau}$, varies positively with γ which represents the competitive output share of human capital. This is justified because the tax revenue is spent as educational subsidy. Here, $\hat{\tau} = \gamma$ when $\beta = \delta = 0$; and $\beta > 0$ implies $\hat{\tau} > \gamma$. So, in the absence (or presence) of the negative effect of environmental degradation on the human capital accumulation, the optimum tax rate is equal to (or greater than) the competitive output share of human capital.

Proposition 5 If the revenue obtained from the output tax is spent as educational subsidy and if $\beta\delta < (1-\gamma)\phi$, then there exists a unique positive tax rate, $\hat{\tau}$, satisfying $0 < \hat{\tau} < 1$ and maximizing the steady-state equilibrium growth rate. However, this growth rate varies positively with the tax rate when $\beta\delta > (1-\gamma)\phi$.

7 Growth and Social Welfare

If the social welfare is a positive function of the balanced growth rate, then there is no conflict between the growth rate maximization and the social welfare maximization. Fortunately, this is true for all the models described in the earlier sections.

Here, W stands for the level of social welfare. So,

$$W = \int_0^\infty lnCe^{-\rho t}dt = g\int_0^\infty te^{-\rho t}dt + lnC(0)\int_0^\infty e^{-\rho t}dt = \frac{lnC(0)}{\rho} + \frac{g}{\rho^2}$$

because $C(t) = C(0)e^{gt}$ along the balanced growth path. In Sects. 2 and 4, where the tax revenue is returned to the individuals as lump sum income transfer, we have $C(0) = Y(0) - gK(0)$. Hence, we have

$$\frac{dW}{dg} = \frac{1}{\rho^2} - \frac{K(0)}{\rho\{Y(0) - gK(0)\}} = \frac{1}{\rho^2}\left[\frac{\frac{C(0)}{K(0)} - \rho}{\frac{C(0)}{K(0)}}\right].$$

Here, $\frac{dw}{dg} > 0$ because

$$\frac{C(0)}{K(0)} - \rho = (mE_0{}^{\delta})^{\frac{-\gamma}{\beta\delta-\gamma+\gamma\beta\delta}}\rho^{\frac{\gamma\beta\delta}{\beta\delta-\gamma+\gamma\beta\delta}}(A(1-\tau)(1-\gamma))^{\frac{\beta\delta}{\beta\delta-\gamma+\gamma\beta\delta}}$$
$$\left[\frac{1-(1-\tau)(1-\gamma)}{(1-\tau)(1-\gamma)}\right] > 0.$$

[8] Second-order condition is also satisfied. See Appendix D.

In Sects. 5 and 6, we have $C(0) = (1 - \tau)Y(0) - gK(0)$. In this case also, we have

$$\frac{dW}{dg} = \frac{1}{\rho^2}[\frac{\frac{C(0)}{K(0)} - \rho}{\frac{C(0)}{K(0)}}]$$

In these two cases, along the steady-state equilibrium growth path, $\frac{C(0)}{K(0)}$ is given by

$$\frac{C(0)}{K(0)} = \rho + \frac{\gamma\rho}{(1-\gamma)a^*}.$$

Hence, in Sect. 5, $\frac{C(0)}{K(0)}$ is given by

$$\frac{C(0)}{K(0)} - \rho = \frac{\gamma\rho}{(1-\gamma)}[\frac{\rho^{\gamma-\beta\delta+\theta\delta(1-\gamma)}\{A(1-\tau)(1-\gamma)\}^{(\beta\delta+\theta\delta(\gamma-1))}}{(mE_0{}^{\delta}A\tau^{\theta\delta)})^{\gamma}}]^{\frac{-1}{\gamma+\{\theta-(1+\gamma)\beta)\}\delta}} > 0$$

and in Sect. 6, $\frac{C(0)}{K(0)}$ is given by

$$\frac{C(0)}{K(0)} - \rho = \frac{\gamma\rho}{(1-\gamma)}[\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\tau)(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^{\delta}(A\tau)^{\phi})^{\gamma}}]^{\frac{-1}{\gamma+\phi-(1+\gamma)\beta\delta}} > 0.$$

Hence, $\frac{dw}{dg} > 0$. So the social welfare along the balanced growth path varies positively with the balanced growth rate, g. Hence, there is no difference between the growth rate-maximizing tax rate and the social welfare-maximizing tax rate. However, we cannot derive the socially optimal tax rate along the transitional growth path.[9]

8 Conclusion

We have developed an endogenous growth model where the environmental quality varies negatively with the stock of physical capital and varies positively with the size of human capital. The rate of human capital accumulation is positively affected by the external effect emanating from environment. The interesting results obtained in this model are as follows. Firstly, the steady-state equilibrium rate of growth, in this model, varies positively with the proportional tax rate imposed on output or on capital income when tax revenue is spent as lump sum payment. This result holds even if the environmental quality is positively related to the abatement expenditure and the entire tax revenue is spent as abatement expenditure. The optimum (growth rate maximizing) tax rate may appear to be positive and less than unity when the tax revenue is spent as educational subsidy. However, this rate of growth is independent of the tax rate imposed on labour income. In Lucas (1988), this rate of growth is independent of the tax rate imposed either on output or on capital income. In Rebelo (1991), Mohtadi (1996), etc., the rate of growth varies inversely with the tax rate.

[9] The detailed derivations are shown in Appendix E.

Garcia-Castrillo and Sanso (2000), Gomez (2003), etc. find the optimal physical capital tax rate to be zero and the optimal labour tax rate to be positive in the Lucas (1988) model when tax revenue is spent as educational subsidy. None of these models considers the negative effect of environmental degradation on the human capital accumulation. However, in this model, we have considered the negative effect of environmental degradation on the human capital accumulation and have shown that the steady-state equilibrium growth rate would receive a positive effect from taxation either on output or on capital income. Existing literature does not point out such a possibility.

9 Appendix A

Here, the Jacobian matrix corresponding to the system of differential Eqs. (12.16)–(12.18) is given by

$$J = \begin{bmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial y} & \frac{\partial \dot{x}}{\partial a} \\ \frac{\partial \dot{y}}{\partial x} & \frac{\partial \dot{y}}{\partial y} & \frac{\partial \dot{y}}{\partial a} \\ \frac{\partial \dot{a}}{\partial x} & \frac{\partial \dot{a}}{\partial y} & \frac{\partial \dot{a}}{\partial a} \end{bmatrix};$$

and the elements of the Jacobian matrix evaluated at the steady-state equilibrium values of the variables are given as follows:

$$\frac{\partial \dot{x}}{\partial x} = x^*;$$

$$\frac{\partial \dot{x}}{\partial y} = Aa^{*\gamma} x^* \gamma y^{*y-1}\{(1-\tau)(1-\gamma)-1\};$$

$$\frac{\partial \dot{x}}{\partial a} = Aa^{*\gamma-1} x^* \gamma y^{*\gamma}\{(1-\tau)(1-\gamma)-1\};$$

$$\frac{\partial \dot{y}}{\partial x} = y^*;$$

$$\frac{\partial \dot{y}}{\partial y} = m(E_0)^{\delta}\,(1-a)\,\beta\delta y^{\beta\delta} - Aa^{\gamma}\gamma y^{\gamma};$$

$$\frac{\partial \dot{y}}{\partial a} = -A\gamma a^{\gamma-1} y^{\gamma+1} - m(E_0)^{\delta} y^{*(\beta\delta+1)};$$

$$\frac{\partial \dot{a}}{\partial x} = -\frac{(1-\gamma+\beta\delta)a^*}{(1-\gamma)};$$

$$\frac{\partial \dot{a}}{\partial y} = \frac{\{1-(1-\gamma+\beta\delta)(1-a)\}}{(1-\gamma)} m(E_0)^{\delta} y^{\beta\delta-1} a\beta\delta + \{\tau + \frac{\beta\delta}{(1-\gamma)}\}Aa^{\gamma+1}\gamma y^{\gamma-1};$$

and

$$\frac{\partial \dot{a}}{\partial a} = \frac{(1-\gamma+\beta\delta)}{(1-\gamma)} m(E_0)^{\delta} y^{\beta\delta} a + \{\tau + \frac{\beta\delta}{(1-\gamma)}\} A\gamma a^{\gamma} y^{\gamma}.$$

The characteristic equation of the J matrix is given by

$$|J - \lambda I_3| = 0,$$

where λ is an Eigen value of the Jacobian matrix with elements being evaluated at the steady-state equilibrium values. The three charateristic roots can be solved from the equation

$$a_0\lambda^3 + b_0\lambda^2 + a_1\lambda + b_1 = 0,$$

where

$$a_0 = -1,$$

$$b_0 = \text{Trace of } J,$$

$$a_1 = \text{sum of the minors of diagonal terms of } J$$

and

$$b_1 = \text{Determinant of } J$$

Clearly, a_0 is negative. We can derive that

$$\begin{aligned}
b_0 &= \text{Trace of } J \\
&= J_{xx} + J_{aa} + J_{yy} \\
&= x^* + \frac{(1-\gamma+\beta\delta)}{(1-\gamma)} m(E_0)^{\delta} y^{\beta\delta} a + \{\tau + \frac{\beta\delta}{(1-\gamma)}\} A\gamma a^{\gamma} y^{\gamma} \\
&\quad + m(E_0)^{\delta}(1-a)\beta\delta y^{\beta\delta} - Aa^{\gamma}\gamma y^{\gamma} \\
&= x^* + A(a^* y^*)^{\gamma}\gamma[\tau + \frac{\beta\delta}{(1-\gamma)} - 1] \\
&\quad + m(E_0 y^{\beta})^{\delta}[(1-a^*)\beta\delta + \frac{(1-\gamma+\beta\delta)}{(1-\gamma)} a^*].
\end{aligned}$$

Also, it can be shown that

$$\begin{aligned}
b_1 &= \text{Determinant of } J \\
&= J_{xx}[J_{yy}J_{aa} - J_{ya}J_{ay}] - J_{xy}[J_{yx}J_{aa} - J_{ya}J_{ax}] + J_{xa}[J_{yx}J_{ay} - J_{yy}J_{ax}] \\
&= m^2 {E_0}^{2\delta} a(1-a) y^{2\beta\delta} x\beta\delta \frac{(\beta\delta+1-\gamma)}{(1-\gamma)} - Am{E_0}^{\delta} a^{\gamma+1} y^{\gamma+\beta\delta} x(1-\tau)\gamma(1-\gamma) \\
&\quad + m{E_0}^{\delta} y^{\beta\delta+1} x[m{E_0}^{\delta}\beta\delta y^{\beta\delta-1} a \frac{\{1-(\beta\delta+1-\gamma)(1-a)\}}{(1-\gamma)} \\
&\quad + A\gamma a^{\gamma+1} y^{\gamma-1}(1-\tau)(\beta\delta-\gamma)]
\end{aligned}$$

From the steady-state equilibrium values x^*, y^* and a^* obtained from Eqs. (12.16)–(12.18), we have the following equations:

$$Aa^{*\gamma+1}y^{*\gamma} = \frac{\rho}{(1-\tau)(1-\gamma)};$$

and

$$mE_0{}^{\delta}y^{*\beta\delta} = \frac{Aa^{*\gamma}y^{*\gamma}(1-\tau)(1-\gamma)-\rho}{(1-a^*)}.$$

Using the above equations, we have

$$b_1 = \text{Determinant of } J = \frac{mE_0{}^{\delta}y^{*\beta\delta}x^*\rho}{(1-\gamma)}[\beta\delta(1+\gamma)-\gamma]$$

10 Appendix B

The current value Hamiltonian function is given by

$$Z = lnC + \mu_K[(1-\tau_K)rK + (1-\tau_l)waH - C + P] + \mu_H[m(E)^{\delta}(1-a)H],$$

where μ_K and μ_H are the costate variables.

The first-order optimality conditions are given by the following:

$$\frac{\partial Z}{\partial C} = \frac{1}{C} - \mu_K = 0, \tag{B.1}$$

and

$$\frac{\partial Z}{\partial a} = \mu_K(1-\tau_l)wH - \lambda_H m(E)^{\delta}H = 0. \tag{B.2}$$

Time behaviour of the costate variables along the optimum growth path should satisfy the following:

$$\dot{\mu}_K = \rho\mu_K - \mu_K(1-\tau_K)r, \tag{B.3}$$

and

$$\dot{\mu}_H = \rho\mu_H - \mu_K(1-\tau_l)wa - \mu_H m(E)^{\delta}(1-a) \tag{B.4}$$

Transversality conditions are given by the following:

$$\lim_{t\to\infty} e^{-\rho t}\lambda_K(t)K(t) = \lim_{t\to\infty} e^{-\rho t}\lambda_H(t)H(t) = 0$$

From Eqs. (B.1) and (B.3), we have

$$\frac{\dot{C}}{C} = \frac{(1-\tau_K)r-\rho}{\sigma}.$$

From Eq. (B.2), we have

$$\frac{\mu_K}{\mu_H} = \frac{mE^\delta}{(1-\tau_l)w}.$$

Since w is constant in the steady state, we have

$$\frac{\dot{\mu}_K}{\mu_K} = \frac{\dot{\mu}_H}{\mu_H}.$$

From the above equation, we have

$$(1-\tau_K)r = m(E)^\delta.$$

11 Appendix C

The dynamic optimization problem of the representative individual in this model is to maximize

$$\int_0^\infty U(C)e^{-\rho t}dt$$

with

$$U(C) = lnC$$

subject to the Eqs. (12.1), (12.4), (12.26) and (12.30). While maximizing their present discounted value of utility the individual would consider G to be given.

The first-order optimality conditions are given by the following:

$$\frac{\partial Z}{\partial C} = \frac{1}{C} - \lambda_K = 0, \tag{C.1}$$

and

$$\frac{\partial Z}{\partial a} = \lambda_K A(1-\tau)\gamma K^{1-\gamma}a^{\gamma-1}H^\gamma - \lambda_H m(E)^\delta H(\frac{\tau Y}{H})^\phi = 0. \tag{C.2}$$

Time behaviour of the costate variables along the optimum growth path should satisfy the following:

$$\dot{\lambda}_K = \rho\lambda_K - \lambda_K A(1-\tau)(1-\gamma)(aH)^\gamma K^{-\gamma} \tag{C.3}$$

and

$$\dot{\lambda}_H = \rho\lambda_H - \lambda_K A(1-\tau)\gamma(a)^\gamma H^{\gamma-1}K^{1-\gamma} - \lambda_H m(E)^\delta\,(1-a)\,(\frac{\tau Y}{H})^\phi. \tag{C.4}$$

From Eqs. (C.1) and (C.3), we have

$$\frac{\dot{C}}{C} = (1-\tau)(1-\gamma)A(aH)^{\gamma}K^{-\gamma} - \rho. \tag{C.5}$$

From Eq. (C.2), we have

$$\frac{\lambda_K}{\lambda_H} = \frac{mE^{\delta}(\frac{\tau Y}{H})^{\phi}}{(1-\tau)A\gamma a^{\gamma-1}H^{\gamma-1}K^{1-\gamma}}.$$

Taking logarithm and differentiating both sides of the above-mentioned equation with respect to time, t, and then using Eqs. (12.26), (12.30), (C.3) and (C.4), we have

$$\frac{\dot{a}}{a} = \frac{\{1-\{(1-\gamma)(1-\phi)+\delta\beta\}(1-a)\}}{\{\phi\gamma+(1-\gamma)\}} m(E_0)^{\delta}(\tau A a^{\gamma})^{\phi}(\frac{H}{K})^{(\beta\delta+(\gamma-1)\phi)}$$

$$+ A(1-\tau)a^{\gamma}(\frac{H}{K})^{\gamma}\frac{[\beta\delta-\phi(1-\gamma)]}{\{\phi\gamma+(1-\gamma)\}} - \frac{\{(1-\gamma)(1-\phi)+\delta\beta\}}{\{\phi\gamma+(1-\gamma)\}}\frac{C}{K}. \tag{C.6}$$

Using Eqs. (12.26) and (C.5), we obtain Eq. (12.31). Using Eq. (C.6) and the definitions of x and y, we obtain Eq. (12.33). Using Eqs. (12.26) and (12.30), we obtain the Eq. (12.32).

12 Appendix D

The steady-state equilibrium value of a is

$$a^* = [\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\tau)(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^{\delta}(A\tau)^{\phi})^{\gamma}}]^{\frac{1}{\gamma+\phi-(1+\gamma)\beta\delta}}.$$

Here,

$$\frac{da^*}{d\tau} = -[\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^{\delta}A^{\phi})^{\gamma}}]^{\frac{1}{\gamma+\phi-(1+\gamma)\beta\delta}}(1-\tau)^{\frac{\beta\delta(2+\gamma)+(\gamma-2)\phi-\gamma}{\gamma-(1+\gamma)\beta\delta+\phi}}$$

$$\tau^{\frac{(1+\gamma)\beta\delta-\phi-\gamma(1+\phi)}{\gamma-(1+\gamma)\beta\delta+\phi}}[\frac{(\beta\delta-\phi)\tau+\phi\gamma}{\gamma-(1+\gamma)\beta\delta+\phi}]$$

Hence,

$$\frac{da}{d\tau} = 0$$

when

$$\tau = \frac{\phi\gamma}{\phi - \beta\delta} = \tau*$$

$$\frac{d^2 a^*}{d\tau^2} at(\tau = \tau^*) = [\frac{\rho^{\gamma-\beta\delta+\phi(1-\gamma)}\{A(1-\gamma)\}^{(\beta\delta+\phi(\gamma-1))}}{(mE_0{}^{\delta}A^{\phi})^{\gamma}}]^{\frac{1}{\gamma+\phi-(1+\gamma)\beta\delta}}$$
$$\frac{\phi\gamma(\phi-\beta\delta-\phi\gamma)}{(\phi-\beta\delta)\{\gamma-(1+\gamma)\beta\delta+\phi\}}$$

$$(1-\tau*)^{\frac{\beta\delta(2+\gamma)+(\gamma-2)\phi-\gamma}{\gamma-(1+\gamma)\beta\delta+\phi}-1}\tau*^{\frac{(1+\gamma)\beta\delta-\phi-\gamma(1+\phi)}{\gamma-(1+\gamma)\beta\delta+\phi}-1}.$$

This is positive when $\beta\delta < (1-\gamma)\phi$.

13 Appendix E

In Sects. 2 and 4, we have

$$\dot{K} = Y - C.$$

Hence, along the steady-state equilibrium growth path,

$$\frac{\dot{K}}{K} = g = \frac{Y(0)}{K(0)} - \frac{C(0)}{K(0)}.$$

Now, from Eq. (12.1), we find that along the steady-state equilibrium growth path

$$\frac{Y(0)}{K(0)} = Aa^{*\gamma}(\frac{H}{K})^{*\gamma}.$$

Using Eq. (12.10), we have

$$g = A(1-\tau)(1-\gamma)a^{*\gamma}(\frac{H}{K})^{*\gamma} - \rho.$$

Hence,

$$\frac{C(0)}{K(0)} = \frac{Y(0)}{K(0)} - g = \rho + Aa^{*\gamma}(\frac{H}{K})^{*\gamma}[1-(1-\tau)(1-\gamma)].$$

Substituting the values of a^* and $(\frac{H}{K})^*$ from Sect. 3, we obtain the expression of $(\frac{C(0)}{K(0)} - \rho)$ in Sect. 7.

In Sects. 5 and 6, we have

$$\dot{K} = (1-\tau)Y - C.$$

Hence,

$$\frac{C(0)}{K(0)} = (1 - \tau)\,\frac{Y(0)}{K(0)} - g.$$

In these sections also, we find that

$$g = A(1 - \tau)(1 - \gamma)a^{*\gamma}(\frac{H}{K})^{*\gamma} - \rho.$$

The expression of $\frac{Y(0)}{K(0)}$ remains same as above. Hence,

$$\frac{C(0)}{K(0)} - \rho = \frac{\gamma\rho}{(1 - \gamma)a^{*}}.$$

References

Alonso-Carrera J, Freire-Seren MJ (2004) Multiple equilibria, fiscal policy and human capital accumulation. J Econ Dyn Contr 28:841–856

Ben-Gad M (2003) Fiscal policy and indeterminacy in models of endogenous growth. J Econ Theory 108:322–344

Benhabib J, Perli R (1994) Uniqueness and indeterminacy: on the dynamics of endogenous growth. J Econ Theory 63:113–142

Bretschger L, Smulders S (2007) Sustainable resource use and economic dynamics. Environ Resour Econ 36:1–13

Caballe J, Santos M (1993) On endogenous growth with physical and human capital. J Polit Econ 101:1042–1067

Chamley C (1993) Externalities and dynamics in models of learning or doing. Int Econ Rev 34: 583–610

Clarke M, Maantay J (2006) Optimizing recycling in all of New York City's neighborhoods: using GIS to develop the REAP index for improved recycling education, awareness, and participation. Resour Conserv Recycl 46(2):128–148

Dinda S (2005) A theoretical basis for environmental Kuznets Curve. Ecol Econ 53:403–413

Endress LH, Roumasset JA, Zhou T (2005) Sustainable growth with environmental spillovers. J Econ Behav Org 58:527–547

Faig M (1995) A simple economy with human capital: transitional dynamics, technology shocks, and fiscal policies. J Macroecon 17:421–446

Garcia-Castrillo P, Sanso M (2000) Human capital and optimal policy in a Lucas-type model. Rev Econ Dyn 3:757–770

Gradus R, Smulders S (1993) The tradeoff between environmental care and long term growth—pollution in three prototype growth models. J Econ 1:25–51

Gomez MA (2003) Optimal fiscal policy in the Uzawa-Lucas model with externalities. Econ Theory 22:917–925

Goulder LH, Mathai K (2000) Optimal CO2 abatement in the presence of induced technological change. J Environ Econ Manage 39:1–38

Greiner A, Semmler W (2002) Externalities of investment, education and economic growth. Econ Model 19:709–724

Grimaud A (1999) Pollution permits and sustainable growth in a Schumpeterian Model. J Environ Econ Manage 38(3):249–266

Grimauda A, Tournemaineb F (2007) Why can an environmental policy tax promote growth through the channel of education? Ecol Econ 62:27–36

Hart R (2004) Growth, environment and innovation: a model with production vintages and environmentally oriented research. J Environ Econ Manage 48(3):1078–1098

Hettich F (1998) Growth effects of a revenue-neutral Environmental Tax Reform. J Econ 67(3): 287–316

Kauppi L (2006) Mercury-background for abating mercury pollution-presentation for the International Mercury Conference, Brussels, 26–27 Oct 2006

Lucas RE (1988) On the mechanics of economic development. J Monet Econ 22:3–42

Lucas RE (1990) Supply-side economics: an analytical review, Oxford Eco Papers, 1990, 42(2): 293–316

Margulis S (1992). Back-of-the-envelope estimates of environmental damage costs in Mexico, Policy Research Working Papers, Agriculture Operations, Country Department II, Latin America and the Caribbean Regional Office, The World Bank, WPS 824

Mino K (1996) Analysis of a two-sector model of endogenous growth with Capital Income Taxation. Int Econ Rev 37:227–51

Mohtadi H (1996) Environment, growth and optimal policy design. J Public Econ 63:119–140

Ortigueira S (1998) Fiscal policy in an endogenous growth model with human capital accumulation—transitional dynamics with multiple equilibria. J Monet Econ 42:323–355

Perez R, Ruiz J (2007) Global and local indeterminacy and optimal environmental public policies in an economy with public abatement activities. Econ Model 24:431–452

Petrosillo I, Zurlini G, Corlianò ME, Zaccarelli N, Dadamo M (2007). Tourist perception of recreational environment and managemAuthor: ent in a marine protected area. Landscape and Urban Plan 79(1):29–37

Rebelo S (1991) Long-run policy analysis and long run growth. J Polit Econ 99:500–521

Ricci F (2007) Channels of transmission of environmental policy to economic growth: a survey of the theory. Ecol Econ 60:688–699

Rosendahl KE (1996) Does improved environmental policy enhance economic growth? Environ Resour Econ 9:341–364

Torras M, Boyce J (1998) Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve. Ecol Econ 25:147–160

Chapter 13
Labour Supply Schedule of the Poor: A Commonsense Approach

Pradipta Chaudhury

1 The Problem

When I lived in Kolkata during the early 1980s, I was struck by the poverty, paucity of employment opportunities for adults and the pervasive use of child labour. A large amount of open as well as disguised unemployment of unskilled and low-skilled labour was evident. At the same time, a large number of children were in the labour force instead of attending schools. A large number of poor women worked, while their small children and even babies lay neglected at their work sites or in their homes. These phenomena puzzled me a great deal as they did not gel with the labour supply curve given in textbooks of economics.

This chapter began as an attempt to understand such phenomena. Its first draft, deriving a downward-sloping labour supply curve of the poor from some striking facts and common sense, and drawing some of its implications, was presented in a conference at the Department of Economics, Presidency College, Kolkata, in March 1987 (Chaudhury 1987). Since then, over a quarter century, the issues have widened in scope as my understanding has broadened (Chaudhury 2008). In the 1990s, I began using the simple idea of the downward-sloping labour supply curve of the poor in a backward/traditional economy to understand the empirical relation between caste and class in India for the first half of the twentieth century while using indirect evidence.[1]

The basic problem and some of my understanding were formulated during my tenure as a Research Associate in the Department of Economics of the Presidency College, Kolkata. It was my first employment. Biswajit was a colleague. He became my first friend in the profession. We have remained friends since then. He is a remarkable man, endowed with many admirable qualities. I have great pleasure in writing this paper in his honour.

[1] Some results of my analysis of caste and class are published in Chaudhury 2004.

P. Chaudhury (✉)
Centre for Economic Studies and Planning, School of Social Sciences,
Jawaharlal Nehru University, New Delhi, India
e-mail: pradipta.chaudhury@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_13,

This chapter emphasizes the dissimilarities between the realities of poor societies and the economic conditions analysed in the textbooks while drawing a labour supply schedule. There are three fundamental differences: (1) In poor societies, the family is the unit of labour supply, not an individual. (2) The first objective of a poor family is to meet its subsistence needs by selling labour. This gives rise to the idea of a target income, which in turn leads to a rectangular hyperbola shape of the family's labour supply curve. (3) The downward slope of the labour supply schedule is for low wages, unlike the backward-bending part of a labour supply curve in a standard textbook, which is for high wages.

There are some interesting implications of these three ideas; (1) and (2) lead to a typical sequence in the supply of labourers: Usually, in a poor family first the male, then the female and then the child join the labour force. Consequently, we observe a corresponding wage differential: male wage rate is higher than female wage rate which in turn is higher than the child wage rate, even when there is no difference in the productivity. Under such conditions, a rise in the wage rate, or an increase in employment opportunities for adults, caused either by economic change or by public policy and funding leads to a decline in supply of labour, particularly female and child labour. This may be unwelcome for the employers who are enamoured by the abundance of cheap labour. But more and regular adult employment and better wages will improve education and nutrition which will lead to increases in skills and efficiency.

In the following section, some conspicuous facts suggesting the shape of the labour supply schedule of the poor are discussed. In Sect. III, the labour supply schedule is derived from these facts and common sense; several implications of the curve are drawn. In Sect. IV, our discussion is extended to the issue of differential wage rates of male, female and child labour in a traditional or backward economy. Section V summarises the central message of the chapter.

2 Empirical Basis of the Central Idea

In many economies including India, labour is extremely cheap. What makes labour so cheap? The standard answer is that there is excess supply of labour. But that leads us to more questions: Why is supply of labour so large? Why is unemployment so high? Why are the lowest wage rates so low? Unemployment has been examined and theorised by economists for long, while the existence of open as well as disguised unemployment in traditional agriculture has been questioned both on theoretical and on empirical grounds. On the one hand, there is a whole body of influential literature on development economics, following Arthur Lewis (1954), which takes the existence of (unlimited) surplus labour in traditional agriculture as a basic premise. On the other hand, there are others, led by Theodore Schultz, who see no theoretical or empirical basis for its existence.

For Schultz (1964), the issue is essentially technological: If labour is surplus in traditional peasant agriculture, then its marginal product must be zero. However, each unit of labour, which is used in agricultural operations, must make a positive

Table 13.1 Percentage of dependents in total population (% of dep. in populn.) in an occupation, and the number of female workers per 1,000 male workers (Fem:1,000 male workers) in the occupation in Uttar Pradesh (UP) in 1911. (Source: Census of India 1911, Vol. 15, Pt. 1, Report, pp. 401–402.)

Occupation	% of dep. in populn.	Fem:1,000 male worker
Law	72	
Public administration	67	
Trade in textiles	62	
Proprietors	61	375
Income from land	60	335
Industries connected with luxury (mainly goldsmiths)	59	160
Cultivators	51	402
Milk sellers	41	1,293
Washermen	41	825
Oil pressers	41	
Indoor servants	39	752
Trade in betel, etc.	36	1,364
Scavengers, etc.	35	1,340
Grain parchers	35	1,249
Farm servants and field labourers	35	845
Trade in grass, etc.	33	1,353
Flour grinders (largely old women)	22	1,977
Raising of livestock (majority are herdsmen, largely male children)	19	

contribution to output. Otherwise, it would not be employed. Furthermore, according to Schultz, it is not possible that the use of an additional unit of labour will not increase output in some way or the other. However, Amartya Sen (1966) has shown that the marginal product of labour being equal to zero is neither a necessary nor a sufficient condition for the existence of surplus labour in traditional peasant agriculture. According to Sen, surplus labour exists if the utility (from income or output) and disutility (from work) functions of workers belonging to peasant families are of a particular type or meet certain conditions. Thus, for Sen, the issue is essentially psychological; economic conditions do not seem to matter. I did not find this state of knowledge fully satisfying. Something vital was amiss if the material conditions were not crucially and explicitly relevant to the issue of supply of labour.

Some conspicuous evidence which is crude but highly suggestive helped to understand the phenomena better. In the Census of Uttar Pradesh (UP) of 1911, a close link between the occupations and the work participation rates (WPRs) or worker to dependent ratios is evident. Occupations associated with higher (lower) incomes have higher (lower) dependent to worker ratios. Furthermore, in poorer occupations, the workforce contains a greater proportion of females (see Table 13.1). The occupations with the highest percentage of dependents are law, public administration, trade in textiles, proprietors, income from land and industries connected with luxury (mainly goldsmiths) which are 'occupations followed chiefly by the upper classes, whose women do not work and whose children are at school, so that dependents are numerous' (Census of India 1911, Vol. 15, Pt. 1, p. 401).

In the same Census of UP, the poorer eastern districts show a considerably higher WPR, higher number of workers per unit of area cultivated and lower wage rates compared to the better-off western districts of the state. This is so from about 1880. But it was the reverse earlier. Before 1850, eastern UP was more prosperous. Deindustrialisation under aggressive colonial policies and a decline of river trade in the east led to its poverty and an excess supply of labour (in agriculture). Similarly, during that period, there is a contrast between the prosperous eastern Bengal districts and the districts of Gangetic Bihar. Districts of eastern Bengal showed lower WPRs and higher wage rates compared to the Bihar districts which were supplying labour to the industries of western Bengal as well as the agricultural fields of eastern Bengal.[2]

Thus, it appears that the worker to dependent ratio in a household varies inversely with its economic status. In fact, we notice that in very poor families, all the able-bodied male adults, women and even children join the labour force. On the other hand, in lower middle-class families, children attend school and women are mostly confined to domestic work. This evidence, though of an indirect nature, suggests a negative slope of the labour supply schedule of the poor. Obviously, there is a difference between labourer and labour unit (or labour time). Labour supply is measured in terms of the latter, but the evidence cited here is in terms of the former. Even then, this evidence suggests that when the average earnings of labour are lower, the intended supply of labour is higher. As and when the wage rate (or the average earnings of workers in a family) falls, the household has to augment its labour supply in order to maintain its income level. There are alternative ways of increasing the supply of labour. The existing workers may work longer and/or dependents may be turned into workers. Women, hitherto confined to domestic work, can be sent to take up wage work and/or children can be withdrawn from schools and sent to work.

3 Labour Supply Curve of the Poor

Now, let us begin from the fundamentals. Unemployment or excess supply of labour or surplus labour would exist if labour supply exceeds demand for labour. Why does supply of labour typically exceed demand in traditional agriculture or in poor economies? Why is labour supply so large as to include child labour? This leads us to an investigation into the determinants of the supply of labour in traditional agriculture or poor economies. It is amazing that none of the theorists seem to notice that the overwhelming numbers of victims of unemployment, open and disguised, in traditional agriculture belong to particular income classes, namely the landless labourers and the small peasants. It is imperative that we discuss the labour supply schedule of the poor and explore its implications.

A large section of our society is on or around the 'subsistence' level. In this section, we can broadly distinguish between two groups. One group comprises property-less

[2] See Chaudhury 1992 (Table IV.5 and Table IV.6). Also see Census of India 1911; Vol. 5, Bengal, Bihar and Orissa, Pt. 2 Table XV, Pt. 3 Table XV, and Vol. 15, U.P., Pt. 2 Table XV.

labour households, entirely dependent on the labour market for a livelihood. The other group consists of households that earn a part of their livelihood by means of cultivation of their own farm, sharecropping, handcrafts, etc. and part of their livelihood from the labour market.

While discussing the labour supply curve of the poor, what should be taken as the unit of analysis? The textbooks uniformly take the individual as the unit of labour supply. This is clearly inappropriate in our context. Typically, a child or even a poor woman does not enter the labour market only on his\her own wish. It is the family or its head that takes such decisions. Therefore, it is appropriate to consider the household, rather than the individual, as the labour supplying unit.[3] The reason for making this departure from textbook economics should be clear at least to those familiar with the Indian situation.

Another departure we should make from the textbook is with respect to the factors which influence the decision to work or regarding the motivation behind work. In textbooks, it is stated that an individual has a choice between labour and leisure and she/he works because the utility from income is higher than the disutility from the effort. Some amounts of leisure, for example, sleep and rest, is necessary for the human body to keep it functioning. Beyond that, a poor household does not have the ability to choose leisure over work. Some or all of its members may have to work in order that the family stays alive. Thus, it is asset poverty or inadequacy of income from land and other assets that drives the poor to work or sell labour power.

Let us consider a household belonging to the first group, comprising of property-less labour households. In any period, say 1 year,[4] it would try to earn a certain minimum level of income. This level of income, which can be designated a target income[5] (M), is determined by biological, demographic, cultural and social factors. M may be thought of as what the household considers a 'decent' level of income. However, in reality, the household may or may not achieve its target income. In the not-so-unlikely event of the household's actual income falling short of its target income, it may still attain the basic minimum subsistence level of living, if the gap is not large. However, when its realised income falls considerably short of its target income, the household is likely to be reduced to a miserable existence below the subsistence level or the poverty line. Usually, a property-less household finds it extremely difficult to borrow, since they cannot offer any collateral which is attractive to a lender. Consequently, its members suffer from malnutrition, morbidity and other malaise.

---

[3] The head of the family may decide which member is to seek what type of employment.

[4] One can also take a month as the unit period. But because of the importance of agriculture in our economy and other factors, 1 year seems to be a better proposition.

[5] A Marathi woman, aged 30, with two children to look after, employed in a cotton mill, told the Indian Factory Commission of 1890 that she worked because the income of her husband, a porter in the Railways, was not 'sufficient'. The working hours then were much longer than at present. Having joined the factory, she had no option of working for fewer hours. However, she expressed a preference for 'less hours and less pay'. See The Indian Factory Commission 1980.

A property-less household tries to earn its target income by selling the labour power of its members in the labour market. Usually, the workforce of a poor household consists of male, female and even child members. The wage rates of male, female and child labourers differ even when the workers do the same type and the same amount of work. Furthermore, for the workers of the same sex and age group, the wage rate may not be uniform due to difference in working skills. Thus, a family of labourers does not face a uniform wage rate. However, to keep matters simple, let us assume that the labour supplied by the household is of uniform quality. This assumption makes it easy for us to obtain our first and the most powerful insight into the nature of the labour supply schedule of the poor.

Let W be the wage rate of labour (of uniform quality) supplied by the household.[6] Let L be the amount of labour[7] the household has to supply during a period, in order to achieve its target income M.

Therefore,

$$L = M/W. \tag{13.1}$$

From (1), it follows that the lower (higher) the wage rate (or the average earnings of labour), the larger (smaller) would be the amount of labour supply required to attain the target income. Thus, the labour supply schedule of a property-less household is a rectangular hyperbola, as shown in Fig. 13.1. It should be obvious that the negative slope of the curve is not contingent upon our simplifying assumption regarding the quality of labour.

The continuity of the labour supply schedule, as drawn in Fig. 13.1, implies that labour can be supplied in divisible units. Actually, whether labour supply can be in divisible or indivisible units depends, to some extent, on the demand conditions. In modern factories and offices, labour is supplied largely in indivisible units of a certain number of hours per day for a fixed number of days in a week and throughout the year. In such a case, the worker faces a zero to one choice between working for a certain amount of time in a year and not working at all. Likewise, an attached farm servant employed by a landowner is normally engaged for a year; he cannot vary his labour supply within that period.

However, there are reasons for making the assumption of continuity of the labour supply schedule of a poor family in a developing economy. Such a household can vary its total supply of labour, at least within certain limits, while some of its working members may not face such a choice. Consider domestic service undertaken by

---

[6] Alternatively, W may be treated as the average earnings of labour belonging to the household, to do away with the assumption of uniform quality of labour. But this would complicate matters in other ways.

[7] L is measured in labour units. A working day of an average length may be taken as one unit of labour. In agriculture, an average working day has 10 h (if not more), whereas in modern industry it has about 8 h. This difference in the length of working days becomes relevant when we specify the household as completely rural or as completely urban, or as one which combines rural and urban occupations. Also, see the discussion in the text regarding possible variation in the amount of labour supply.

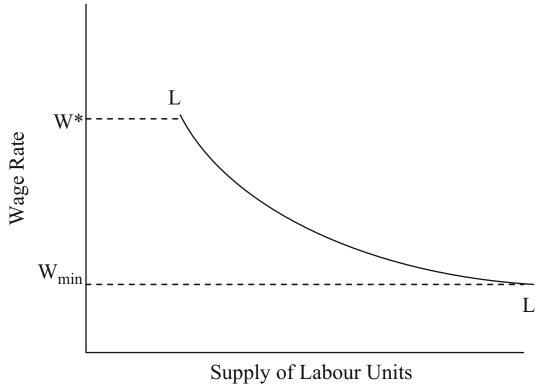


Fig. 13.1 The labour supply schedule of a poor household

females as well as males. Though it is indivisible to some extent, say for 1 or 2 hours per day for one employer, by choosing the number of households to work for, a worker can still vary the supply of his/her labour hours. There is also another kind of work, namely, piecework (for example, bidi making) where the supply of labour hours can be varied. The same applies to porters, rickshaw pullers, etc., who can, within limits, choose the number of their working hours. Besides, there are types of employment, such as construction work, casual labour, etc., where the supply of labour days by a worker can be varied.[8] Thus, the continuity of the labour supply schedule is more likely in the context of a poor household in a developing economy than for an individual in an advanced industrialised economy which is depicted in the textbooks.

As and when the wage rate (or the average earning per labour unit) falls, the family's income and its level of nutrition are likely to fall. Consequently, on the one hand, the physical ability of the household to supply labour declines, while on the other hand, its economic need to supply labour actually increases. If the wage rate (or the average earning per labour unit) were to fall below a certain (very low) level, the family's level of nutrition would decline below a critical minimum and, hence, the household would be unable to supply any labour. Thus, it is reasonable to assume a (positive) minimum wage rate (W_{min} in Fig. 13.1), below which the family's labour supply curve cannot be defined.

Conversely, one can visualise a (higher) rate (W^* in Fig. 13.1) such that, when the average earnings per labour unit (or the wage rate) approaches it from below, the supply of labour declines, but as the average earnings rise above that rate, the supply of labour does not fall further. That there is a reduction in the supply of labour as the earnings increase from an initial low level is evident from the fact that in very poor labour households children and women, in addition to the men, sell their labour in the market, whereas in lower middle-class families children invariably attend school and even women are mostly confined to domestic work. So long as the household

[8] Suppose the household wants to supply only 275 units of labour but the demand conditions are such that it can sell either 250 or 300 units. In such a situation, it may sell 300 units. In this connection also see footnote 5 earlier.

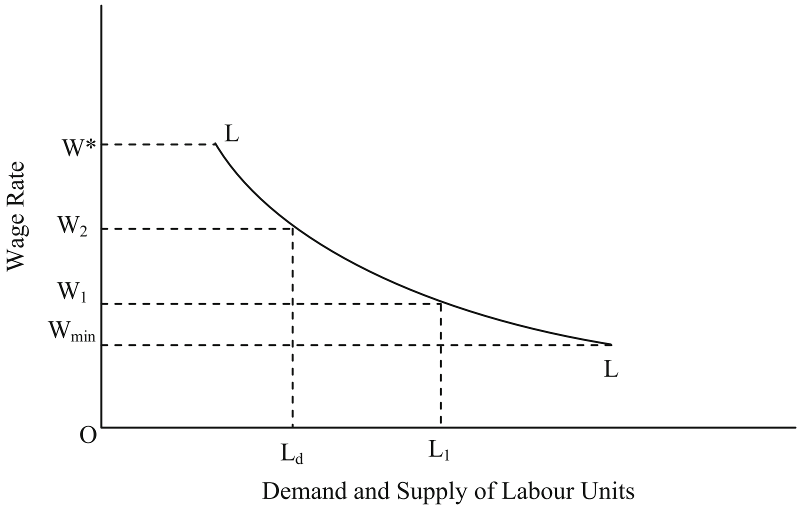


Fig. 13.2 Supply of labour, employment and excess supply in a pure labour household

does not revise its target income, its labour supply declines in response to an increase in the wage rate.

When a household manages to earn what it considers a decent income by supplying an amount of labour which is not excessive, suppose without having to sell the labour power of its children (and women), it is likely to revise its target income upwards. The family would like to acquire comforts first and then luxuries. Accordingly, in a situation where only the adult and able-bodied male members of the household are working for about 8 hours per day, it is unlikely that the family's labour supply would decline as the average earnings per labour unit (or the wage rate) increase.

As the wage rate increases beyond W*, whether labour supply will first increase and then decline is not our immediate concern here; the poor do not face such high wage rates. Those who face such high wage rates belong to entirely different socio-economic conditions. What we would like to stress here is that from common sense it appears that for wage rates below a certain wage rate (W*) the labour supply curve of a household slopes downwards. The poor, who constitute a large section of our society, evidently face wage rates below this (W*).

Let us begin by assuming that the prevailing market wage rate of the uniform quality of labour supplied by the household is W_1 (see Fig. 13.2). Thus, the household would like to sell OL_1 amount of labour in order to achieve its target income. However, suppose all the (male, female and child) workers of the household together can manage only OL_d ($< OL_1$) amount of employment. Then, this household has an excess supply of labour to the extent of $L_d L_1$. The excess supply may exist either in the form of partial unemployment of workers or in the form of complete unemployment of workers or in both forms. Note that this excess supply will disappear if the wage rate can be raised from W_1 to W_2, with no change in employment.

We can now extend our discussion to those families, which earn a part, not the whole, of their living by selling labour. It appears that the most significant constituents of this group are the small and very small peasant households who partly depend on the wage incomes of some of their members. Such a household, being poor, like a family which is a pure seller of labour, has a minimum target income. However, unlike the latter, it uses a part of its labour power to cultivate its own farm[9] and sells another part of its labour in the market. The rates of reward to labour are expected to be dissimilar for the two parts. However, if we characterize W of Fig. 13.1 as the mean income of labour of the family and maintain the convenient assumption of uniform quality of labour, the rectangular hyperbola in Fig. 13.1 would represent the supply curve of labour of this household. In addition, the ideas of W_{min} and W^* would still be applicable.

In contrast with a pure labour household, this family has to choose how much labour to apply in its home farm and how much labour to provide in the market. Till the marginal product of labour on the home-farm is more than the market wage rate, the family should first use its labour in the home farm.[10] After that, it would sell some labour in the market. If by doing both it reaches its target income, then it discontinues its supply of labour at that point. On the other hand, if after selling as much labour in the market as it can[11] its total income from both sources fails to reach its target income, it will apply more labour in its home farm (even if the marginal return to labour on home farm is below the market wage rate) till its target income is achieved. However, it is probable that after applying labour in the home farm to the technically maximum point (that is, to the point where marginal product of labour equals zero) in addition to selling some labour in the market, its actual income falls short of its target income. In such a condition, the family would intend to supply more labour in the market. But its plans cannot be realised because of insufficiency of demand. Hence, there will be some glut of labour.

Figure 13.3 graphically describes the situation of a cultivator-cum-labour household, which fails to realise its target income. LL is the labour supply curve. RR is the curve representing average returns to labour.[12] The flattest part of this curve, corresponding to the segment AB of the labour coordinate, refers to wage employment. After selling AB amount of labour in the market and using the technically maximum possible amount of labour ($= OA + BC$) in the home farm, that is, a total realised supply of OC units of labour, the household manages an average return of $OW_{1.}$ But the intended supply of labour corresponding to that average income is $OL_{1.}$ Thus, the household has an excess supply of labour to the extent of CL_1 which may exist either in the form of open or in the form of disguised unemployment or in both forms.

---

[9] It does not make any significant difference if this household cultivates leased-in land.

[10] If the market wage rate is higher than the return to the first unit of labour in the home farm, the household should first like to supply labour in the market.

[11] The household may not be able to sell all the labour it intends to, because of demand constraint.

[12] The average returns to labour curve is drawn with the following assumptions. The marginal product of labour on home farm declines as more labour is used and the household can sell a certain amount (AB in Fig. 13.3) of labour at a constant wage rate.

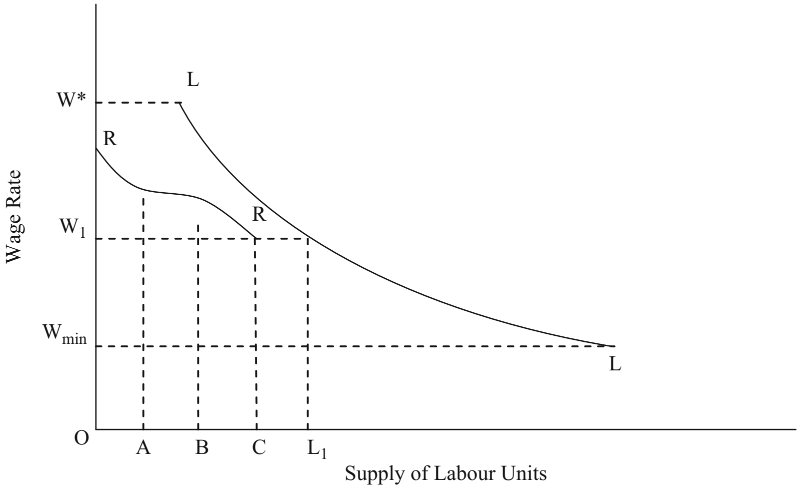


Fig. 13.3 Labour supply and average earnings of a cultivator-cum-labour household

The supply curve of labour links wage rate to the intended supply of labour time rather than labourers. Still, from this schedule, it is possible to infer a few things about the supply of labourers. At the current wage rate in the market, the supply of labour time can be known, given a household's target income to be earned from the labour market. When the wage rate is low, the number of labourers may have to be large for the following reasons. There is a limit to the labour time one worker can supply. In addition, low wages mean low nourishment, which implies diminished ability to work.

Regularity of employment or demand conditions also influences the number of labourers seeking work. Given the wage rate, if some or all workers can only get occasional employment, the number of workers is larger compared to a situation where all workers can find continuous round-the-year employment. For instance, an agricultural labour household in traditional agriculture, where employment tends to be seasonal, is likely to have a large number of workers. The shorter (longer) the length of the employment season, the number of workers is likely to be larger (smaller). Thus, the number of members of a household who have to seek work depends on some aspects of demand for labour, contrary to the assumption of independence of demand and supply.

From the earlier discussion, it follows that when the intensity of cropping in agriculture increases, the supply of local agricultural labourers is likely to decline, even if the wage rate remains unchanged. Now, a labourer can get more regular employment and she/he can earn more in a year. Consequently, some, if not all, child and female members of a labour household are likely to be withdrawn from the labour market. This hypothesis finds support in the case of Punjab and Haryana, which, according to some studies,[13] had surplus agricultural labour before the onset of the Green Revolution but have been experiencing shortage of local labour for

[13] See Mehra 1972; Mathur 1964.

quite some time now. Recently, studies and observers have found that the supply of labour has declined in the districts where the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA) has been relatively successful. However, some economists have considered this to be the result of a backward-bending labour supply curve. They have forgotten to notice that the range of wages prevailing in the rural areas is on the lowest side of possible wage rates.

An inescapable implication of the earlier analysis is that in such conditions an increase in wage rate will reduce labour supply, unlike what is described in textbooks, and it will reduce child labour. In fact, child labour exists because of adult unemployment and low wages.

The downward-sloping labour supply schedule of the poor can be very useful in understanding the economic position of a household or a group in a particular context. In a situation where the economy is traditional or backward and the size of the modern sector in the economy is insignificant, and consequently an upward slope of the labour supply schedule is not observed, the WPR of a group or a household is a very good inverse indicator of its economic status. The higher the WPR, the lower is the economic status. This simple but robust indirect indicator is very useful when direct evidence on economic status, namely income and wealth, is not available.[14]

4 Differential Wages of Male, Female and Child Labour

Our discussion can be easily related to the issue of differential wage rates of male, female and child labour of low-skilled or unskilled variety, in spheres of economic activity such as agriculture, construction work, domestic service and even factories. Suppose a property-less household has three members, a man, his wife and their child, more than 5 years of age, all unskilled. In general, there is a sequence in which the members of the household may join the labour market. The man is socially assigned the role of the 'provider'.[15] Accordingly, it is he who first looks for work in order to earn the subsistence of the family. In the event of his earnings falling short of the family's target income, either because of low wage rate or because he cannot get regular employment, his wife enters the labour market.[16] If their combined income falls short of their target income, their child too joins the labour market.

[14] I have used this to analyse and understand the link between ritual and economic status of castes at the macro-level (of provinces) in India during the early decades of the twentieth century. See Chaudhury 2005a, b. More results, at the country (India) level, were presented in my Vera Anstey Memorial Lecture. See Chaudhury 2009.

[15] This is an age-old practice which is manifested in the words used for husband and wife in ancient Sanskrit. 'Bhartr is generally used for "husband", but is also used in its root meaning and then connotes "one who feeds"... Bharya occurs in the later texts of this period and is the counterpart of bhartr, the husband. Bharta [nominative of bhartr] is the protector, the feeder; bharya is the protected or the fed female, therefore, the wife' (Karve 1953).

[16] A Marathi woman aged 30, with two children to look after, employed in a mill, told the Indian Factory Commission of 1890 that she worked because the income of her husband, a porter in the Railways, was not 'sufficient'. The working hours then were much longer than at present. Having joined the factory, she had no option of working for fewer hours. However, she expressed

This sequence, however, is extremely difficult to observe for a particular family at a point of time. Evidence, though of an indirect nature, can be found from observations, noted earlier, relating to cross sections of society. Again, in the sequence of joining the labour market, sometimes the child may precede the female adult. This may happen due to demand conditions, i.e. if employment for child labour is more easily available than for adult female labour, e.g. child labour in Sivakasi.

When a female goes in search of a job, it implies that the man of the family does not earn enough for its subsistence. Hence, participation of a female in the labour market reflects that her family is in distress. This fact can be easily appreciated by employers[17] to press down the wages of a female worker vis-à-vis a male worker even when there is no difference in their productivities, e.g. many types of agricultural operations. There are reasons which make it easy for the employers to impose gender-based wage differences. In a male-dominated society, both males and females accept that a female is inferior to a male in all respects including the ability to earn. Besides, the female worker's objective is to bridge the gap between the family's target income and her husband's income. Thus, her income is seen as supplementary or secondary.[18] This factor also contributes to the acceptability of low wages.

A child selling his/her labour in the market reflects the extreme distress of his/her family, i.e. his/her parents are unable to earn the subsistence for the family. This fact is used by the employers to severely press down the wage rate of child labour vis-à-vis adult labour. It is true that in many types of jobs children are no match for adults. At the same time, there are many kinds of work at which a child is no worse, in any case not to the degree of the wage differences, than an adult. It can be easily noticed that child labour is relatively rarely employed in the first kind and frequently employed in the second kind of jobs, in fact, so much so that in many instances child labour largely eliminates adult labour causing considerable unemployment among the latter. This gives a firm impression that the wages of child labour are kept abysmally low because of the extreme distress and the consequent low bargaining position of the families which supply child labour and not due to low productivity of child labour.

5 Conclusion

Distribution of income and employment are critically linked. Poverty creates a large supply of unskilled or low-skilled labour, as education is low or absent for the poor. (There can be skilled though uneducated artisans among the poor.) Mass poverty, low wages and unemployment generate low demand for food, clothing and shelter,

---

a preference for 'less hours and less pay'. See The Indian Factory Commission 1890. This is also a possible evidence for the existence of a fixed target income assumed in this chapter.

[17] In many instances, the employer may even have personal knowledge of the economic situation of such a family. However, it is not necessary. Given the gender-based division of labour in our society, the participation of a female in the market for unskilled or low-skilled labour is a clear indication of the economic distress faced by the household. However, the above argument need not apply to female workers belonging to higher income groups, who are usually well-educated professionals.

[18] In some cases, the reverse may be true; but these cases appear to be exceptional.

the goods of mass consumption originating in agriculture and industry. That, in turn, keeps the demand for labour (involved in the production of these goods) low. Hence, from both sides, that is demand for and supply of labour, there is a pressure for unemployment to be high and wages to be low. Thus, market forces are unlikely to reverse the existing conditions. Under these circumstances, it is the responsibility of the state to implement policies to increase wage rate and employment opportunities for adults, which are likely to be unwelcome to the rural as well as urban employers who are enamoured by cheap labour. However, successful implementation of such policies will reduce poverty, deprivations and indignities for the majority of the people. And, it will improve education, health and efficiency. As a result, the society will have more harmony.

References

Census of India (1911) Vol. 5, Bengal Bihar and Orissa, Pts. 2 and 3; Vol. 15, United Provinces, Pts. 1 and 2

Chaudhury P (1987) Labour supply schedule of the poor and some related issues. Presented at a National Seminar on Approaches to Issues in Economic Development, Department of Economics, Presidency College, Calcutta, 6–8 March 1987

Chaudhury P (1992) Labour Migration from the United Provinces, 1881–1911. Studies in history, new series, vol 8, no. 1, pp 13–41

Chaudhury P (2004) The "Creamy Layer": Political Economy of Reservations. Econ Polit Weekly 39(20):1989–1991 (May 15)

Chaudhury P (2005a) Does caste indicate deprivation? Seminar, May 2005, no. 549 , pp 26–29

Chaudhury P (2005b) Caste Quotas in India: some basic issues. Bhartiya Samajik Chintan (A Quarterly Journal of the Indian Academy of Social Sciences) April–June 2005, pp 25–34

Chaudhury P (2008) Unemployment in a Traditional Agrarian Economy: Theory and Evidence. in Bhaumik S K (ed.) Reforming Indian Agriculture Towards employment generation and poverty reduction, essays in honour of G K Chadha, Sage Publications, New Delhi, pp 345–362

Chaudhury P (2009) Caste, Class and Public Policy in India: A macro- level quantitative analysis. Vera Anstey Memorial Lecture in Economic History, delivered at The Indian Economic Association's Annual Conference in Bhubaneswar in December 2009. The lecture was printed and distributed by the I. E. Association at the conference

Karve I (1953) Kinship organisation in India. Deccan College, Poona, pp 30–1

Lewis WA (1954) Economic development with unlimited supplies of labour. Manch Sch 22(2): 139–191

Mathur A (1964; July) The anatomy of disguised unemployment. Oxford Economic Papers

Mehra S (1972) Surplus labour in Indian agriculture. In: Chaudhuri P (ed) Readings in Indian Agricultural Development. Allen & Unwin, London

Schultz TW (1964) The doctrine of agricultural labour of zero value. Transforming Traditional Agriculture, New Haven

Sen AK (1966) Peasants and dualism with or without surplus labour. J Polit Econ 74(5):425–450

The Indian Factory Commission (1890) Evidence of workers, first witness, p 22

Chapter 14
Switching as an Investment Strategy: Revisiting Parrondo's Paradox

Avik Chakrabarti

1 Introduction

An essential characteristic of most investment decisions is uncertainty over future rewards from an investment.[1] The best an investor can usually do is to assess the probabilities of the alternative outcomes that can mean greater or smaller profit (or loss) for the venture. Turbulence in the economic environment prompts investors to often seek the protection of strategies that tend to minimize the risk of loss caused by idiosyncratic downdrafts in some investment.

While changes in the prices of any particular asset cannot be predicted with certainty, conventional wisdom theorizes that a diversified asset is less risky. One of the pillars of financial economics and investment theory is that asset diversification reduces risk without affecting returns. An important aspect of asset diversification is correlation, whereby stocks are selected to offset each other's vulnerabilities. Careful matching of individually volatile assets can work to reduce combined volatility.

The question that motivates this chapter is: Can an investment strategy combine a set of losing assets in a way that they become gaining? The results presented in this chapter demonstrate that there may be situations when *randomly switching* between assets with negative expected return can itself prove to be a useful strategy for the investor by generating an asset with positive expected return. To that end, this chapter draws heavily on the literature spawning from Parrondo's paradox (where the combination of two losing games can result in a winning game) as it maps into any gamble through a *ratchet effect:* it is possible to use random fluctuations to generate ordered motion against a potential barrier in disequilibrium.

[1] See Katzner (1998).

A. Chakrabarti (✉)
Department of Economics, University of Wisconsin-Milwaukee,
Milwaukee, WI, USA
e-mail: chakra@uwm.edu

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_14,

2 Parrondo's Paradox

Parrondo's paradox has its roots in the works of Ajdari and Prost (1993) who had demonstrated that a Brownian particle, moving in the direction of the force when the potential is "on" or "off," can move in the opposite direction if the potential is "flashing." The dynamics of this particle have been applied to gambling games, a broad class of which has come to be known as Parrondo's games, where two losing games (that mimic the behavior of the Brownian particle in a flat and a ratchet potential) can generate, when alternated, a winning game. While it is possible to conceptualize a wide range of physical processes drifting in a counterintuitive direction, Parrondo's paradox has gained significant mileage as the first game-theoretic realization of such processes.[2]

Consider two games: one memory-less game (e.g., simple coin tossing) and another history-dependent game with the structure of a Markov chain (e.g., several coins, one of which is chosen for a toss conditional on the outcome of previous coin tosses). The expected return from the memory-less game depends only on the probability of landing heads, but the expected return from the history-dependent game depends not only on the probabilities with which the various coins land heads but also on how often each coin is tossed, which is determined by the steady-state probabilities of the Markov chain. When played separately, each game is set up to yield a negative expected return. However, randomly switching between these games changes the transition probabilities and, therefore, the steady-state probabilities for the Markov chain game. In sum, a player's fortune can be reversed when an unconditional probability enters as a conditional probability in the Markov chain game, changing the steady-state probabilities of the latter.

3 Analysis

This section presents a formal analysis based on properties of discrete-time Markov chains with stationary transition probabilities to establish the result that randomly switching between assets with negative expected return can generate an asset with positive expected return.

Definition An investment in any asset is gaining if $\Pr\{X_{i+n} > i|X_i = i\} > \Pr\{X_{i+n} < i|X_i = i\}$; neutral if $\Pr\{X_{i+n} > i|X_i = i\} = \Pr\{X_{i+n} < i|X_i = i\}$; and losing if $\Pr\{X_{i+n} > i|X_i = i\} < \Pr\{X_{i+n} < i|X_i = i\}$.

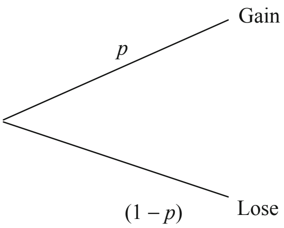
That is, an investment in any asset is *gaining*, *neutral*, or *losing* according to the probability that the value of the asset goes up in n successive rounds is greater than, equal to, or less than the probability that the value goes down in n successive rounds.

It may be noted that a Markov chain is said to have stationary transition probabilities if $P(x, s; t, A) = \Pr\{X_n \in A|X_m\} = x(\forall t > s)$ is a function only of $(t - s)$. For a Markov chain with stationary transition probabilities, $\Pr\{X_{n+1} = j|X_n = i\} = P_{ij}$,

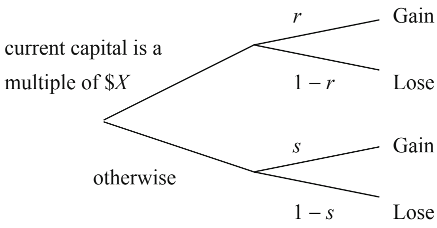
[2] See Parrondo et al. (2000) and Harmer and Abbott (2002).

i.e., the one-step transitional probabilities are independent of the number (n) of rounds (see Karlin and Taylor 1975, p. 30, 45–46).

Consider two assets (A and B). In every round, the value of the asset A grows with a probability p and shrinks with a probability $(1 - p)$. Let p_k be the conditional probability that the investor's capital in asset A reaches zero given that he/she starts with a capital worth \$$k$. Let the probability that the value of B goes up or down in a single round depend on the value of investor's current capital. The probability that the value of B goes up is r and down is $(1 - r)$ if the investor's current capital is a multiple of \$$X$. The probability that the value of B goes up is s and down is $(1 - s)$ if his/her current capital is not a multiple of \$$X$. Let q_k be the conditional probability that the investo's capital in asset B reaches zero given that he/she starts with a capital worth \$$k$.



Asset A



Asset B

It is a consequence of the Markov chain theory (Karlin and Taylor 1975, pp. 49) that if ϕ_k be the conditional probability that the investor's capital reaches zero given that he/she starts with a capital worth \$$k$, then either

a. $\phi_k = 1 \ \forall \ k \geq 0$, i.e., the investment is neutral or losing or
b. $\phi_k < 1 \ \forall \ k > 0$, i.e., the investment is gaining.

Proposition 1 An investment in A is gaining if $p >$ ½, losing if $p <$ ½, and neutral if $p =$ ½.

Proof The set of probabilities $\{p_k\}$ is the minimal nonnegative solution to the set of equations:

$$p_k = p p_{k+1} + (1 - p) p_{k-1} \ \forall \ k \geq 1 \tag{14.1a}$$

$$p_o = 1. \tag{14.1b}$$

The general solution to Eq. (14.1a) is of the form

$$p_k = \mathrm{A}_0 \ (1/p \ - 1)^k + \mathrm{A}_1 \tag{14.1c}$$

where A_0 and A_1 are constants.

Using the boundary condition (14.1b), expression (14.1c) reduces to

$$p_k = \mathrm{A}_0 \left[(1/p - 1)^k - 1 \right] + 1 \tag{14.1d}$$

If $(1/p-1)<1$, the minimal nonnegative solution to Eq. (14.1a) is obtained when $A_0=1$, in which case $p_k=(1/p-1)^k\ \forall\ k>0$.

Therefore, $p_k = min\left[(1/p-1)^k, 1\right]$ and the investment is gaining if

$$(1/p-1)<1,\ \text{i.e., if}\ p>½.$$

By symmetry, the investment is losing if $p<$ ½ and neutral if $p=$ ½.

Proposition 2 An investment in B is gaining if $(1-r)(1-s)^{X-1}/rs^{X-1}<1$, losing if $(1-r)(1-s)^{X-1}/rs^{X-1}>1$, and neutral if $(1-r)(1-s)^{X-1}/rs^{X-1}=1$.

Proof For $k\in(1,\ldots,X-1)$ and $m>1$, the set of probabilities $\{q_k\}$ satisfies the set of equations:

$$q_{mX} = rq_{mX+1} + (1-r)q_{mX-1} \tag{14.2a}$$

$$q_{mX+k} = sq_{mX+k+1} + (1-s)q_{mX+k-1} \tag{14.2b}$$

$$q_o = 1 \tag{14.2c}$$

For $k\in(1,\ldots,X-1)$, the general solution to Eq. (14.2b) is of the form

$$q_k = B_0(1/s-1)^k + B_1 \tag{14.2d}$$

where $B_0 = \left[q_{mX} - q_{(m+1)X}\right]/\left[1-(1/s-1)^X\right]$

and $B_1 = \left[q_{(m+1)X} - q_{mX}(1/s-1)^X\right]/\left[1-(1/s-1)^X\right]$

Using Eq. (14.2d), Eq. (14.2b) reduces to

$$\left[(1-r)(1-s)^{X-1}\right]q_{(m-1)X} - \left[rs^{X-1} + (1-r)(1-s)^{X-1}\right]q_{mX} + rs^{X-1}q_{(m+1)X} = 0\ \ \forall\ \ m\geq 1. \tag{14.2e}$$

The general solution to Eq. (14.2e) is of the form

$$q_{mX} = B_2\left((1-r)(1-s)^{X-1}/rs^{X-1}\right)^m + B_3\ \ \forall\ \ m\geq 0. \tag{14.2f}$$

Using the boundary condition (14.2c), expression (14.2f) reduces to

$$q_{mX} = B_2\left[\left((1-r)(1-s)^{X-1}/rs^{X-1}\right)^m - 1\right] + 1 \tag{14.2g}$$

Therefore, $q_{mX} = min\left[\left((1-r)(1-s)^{X-1}/rs^{X-1}\right)^m, 1\right]$ and the investment is gaining if

$$(1-r)(1-s)^{X-1}/rs^{X-1}<1.$$

By symmetry, the investment is losing if $(1-r)(1-s)^{X-1}/rs^{X-1}>1$ and neutral if $(1-r)(1-s)^{X-1}/rs^{X-1}=1$.

Proposition 3 Even when investments in assets A and B are individually losing ∃, a randomized asset C is gaining.

Proof Suppose, in each round, the investor invests in asset A with probability δ;and in asset B with probability $(1-\delta)$. Then the probability that the value of this randomized asset (C) will go up is given by $(\delta p + (1-\delta)r)$ if capital is a multiple of \$*X* and $(\delta p + (1-\delta)s)$ if capital is *not* a multiple of \$*X*.

Let g_k be the conditional probability that the investor's capital reaches zero given that he/she starts with a capital worth \$*k*.

For $k \in (1, \ldots, X-1)$ and $m > 1$, the set of probabilities $\{g_k\}$ satisfies the set of equations:

$$g_{mX} = (\delta p + (1-\delta)r)g_{mX+1} + (1-(\delta p + (1-\delta)r))g_{mX-1} \tag{14.3a}$$

$$g_{mX+k} = (\delta p + (1-\delta)s)g_{mX+k+1} + (1-(\delta p + (1-\delta)s))g_{mX+k-1} \tag{14.3b}$$

$$g_o = 1. \tag{14.3c}$$

For $k \in (1, \ldots, X-1)$, the general solution to Eq. (14.3b) takes the form

$$g_k = \mathrm{C}_0(1/(\delta p + (1-\delta)s) - 1)^k + \mathrm{C}_1 \tag{14.3d}$$

where $\mathrm{C}_0 = \left[g_{mX} - g_{(m+1)X}\right] / \left[1 - (1/(\delta p + (1-\delta)s) - 1)^X\right]$

and $\mathrm{C}_1 = \left[g_{(m+1)X} - g_{mX}(1/(\delta p + (1-\delta)s) - 1)^X\right]/1 - (1/(\delta p + (1-\delta)s) - 1)^X.$

Using Eq. (14.3d), Eq. (14.3b) reduces to

$$\begin{aligned}&\left[(1-(\delta p + (1-\delta)r))(1-(\delta p + (1-\delta)s)^{X-1}\right] g_{(m-1)X}\\ &\quad - \left[(\delta p + (1-\delta)r)(\delta p + (1-\delta)s)^{X-1}\right.\\ &\quad \left. + (1-(\delta p + (1-\delta)r))(1-(\delta p + (1-\delta)s))^{X-1}\right] g_{mX}\\ &\quad + (\delta p + (1-\delta)r)(\delta p + (1-\delta)s)^{X-1} g_{(m+1)X} = 0 \;\; \forall \;\; m \geq 1.\end{aligned} \tag{14.3e}$$

The general solution to Eq. (14.2e) takes the form

$$\begin{aligned}g_{mX} = \mathrm{C}_2\Big((1-(\delta p + (1-\delta)r))(1-(\delta p + (1-\delta)s))^{X-1}/(\delta p + (1-\delta)r)\\ (\delta p + (1-\delta)s)^{X-1}\Big)^m + \mathrm{C}_3 \;\; \forall \;\; m \geq 1.\end{aligned} \tag{14.3f}$$

Using the boundary condition (14.2c), expression (14.2f) reduces to

$$\begin{aligned}g_{mX} = \mathrm{C}_2\Big[\Big((1-(\delta p + (1-\delta)r))(1-(\delta p + (1-\delta)s))^{X-1}/(\delta p + (1-\delta)r)\\ (\delta p + (1-\delta)s)^{X-1}\Big)^m - 1\Big] + 1\end{aligned} \tag{14.3g}$$

Thus,

$$\begin{aligned}g_{mX} = min\Big[\Big((1-(\delta p + (1-\delta)r))(1-(\delta p + (1-\delta)s))^{X-1}/(\delta p + (1-\delta)r)\\ (\delta p + (1-\delta)s)^{X-1}\Big)^m, 1\Big].\end{aligned} \tag{14.3h}$$

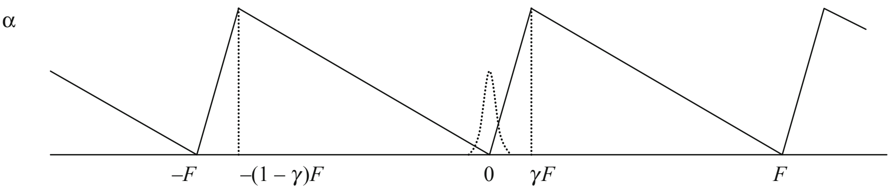


Fig. 14.1 Flashing Ratchet

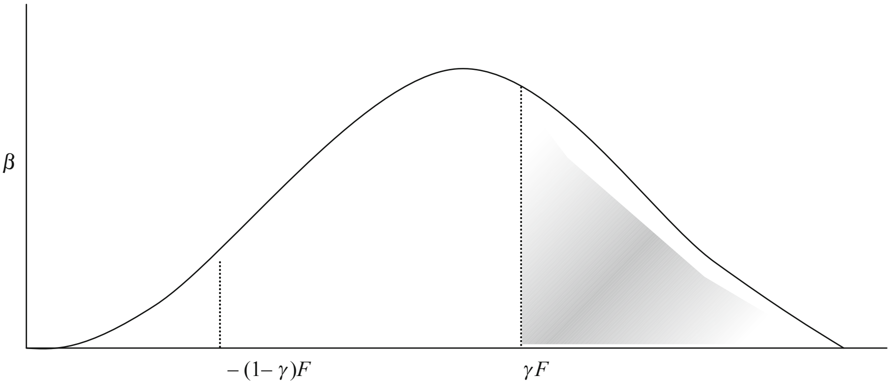


Fig. 14.2 Flashing Ratchet

Therefore, investment in asset C is gaining if

$$(1-\gamma_1)(1-\gamma_2)^{X-1}/\gamma_1{\gamma_2}^{X-1} < 1,$$

where $\gamma_1 = (\delta p + (1-\delta)r)$ and $\gamma_2 = (\delta p + (1-\delta)s)$.

Combining this with propositions (1) and (2), proposition (3) holds as long as the following conditions are met:

$$p < ½ \tag{14.4a}$$

$$(1-r)(1-s)^{X-1}/rs^{X-1} > 1 \tag{14.4b}$$

$$(1-\gamma_1)(1-\gamma_2)^{X-1}/\gamma_1\gamma_2^{X-1} < 1 \tag{14.4c}$$

where $\gamma_1 = (\delta p + (1-\delta)r)$ and $\gamma_2 = (\delta p + (1-\delta)s)$.[3]

An analog of this result may be found, drawing on Ajdari and Prost (1993), in Brownian ratchets (Feynman et al. 1963). Consider a system (see Figs. 14.1–14.6) with two one-dimensional potentials, α (on) and β (off). The asymmetry of the potential is determined by $\gamma \in [0, 1] : \gamma = ½$ generates a triangular symmetric potential

[3] For example, consider the following set of values for the parameters $\{p, r, s, X\} \equiv \{0.454, 0.091, 0.008, 0.909, 0.5, 3\}$. Then, $p = 0.454 < ½, (1-r)(1-s)^{X-1}/rs^{X-1} = 1.2 > 1$, and $(1-\gamma_1)(1-\gamma_2)^{X-1}/\gamma_1\gamma_2^{X-1} = 0.723 < 1$.

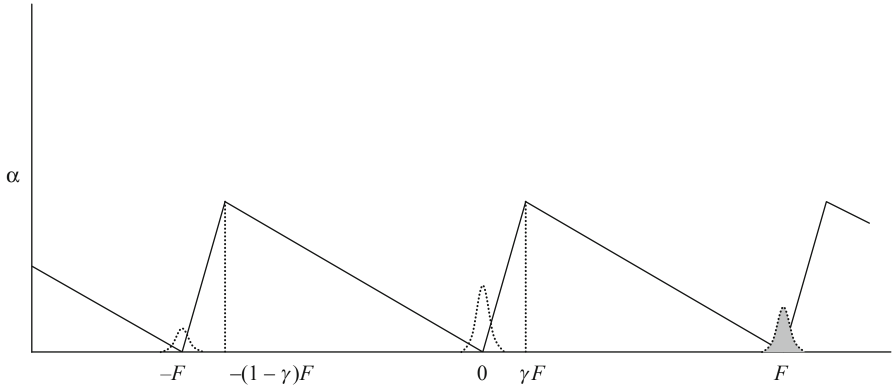


Fig. 14.3 Flashing Ratchet

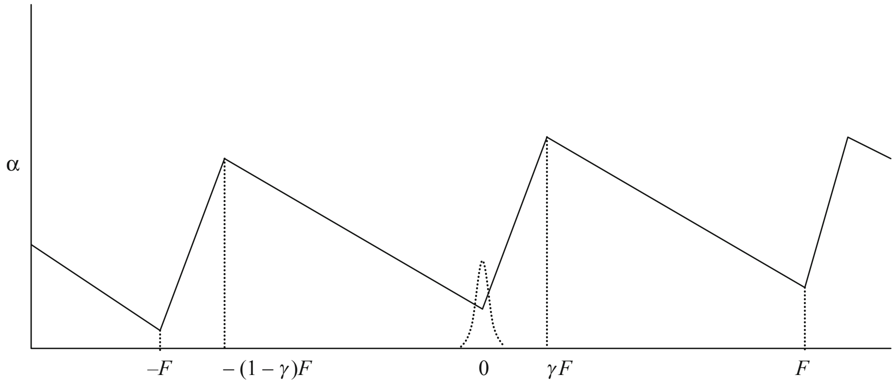


Fig. 14.4 Flashing Ratchet

which otherwise is tilted. Let there be Brownian particles existing in the potential. Time modulating the potential α and β can induce motion (flashing ratchets). When α is applied, the particles are trapped in the minima of the potential so that the concentration of the particles is peaked. Switching the potential off to β allows the particles to diffuse freely so that the concentration is a set of normal curves centered around the minima. When α is switched on again there is a probability (p_+) that some particles are to the right of γF: p_+ is proportional to the shaded area in Fig. 14.2. These particles move forward to the minima located at F. Similarly, there is a probability (p_-) that some particles are to the left of $-(1-\gamma)F$: p_- is proportional to the shaded area in Fig. 14.5. These particles move backward to the minima located at $-F$. When $\gamma <$ ½, as in the figures, $p_+ > p_-$, i.e., the net motion of the particles is to the right. If a tilted periodic potential is held in either the α state or the β state, the particles move downhill. When the tilted periodic potential is switched between the α state or the β state it can be shown, by solving the Fokker–Planc equation of the system, that Brownian particles move uphill (Hanggi and Bartussek 1996). Thus, by switching

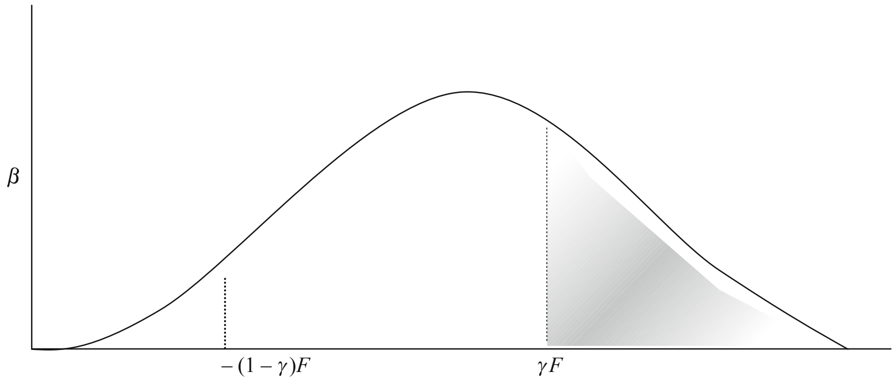


Fig. 14.5 Flashing Ratchet

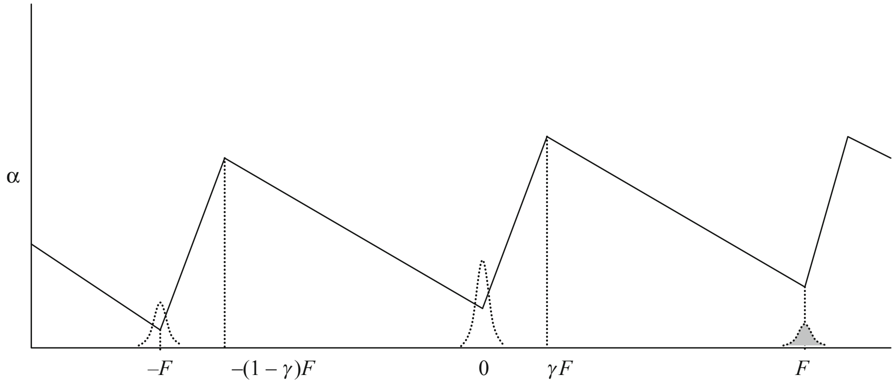


Fig. 14.6 Flashing Ratchet

between two states a flashing Brownian ratchet can move particles uphill (up in potential) even if particles ordinarily move down in each state. The Brownian ratchet is continuous in time and space in that particles can exist at any real displacement along the potential which can be flashed on or off at any real time.

Assets A and B essentially emulate the two potentials in the Brownian ratchet but for the fact that the ratchet now is discrete in both the analogous time and space: The individual states are like the losing assets but when they are switched between gaining expectations are generated. In equilibrium, the expected gain from switching between the losing assets will indeed be equal to the transaction cost of switching. Intuitively, a flashing ratchet can be visualized as an uphill slope that switches back and forth between a linear and sawtooth-shaped profile. Brownian particles on a flat or sawtooth slope are expected to drift downwards. However, if one switches between the flat and sawtooth slope, the particles are "massaged" uphill. The key lies in the asymmetrical shape of the sawtooth that favors particles spilling over a higher tooth. The flat slope resembles one of the assets and the sawtooth slope is analogous to the other.

4 Conclusion

The recent global financial crises, triggered by the bursting of the real estate bubble and amplified by the concentration of risk in a highly leveraged financial sector, remind us that real-world investors bear little resemblance to those doodled by economists on academic manuscripts. Can an investor win even on the face of catastrophic failures of a market? Drawing on insightful contributions, mostly of researchers not confined to the boundaries of economics, this chapter highlights an important result that randomly switching between assets with negative expected returns can generate an asset with positive expected return. With sufficient room for applying such analyses to any gamble, some natural extensions of this chapter may include more than two assets and/or expand the parameter space.

References

Adjari A, Prost J (1993) Drift Induced by a periodic potential of low symmetry: pulsed dielectrophoresis. Comptes Rendus de l'Académie des Sciences, Série II 315(13):1635–1639

Feynman RP, Leighton RB, Sands M (1963) The Feynman lectures on physics, vol 1. Addison-Wesley

Hanggi P, Bartussek R (1996) Nonlinear physics of complex systems: current status and future trends: lecture notes in physics, vol 476. Springer

Harmer GP, Abbott D (2002) A review of Parrondo's paradox. Fluct Noise Lett 2(2):R71–R107

Karlin S, Taylor HM (1975) A first course in stochastic processes. Academic Press, New York

Katzner DW (1998) Time, ignorance, and uncertainty in economic models. University of Michigan Press

Parrondo JMR, Harmer GP, Abbott D (2000) New paradoxical games based on Brownian ratchets. Phys Rev Lett 85:5226–5229

Chapter 15
Asymmetric Information, Non-cooperative Games and Impatient Agents: Modelling the Failure of Environmental Awareness Campaigns

Somdeep Chatterjee and Asim K. Karmakar

1 Introduction

The primary concern of modelling behaviour of agents under a nonmarket economic framework has been gaining importance with the advent of the study of the environment under the purview of efforts to commoditise environmental amenities. In this regard, a wide variety of literature supports several propositions, apprehensions, results and corollaries. However, certain primary questions have often been omitted and neglected with the notional view of their obviousness. Therefore, this is an attempt to intervene in the literature by making an inroad into the most common and obvious happenings in the environmental goods' market, and try and analyse them from an economic point of view and end up, at times, with not so obvious results and at other times providing foundation to the so-called obviousness.

The starting point of this chapter is to develop a textbook-style framework of environmental commodity usage and present in a lucid way the basic ideas of what is going on without much of formal environmental economics language. We try to stay within the scope of standard classroom microeconomic and macroeconomic theories, so that we can present an idea which even the beginners in the subject would be able to relate to.

However, it is unfair and inappropriate to skip a quick survey of the literature in this field although keeping in mind the simplicity of our framework, it is hard, if not impossible, to enumerate a complete list of relevant work. Nevertheless, we mention some of the ones which attracted us the most for this essay.

There have been ideas, like the incorporation of utility mapping on the environmental conditions and the effects of production pollution and investment for

S. Chatterjee (✉)
Department of Economics, University of Houston,
204 McElhinney Hall, Houston, TX 77204-5019, USA
e-mail: schatterjee2@uh.edu

A. K. Karmakar
Department of Economics, Jadavpur University, Kolkata, West Bengal 700032, India
e-mail: asimkkarmakar@yahoo.co.in

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_15, © Springer India 2014

rehabilitation, to suggest that some degree of impatience on part of consumers and activists is required to reach an interior steady state (Levy 2008). In the model of lifetime utility maximization, which we present as Model 1, we shall delve into the technicalities of an optimal choice of an infinitely lived forward-looking consumer towards expending costs for environmental improvements, and establish that it is lack of 'information' in the traditional economic sense and not lack of 'awareness' that counts in a big way. If we look at some empirical work, there is inherent contradiction in many papers. Attempts have been made at identifying certain parameters for 'environmental concerns' (Dunlap and Mertig 1997), but the results have been contradicted later on (Diekmann and Franzen 1999). Both these studies tried to correlate gross national product (GNP) with the specific parameters, but ended up contradicting each other for many of these cases. The initial formulation of a negative correlation of GNP with Environmental Standard is, however, credited to an earlier work (Dunlap et al. 1993).

We shall establish two stylized models of analysis which shall be purely theoretical. Methodology of model framing is simple introspection and some use of mathematical tools, consistent with pure economic theory.

Model 1 deals with a simple explanation of why even if there is a perfect market setting and property rights defined over an environmental amenity, with the existence of environmentally conscious users, a resource can get depleted. The primary reason that one can possibly get to explain such a phenomenon is that agents are impatient. Also, the role of people's expectations about the renewability of a resource would play a major role in this model. We use a natural 'decay factor', 'δ', and a factor of maintenance of the resource which we call the 'upkeep factor' and is denoted as 'θ', and use standard concave utility functions to maximize lifetime utility subject to resource constraints and capital accumulation equations.

Model 2 is purely a strategic interaction model. We use the simple concepts of game theory to describe a situation why people fail to cooperate among themselves to provide for an improvement in environmental amenity for a localised project, and then suggest a modified game, which would induce everyone to go for the strategy to incur maintenance cost. We once again bring in the role of expectations about the payoffs of an individual and show that under a stylized setting, it is possible to avoid free-riding in a perfectly simultaneous setting, and the Nash equilibrium shall not entail free-riding and rather would lead to the desired solution.

Therefore, Sect. 2 deals with Model 1, Sect. 3 contains Model 2 and Sect. 4 points out Policy Implications and Sect. 5 presents Concluding Remarks.

2 Model 1: Decay Factor and Upkeep Factor

In this section, we consider a forward-looking consumer model (following Ramsey 1928) to demonstrate how in a simple textbook framework we can show the obviousness of depletion of natural resources. Consider agents are infinitely lived and there exists a single representative agent. They derive utility from consumption of a consumer good 'C' and a natural resource 'R'. We assume for simplicity that the environmental commodity is marketed and can be treated as an asset. (This is bound to be a controversial assumption in the Environmental Economics literature but it is innocuous for what we are trying to show.) Let us assume the individual's utility is defined as:

$$U(C_t,\ R_t) = \ln\,(C_tR_t)$$

Also, assume individuals having a time preference parameter, then $0 \le \beta \le 1$. So, we can denote the lifetime utility of the agent as

$$V(.) = \sum_t \beta^t U(C_tR_t) = \sum_t \beta^t \ln\,(C_tR_t),\ t = 0, 1, 2, 3, \ldots (\text{to } \infty)$$

Consider that the production function in the economy is $Y_t = f(K_t)$. Then we can write down a dynamic accumulation equation for capital as

$K_{t+1} = (1 - \varphi)K_t + f(K_t) - C_t$, where $\varphi =$ Depreciation of capital stock.

Further, we can assume that the natural resource has a decay factor, δ, and can be replenished by maintenance captured by an upkeep factor, θ. It must be the case that $\theta \le \delta$ (or in other words the resource cannot be produced, it can only be replenished to the extent of decay). So, we can think of a dynamic resource constraint as follows:

$$R_{t+1} = (1 - \delta)R_t + \theta R_t$$

So, we now have the setup ready for our optimization exercise. All we are doing is analyzing a modified Ramsey model accounting for a new variable (in line with our claim of presenting a simple textbook model). The individual maximizes lifetime utility over C and R subject to the constraints. The Lagrangian is as follows:

$$L = \sum_t [\beta^t U'(C_t, R_t) + \mu_t\{(1 - \varphi)K_t + f(K_t) - C_{t+}(1 - \delta)R_t + \theta R_t - K_{t+1} - R_{t+1}\}]$$

The first order conditions are:

$$C : \beta^t U'(C_t, .) - \mu_t = 0; \tag{15.1}$$

$$K : \mu_{t+1}\,[(1 - \varphi) + f'(K_{t+1})] - \mu_t = 0; \tag{15.2}$$

$$R : \mu_{t+1}\,[1 - \delta + \theta] - \mu_t = 0; \tag{15.3}$$

Let us analyze the above conditions, in brief, before proceeding further to develop the intuition of how the environment can get degraded even if there are efforts to upkeep the environment in the presence of environmentally conscious agents.

The conditions given in (15.1) are nothing but the standard Euler equation of consumption theory. We can combine the condition for adjacent periods to get:

$$C_{t+1}/C_t = \beta(\mu_t/\mu_{t+1}) \tag{15.4}$$

This follows from the fact that individuals have log utility.

Conditions (15.2) and (15.3) give us:

$$\mu_t/\mu_{t+1} = \left[(1 - \varphi) + f'(K_{t+1})\right] = [1 - \delta + \theta] \tag{15.5}$$

The second equality in (15.5) tells us that:

$$f'(K_{t+1}) - \varphi = \theta - \delta \tag{15.6}$$

We can refer to (15.6) as the 'environmental upkeep arbitrage' condition. Notably, the left side of the equality is nothing but returns to the capital. So, what this arbitrage condition highlights is that the agent can choose to save or invest in environmental upkeep interchangeably until the returns from both makes him indifferent between either.

Conditions (15.1) and (15.2) can be combined to note the following:

$$C_{t+1}/C_t = \beta\left[(1 - \varphi) + f'(K_{t+1})\right]. \tag{15.7}$$

2.1 Steady State in the Economy and the Golden Rule

If we think of the economy to be at steady state, then we want consumption to be constant over time. So, $C_{t+1} = C_t$. Thus, (15.7) now implies

$$(1/\beta) - 1 + \varphi = f'(K_{t+1}). \tag{15.8}$$

We know that agents are impatient, and so, $\beta \leq 1$. This would mean, $f'(.) \geq \varphi$. This is the reason that the steady-state level of Capital stock in a standard Ramsey economy has to be below the Golden Rule level. However, in our model this becomes counter intuitive, since we allow for the accumulation of capital as well as environmental resource but put a constraint on the amount of the total resource ($\theta \leq \delta$). If, we were at a Golden Rule level then $f'(.) = \varphi$, and as a result $\beta = 1$. Then, from the arbitrage condition (15.6) we know $\theta = \delta$, or the upgrade should exactly be equal to the decay to allow constant consumption. What happens in reality is something different. We find that in spite of the efforts from agents, the environment actually degrades. This can be explained in the simple model, as mentioned earlier, due to the presence of impatient individuals who discount the future. If $\beta < 1$, then the model breaks down and we cannot attain the Golden rule and we cannot even attain a steady state in this economy.

The only way to have the earlier mentioned model to be consistent and not break down is if $\theta = \delta$ exactly.

2.2 Why is $\theta < \delta$ Often Observed?

Even if a mechanism of utility maximization and a simple cost–benefit analysis existed and the utility cost of upkeep could be linearly weighed against the benefit from non-decay, we would still have depletion. This can be easily explained by the parametric restriction in our model of the nature; $\delta > \theta$. Why can this logically hold?

Suppose there is a lack of information about the magnitude of δ, then the resource may end up getting completely depleted. This is because optimal $\theta^* = \delta$ suggests that the society should choose such a θ that is exactly equal in magnitude to the δ. However, the δ is in most cases unknown, so it is, therefore, not an irregularity that $\theta < \delta$. So, a lack of information about the actual magnitude of degradation caused due to resource usage itself leads to problems and the lack of market and institutions are not even required.

2.3 Two Suggestive Reasons for the Degradation

From the earlier mentioned analysis, we can think of two possibilities which would make life simple to think of why the natural resources can be used up in a trivial setting without going into the rigorous details provided by the vast Environmental Economics literature. It is beyond the scope of this essay to think of the technicalities of nonmarket valuation of the environmental amenities and formally setting-up an environmental problem. Our aim was to make a textbook presentation of how with standard tools taught in the classroom, we can make the evidence pretty apparent.

Firstly, we resorted to impatient agents who discount the future and show that it is enough for not being able to establish a steady state or growth in the economy with environmental resource use being equally important as consumption.

Secondly, in a very brief point we observe that, information asymmetry can play its part too. This second idea can be delved into with greater detail in a strategic setup. We deal with this in the following sections.

3 Model 2: A Game Theoretic Analogy

Any standard analysis in game theory always begins with the classical enunciation of the prisoners' dilemma. For environmental economics, in general, and our chapter, in particular, it shall be no different. We reiterate the fact that providing for the maintenance of an environmental amenity in a community with strategic interaction shall replicate a prisoners' dilemma-type scenario. It is not required to go over the prisoners' dilemma game here, and it is a redundant exercise to try and explain equilibrium strategy choices for a Simultaneous Move game, viz., the Nash equilibrium. Therefore, our starting point in this section is to try and build on the Model 1 mentioned earlier. Let us for the time being assume that there are two (instead of one) representative individuals in the economy, such that their joint action (contributing some θ) determines whether or not an environmental upkeep is possible or carried out. This is done to extend the representative microeconomic decision-making of an individual to an interaction level to show how 'self-maximization' is supposed to confront the overall community interests and then revise the model, with the use of other standard economic technicalities to show how the problem can be overcome.

So, our first step is to establish (in accordance with standard prisoners' dilemma[1] beliefs) that under an interactive setup, self maximization will lead to an outcome that all would have wanted to avoid (it shall lead to non-maintenance, which is not tenable with the presence of 'environmentally conscious' individuals). This is the type of result that is in the literature often referred to as the, 'Classical Constitutional Conundrum' (Bowles 2006).

The Game we design, therefore, has two individuals, both aware and willing to conserve the environment. However, they are also aware of the fact, that the Joint Action of the society only leads to an improvement (which is largely the case) and the source of the problem shall lie exactly here. Each agent is uncertain about the interests of the others (we bring in some pseudo-asymmetry of information here). Each player has two strategies to choose from, either to 'contribute some θ' or 'not contribute any θ'. We denote the two strategies by 'C' and 'NC', respectively. The main purpose of this exercise is to demonstrate, why an environmental upgradation and maintenance is so uncommon, even after awareness about costs and benefits. Therefore, let us notionally define certain values to try and formulate the payoff structure of this game.

Let the benefit from any individual environmental upgrade be equal to a payoff of 30 units. Also, the cost of an environmental upkeep is say 50 units. We deliberately formulate a cost greater than the benefit if only one individual acts benevolently. This should mean in essence, that a single individual should definitely not incur the cost all by himself of an environmental upgrade, thereby endogenizing the idea of not providing for the amenity improvement all by himself unless he is certain of the other ones' motives. To demonstrate that both would have been better off had they contributed; let us assume that dual action leads to a benefit of $30 + 30 = 60$ which exceeds the cost of upkeep for each individual.[2]

So, the normal form of this hypothetical game is as follows:

Player 2

		Contribute a θ (C)	Do Not Contribute a θ (NC)
Player 1	Contribute a θ (C)	10, 10	-20, 30
	Do Not Contribute a θ (NC)	30, -20	0, 0

[1] This is also a replication of the standard 'Public Goods Game', where the problem of free-riding leads to an outcome that by voluntary contribution, a public good is not always provided. One can consider the environmental upkeep as a 'public good' here, whose benefits are rival and non-excludable.

[2] We assume that the benefits accrue in total to both. So, a benefit of 30 units accrues to both indicating a Total Social Benefit of 60 units gross. Similarly, a benefit of 60 units also accrues to both, thus yielding a Total Gross Social benefit of 120 units. So, for a social planner, even if one individual contributes, it is sufficient to yield a Net Total benefit for the society. This somehow also entails the clash of interests of an individual against that of the society as a whole.

This essentially means the following:

Player 2

Player 1		Contribute a θ	Do Not Contribute a θ
	Contribute a θ	(60-50), (60-50)	(30-50), (30-0)
	Do Not Contribute a θ	(30-0), (30-50)	0, 0

Clearly, the Nash equilibrium strategy combination of such a game is (NC, NC). By design of the game, it is now a pattern of this theoretical model that the behaviour of Not Contributing (NC) any positive amount towards environmental protection is actually optimal. This is precisely because of the incentive design. Even though the spillovers to the society are better and the individual gains are higher if both the agents contribute, they end up in the quagmire of not performing such a task. So, once again we formalize a situation, where the presence of environmentally aware and wilful agents is also not enough to drive towards a mass improvement of the environment. This model is typical because, it extends the discussion of Model 1 for a single agent to a strategic interaction setup and provides consistency to the real life situation.

4 Policy Study: How to Escape the Doldrums

The situation described earlier in the two models does not paint a rosy picture at all. It is definitely required, as a result, to try and define incentives and redesign the game in such a manner, that under interactive process, the desired outcome is available. Once again, standard microeconomics comes to help us a lot in this way. As described initially, the basic idea of this chapter is to build on an economic theory to provide some support to the existing ideas in environmental economics. Standard policy economists, so to speak, shall at once try and devise tax and subsidy mechanisms, a well-celebrated yet often-neglected tool of solving the prisoners' dilemma. An intervention to make the suitable outcome more attractive is certainly the way to go about it. Also, in our case, it is easy to do such an analysis. For example, consider that there is a social planner or a government that wants to improve the environmental quality and declares that anyone who undertakes the cost, i.e. the θ in our case, or to be more precise the 50 units to incur in order to achieve environmental benefits, shall be compensated equivalently.[3] Immediately, one can notice that the payoff matrix then takes the following form:

[3] It is as if you spend 50 units on a good and produce the bill to the government, and the government pays you back 50 units. Essentially this means making the effort to improve environment, costless, in some hypothetical way.

	Contribute a θ (C)	Do Not Contribute a θ (NC)
Contribute a θ (C)	60, 60	30, 30
Do Not Contribute a θ (NC)	30, 30	0, 0

Thus, the game now yields a Nash equilibrium of (C, C). This is definitely a most desired solution, both individually and socially. However, the problem with this school of thought is that it essentially calls for a government intervention. It pressurizes the government to act in a 'super-benevolent' manner and puts no emphasis to the much called for awareness campaign. The issue is that if the government is after all called in for an intrusion, then how does the claim that people are environmentally conscious come through. Basically, we do not seem to like this kind of a solution which in all probability most policy economists would prescribe and feel it is an easy escape. Introspecting further into the idea, we offer an alternate, even though complicated, design by which we can attain a similar Nash equilibrium, but without any government financial burden.

Our method, thus, focuses on the 'nature' of the awareness campaign. We already have accepted the fact that the agents in our analysis are responsive to such campaigns and abide by awareness programmes, under the purview of rationality, per se. Let us therefore, devise an awareness programme which advertises that in case no individual contributes to an upkeep of the environment, there is a huge loss to all individuals in absolute terms apart from the social point of view. Therefore, instead of a zero payoff when both choose NC, our modified game shall incorporate a negative payoff.

We make the magnitude of an environmental effect, by similar action, a constant.[4] Another modification is that there is no benefit from single action accruing to anyone. So, if only one person contributes, there is no change in the environmental quality whatsoever (unlike a 30 unit accrual in the earlier case). Thus, the awareness campaign is aimed towards joint action and its importance. We now, once again, impose the assumption that individuals abide by this awareness campaign. This modifies our game as follows:

		Player 2	
		Contribute a θ (C)	Do Not Contribute a θ (NC)
Player 1	Contribute a θ (C)	10, 10	-50, 0
	Do Not Contribute a θ (NC)	0, -50	-60, -60

This new game now yields a Nash equilibrium of (C, C). Therefore, even without any government intervention and without any financial burden of a subsidy on the government, a slight modification in the awareness campaigns adequate to modify the

[4] This means the absolute value of a gain from environmental protection is equal to the absolute value of a loss from its lack thereof. Therefore, the payoff when both do not contribute is −60 to each, just the opposite direction of a positive 60 payoff to both when they together contribute. This is also consistent with heterogeneous action. If player 1 contributes and 2 does not the result replicates the case when player 2 does and 1 does not.

expectations of an individual agent is enough to establish the desirable equilibrium. In any case, we have worked throughout under the assumption of environmentally conscious agents, so there is less reason to be cynical that such a programme will not be abided by.

5 Conclusions

We have tried to include a few issues regarding environmental commodity usage. Basically, we have focussed on modelling the behaviour of agents who use up environmental resource for their day-to-day needs and have throughout worked on the assumption that in general, all agents under our focus are environmentally conscious and care about its upkeep and maintenance in a great way.

The first model has primarily been aimed at unleashing the idea that it is the lack of information or the asymmetry in its discharge therein that leads to a depletion. This is consistent with the claim that agents are also impatient and value the present more than the future. The major idea that comes out of Model 1 is that if the exact value of the decay is known to the agents and they are perfectly patient, then by their simple utility maximization exercise, they shall end up with a maintenance effort which is optimal to upgrade the resource. However, in reality this is rarely the case because in almost every case, the actual value of decay is not quantifiable or unknown.

The second model in essence establishes a strategic interaction between two such environment friendly agents. The idea it establishes is that in general, under a voluntary contribution to a public good game-type scenario, an inefficient equilibrium is attained. However, once the government can intervene with a subsidy waiving off costs of maintenance, all agents choose strategy combinations that are sustainable and efficient. Our idea was to explore an avenue wherein the government would not be called upon to intervene and we suggested that the 'awareness campaign' that is being run is redesigned. This new design creates 'beliefs' that only joint action is beneficial and puts a huge notional penalty towards joint 'inaction', and, hence, naturally, an efficient Nash equilibrium is attained where the strategic interaction leads to environmental upkeep.

References

Bowles S (2006) Microeconomics—behaviour, institutions and evolution. Princeton University Press

Diekmann A, Franzen A (1999) The wealth of nations and environmental concern. Environ Behav 31:540–49

Dunlap RE, Gallup GH, Gallup AM (1993) Of global concern: results of the health of the planet survey. Environ 35(7–15):33–39

Dunlap RE, Mertig AG (1997) Global environmental concern: an Anomaly for Postmaterialism. Soc Sci Quart 78:24–29

Levy A (2008) Environmental concern and rational production, consumption and rehabilitation. Economics Working Papers, School of Economics, University of Wollongong, NSW (wp 08–10)

Ramsey F (1928) A mathematical theory of saving. Econ J 38:543–559

Chapter 16
Government's Role in Controlling Food Inflation

Hiranya Lahiri and Ambar Nath Ghosh

1 Introduction

Food inflation has become a major cause of concern for not only the common man but also the policymakers. Of late, high inflationary pressure, particularly double-digit food inflation since October 2008, is turning out to be a spoilsport in an otherwise robustly growing Indian economy. Food prices in India started spiraling up mid-2008 onward. The spillover effects were visible in other sectors also, and 2010 witnessed overall inflation rate crossing 10 % for 5 months in a row. Inflation based on year-on-year wholesale price index (WPI) of primary food articles, on which the people spend the most, still rules high at double digits (in November 2011). Several factors like drought- or rain-induced shortages in food supply, rising international prices, fragmented value chains resulting in a large price spread of high-value commodities, greater government spending leading to increased money supply, structural changes in demand patterns, etc. are being cited as the main reasons behind this high food inflation.

The problem of food inflation is not new in India. The country has already witnessed many episodes of food inflation. It can be shown that apart from random supply shock, which is held responsible for the recent surge in food inflation, there is also a clear supply–demand gap. Table 16.1 (adapted from Chand 2011) reports the growth rate in various crops from 2004–2005 to 2009–2010. The declining production trend in majority of the items is clear from the table. On the other hand, due to increase in income as well as population pressure, the growth rate in demand has been increasing. Because of this widening gap, many economists have provided some long-term solutions. These mainly involve increasing the productivity of crops and

H. Lahiri (✉)
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: hiranyaeco@gmail.com

A. N. Ghosh
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: ang.juecon@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_16,

Table 16.1 Growth rate in output of major food commodities (in percents)

Item	1993–1994 to 2003–2004	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009	2009–2010
Food grains	0.69	– 6.96	5.16	4.16	6.21	1.34	– 8.00
Oilseeds	– 0.43	– 3.33	14.91	– 13.19	22.52	– 5.38	– 5.00
Sugarcane	– 0.15	1.38	18.60	26.44	– 2.06	– 22.10	– 11.80
Fruits	2.48	7.93	4.26	6.43	6.69	NA	2.50
Vegetables	3.03	8.02	21.62	4.16	5.37	NA	4.80
Total food[a]	2.39	0.55	5.87	4.10	5.39	1.60	– 0.20

[a] Refers to value of all food crops and livestock products at 1999–2000 prices

augmenting the supply of food grains (Chand et al. 2011). Chand (2011) advocates a proper export–import policy that will enable a sustainable availability of food grains every year depending upon the domestic production. Virmani and Rajeev (2001) advocate a lowering of minimum support price (MSP) as a long-term solution for bringing down the overall price level. In their opinion, the word minimum should cover only the variable cost of production. These authors fail to capture the fact that this channel will work only for those items for which there is public procurement. In recent times, the major drivers of food inflation have been onion, sugar, oilseeds, and vegetables for which there is no public procurement. Further, a lowering of MSP may not alter the market price if the amounts the middlemen (big retailers) pay to the farmers enter as a fixed cost in their profit function. In fact, in many situations, the price that the farmers receive from the middlemen is subsistence price, and the increase in market price is often not passed to the direct producer of crops. This chapter adds this aspect in its analysis and shows that for the speculators, a change in MSP may have no effect on the overall price level.

The major problem with the aforementioned studies is that these are only long-term solutions. Since food inflation affects a huge section of the Indian population who live at the brink of starvation, the government has to rely on short-term measures as well. Also, none of these studies captures the reality of the Indian agricultural market, where the entire value chain is comprised of many stages between the initial production stage executed by the farmers and the final stage of sell to the consumers by the small retailers. This chapter brings in this aspect neglected in the aforementioned studies by analyzing the behavior of the big retailers/middlemen/traders who purchase crops from the farmers and hoard them, before they sell it in the final market to either the small retailers or the ultimate consumers. The chapter, by bringing in this aspect, adds to the existing literature on how the behavior of these traders can be affected so that they do not further contribute to supply shortages and, therefore, to food inflation by speculative buffering.

The role of middlemen or speculators has not been comprehensively studied in the literature. The major area of concern of economists in this area has been whether the behavior of the speculators is based on adaptive expectation or on rational expectation (Chavas 1999; Gillespie and Schupp 2002; Holt and McKenzie 2003). While the former study deals with expectation formation in the US pork market, the latter two are concerned with the US ostrich and broiler markets, respectively. Hardly an endeavor has been made in regulating their behavior. Chavas' empirical attempt finds that in the US pork market, expectation formation is mainly backward looking (for 73 % of the players considered) and only 19 % of the players have rational expectation. In Gillespie and Schupp, there is evidence against rational expectation in the US ostrich industry due to lack of information. In the early stage of development of this industry, speculators expected future demand to increase and this led to a price rise for ostrich. However, markets did not grow as expected, and this led to a price crash in the latter stage. In Holt and McKenzie, the authors fit a quasirational model to the US broiler industry and find that in addition to the quasirational forecast, the true supply shock, future price, and ex post commodity price forecast errors have, at times, been influential in the producer's price expectation. This study says that the extent of supply shock affects price expectation.

However, none of these studies can be likened to the Indian market for agricultural products for various reasons. First, in India, the market for pork or ostrich is very thin and supply shock in these markets hardly has an effect on overall food inflation. Second, food inflation in India is mainly attributable to food grains like wheat, rice, pulses, and recently to vegetables. And third, though price for meat and animal protein has increased over the years, it has not been due to a supply shock caused by an increased demand for nutrition led by higher income of people. The kind of asymmetric information adduced as a reason for failure of the expectation process being rationally formed in Gillespi and Schupp is not applicable in the Indian case, as markets for food products are stable (at least in the short run), and food inflation is primarily due to supply shock. This chapter considers this aspect. Further, it assumes that it is the supply shock that affects the expectation of the traders about future price and, thus, buffering by these agents (as in Holt and McKenzie). One relevant study that has focused on the Indian rice market is due to Ramaswamy (2002), where the author points to the absence of rational expectation in the Indian wheat market. He argues that the agents make regular error in predicting future price. However, even this study does not talk about actions that the government can take to regulate their behaviors and cool down rice price.

One common aspect of the studies in US pork, ostrich, and broiler markets is that supply shock affects the price of the product concerned and, in turn, the behavior of the agents. When the agents expect the future price to increase, they tend to hoard more, affecting the present price in turn. In fact, this kind of speculation activity can be serious enough even to cause a famine (Quddus and Becker 2000). According to these authors, speculative hoarding in the rice market is a major cause of the infamous Bangladesh famine, though not the sole one. A point to be noted is that buffering is not always destabilizing as it helps in consumption smoothing. But asymmetric information and speculation can have a destabilizing effect on price as happened in the Philippines (Shively et al. 2002).

Mitra and Boussard (2011) talk about price volatility due to speculation without any reference to supply shock. The authors assume inter-year storage and fit a non-linear cobweb model. Speculation is adaptive in their model and the result shows using a simulation model that in the presence of interannual storage, price shows less variation than without interannual storage. Their study claims that storage does contribute to the volatility of prices. However, the major lacuna of this chapter is that it may not applicable to nonperishable goods, for which interyear storage is not always feasible, a result which this chapter considers.

Deaton and Laroque (1990; 1996) further contend that due to speculative storage, there is autocorrelation between prices of two periods. That is, high price in one period is translated into high price in another period, which is essentially destabilizing. Quite contrary to this result is the study of Heemeijer et al. (2007), who show that for perishable goods, there may be a negative feedback mechanism. That is, high price in one period may lead to lower price in the next period. However, none of these studies talk about regulating the behavior of the middlemen to reduce such correlation, which is the focus of the chapter.

While the market for agricultural products should ideally be characterized by competition (due to a large number of procurements, middlemen, and final consumers), fixed cost, costs to entry, and collusion mark the departure from perfect competition. Laren (1999) argues that in an oligopolistic framework, buffering activity is such that less is stored and prices are more volatile than it is under perfect competition. While this is a standard textbook result, the main contribution of the chapter lies in the fact that a tax on consumption will not only reduce price volatility but also increase the profit that accrues to the oligopolists. The oligopolistic framework is justified on the basis that there might be large fixed cost to entry.

Basu (2011) rests his arguments in a similar oligopolistic framework. He assumes that there are n oligopolists who play the Cournot game and buy an exogenous amount from the farmers. They maximize their profits in a static framework and supply the minimum of the amount they procured or the profit-maximizing Cournot output. Thus, his policy prescription is that the government must distribute grains or crops to a large number of oligopolists, but in small packets. This will have a favorable impact on prices in two ways. By distributing the grains in small packets, the government will make the amount procured the limiting constraint. Secondly, an increase in the number of players will reduce the market power of the oligopolists and their profits will decrease, and output supplied in the market will increase. The amount of buffering (or wastage) being given by the amount procured and profit-maximizing Cournot output should be zero in the ideal case to ensure the optimal use of grains. Therefore, the optimum amount of procurement should be equal to the profit-maximizing output. Undoubtedly, there are many limitations in this analysis. First and foremost, this is a static model and does not capture the reality: Harvest or production takes place once or twice a year, but consumption occurs throughout the year. Thus, this facet must be considered in any study of behavior of speculators and buffering. Second, buffering is not speculative in his model as agents do not carry over grains from one period to the other depending upon their expectation of future price. The amount of buffering is just an outcome of the profit-maximizing exercise.

This chapter develops on Basu (2011) by adding a few dimensions. First, this chapter is dynamic in nature as it allows for speculative hoarding from one period to the next. Second, it analyzes the behavior of speculators (middlemen) both in a perfectly competitive setup as well as in a monopolistic setup (see Appendix 16.1). Third, this chapter considers both the types of grains where there is public procurement (that there is the Public Distribution System (PDS)) as well as no public procurement and, thereby, seeks to find out whether government intervention (through PDS) is stabilizing or not. Finally, this chapter also addresses how the government can influence the behavior of these traders to reduce speculative hoarding in the event of a negative supply shock.

The aim of this chapter is to study how the decisions of the big retailers/traders or speculators affect the open-market price during a supply crunch, and, secondly, to see if government intervention in the form of PDS can have a favorable impact on the open-market price. In reality, we get to see both types of situations: For some crops like rice, wheat, pulses, and sugar, there is government procurement of these items from the farmers, and analogously for other types of food items like vegetables or fruits, the government does not intervene in the functioning of the market. The purpose of this chapter is to see how the middlemen (synonymous with traders or big retailers in this chapter) aggravate food-price inflation in these two cases during a negative supply shock by the act of speculative buffering. Naturally, if the degree of food inflation is lower in the former case, we argue that there is a case for intervention of the government in the latter case as well. This model is cast in a partial-equilibrium framework. Thus, food inflation is synonymous with general inflation. The rest of the chapter is arranged as follows. Section 2 examines the behavior of traders in aggravating food inflation due to their speculative buffering in the absence of government intervention (i.e., PDS). Section 3 examines the same in the case of those crops for which there is public procurement or PDS. Section 4 talks about the policy intervention, that is, what the government can accomplish to reduce the inflationary impact of a negative supply shock. Section 5 draws the conclusion of the chapter.

2 Role of Middlemen in a Perfectly Competitive Setup under Exogenous Expectation and Without Government Intervention

Let us now formulate the behavior of middlemen in creating food-price inflation. Since there are many crops like onion, vegetables, fruits, and others for which there is no public procurement, we concentrate on this case first. We assume that there are very large numbers of traders (middlemen) who purchase crops directly from the peasants at subsistence price $\bar{w}$. This model is cast in a perfectly competitive setup, so that none of the traders has any market power in either the market from which they buy crops from the farmers or the market in which they sell crops to the final consumers (or say, for the matter of fact to the small-retailers; but here, we assume that the big retailers sell directly to the final consumers). Let each middleman purchase $\bar{x}$ amount of food grain, which we assume to be exogenous. We take the marketable surplus as exogenous because the amount of harvest is a function of the decisions taken by the peasants in the previous period. Since there is no public procurement for these type of crops, there is no restriction on the price paid to cultivators being equal to the procurement price. The amount of output purchased by the middlemen is contingent on the decision taken by the cultivators regarding acreage area, input subsidies received, and other market conditions. We further assume that this amount of produce is exogenous to the model. The middlemen behave as follows: They purchase $\bar{x}$ amount of grain at the beginning of period t directly from the cultivators. Out of the total amount they purchase, they decide how much to sell in period t to the final consumers and how much to carry over for selling in the next period, $t+1$. In period $t+1$, no further arrival of crop occurs. Fresh stock comes only at period $t+2$. The reason for this assumption is that while production and harvest occur only once or twice during a year, consumption occurs throughout the year. For simplicity, we assume production and harvest occur at t while consumption occurs twice, at t and at $t+1$.

Let B_t denote the amount of buffering done by a representative trader in period t. Obviously, if q_t is the total amount of grain sold in period t, then $B_t = [\bar{x} - q_t]$. Out of this amount of buffered grain, a trader decides $q_{t+1,}$ that is, how much to sell in $t+1$. Let the cost of buffering be given by $c[\bar{x} - q_t]^2$. Moreover, out of this buffered amount, let θ fraction of the grain be perished in period $t+1$ and the entire amount of grains be perished beyond $t+1$, that is, grains last maximum for two periods. The incorporation of this fraction, θ, is necessary to include transportation cost in the model (a la the famous iceberg model due to Samuelson), apart from capturing the perishable nature of vegetables, fruits, and other food items. This assumption is essentially valid for vegetables, a major source of the recent food inflation (see Appendix 16.2). Thus, in period $t+1$, a trader can sell the maximum amount $(1-\theta)[\bar{x} - q_t]$.

In period t, the trader knows the price; however, in period $t+1$, he is unaware of the price due to lack of perfect foresight. He can only guess the future price and depending upon this expected price, he decides his optimum allocation between periods t and $t+1$. Let p^e_{t+1} be the expected price of food crop in period $t+1$.

The trader faces two constraints: First, total sale $(q_t + q_{t+1})$ and total amount of wastage $[\theta[\bar{x} - q_t] + \{(1-\theta)[\bar{x} - q_t] - q_{t+1}\}]$ must sum up to the total amount of grain purchased $(\bar{x})$. Second, the total sale in $t+1$ must be less than or equal to effective buffering, that is, $q_{t+1} \leq (1-\theta)[\bar{x} - q_t]$.

A representative trader will maximize his profit subject to the previous two constraints.

Mathematically,

$$\text{Max.}\, \pi = p_t q_t + p^e_{t+1} q_{t+1} - c[\bar{x} - q_t]^2 - \bar{w}\bar{x}$$

s.t.

$$q_{t+1} \leq (1-\theta)[\bar{x} - q_t]. \tag{16.1}$$

$$q_t + q_{t+1} + \theta[\bar{x} - q_t] + \{(1-\theta)[\bar{x} - q_t] - q_{t+1}\} = \bar{x}. \tag{16.2}$$

At equilibrium, Eq. (16.1) implies Eq. (16.2) and, thus, we neglect constraint Eq. (16.2) from the maximizing problem.

The Lagrangian is given by

$$L = p_t q_t + p^e_{t+1} q_{t+1} - c[\bar{x} - q_t]^2 - \bar{w}\bar{x} - \mu\{q_{t+1} - (1-\theta)[\bar{x} - q_t]\}$$

with the Kuhn–Tucker conditions being $\frac{\partial L}{\partial q_t} = 0, \frac{\partial L}{\partial q_{t+1}} = 0, \frac{\partial L}{\partial \mu} \leq 0, \mu \frac{\partial L}{\partial \mu} = 0.$

Now, $\frac{\partial L}{\partial q_t} = 0$ implies

$$p_t + 2c[\bar{x} - q_t] = (1-\theta)\mu \tag{16.3}$$

$$\frac{\partial L}{\partial q_{t+1}} = 0 \text{ which implies } p^e_{t+1} = \mu > 0 \tag{16.4}$$

and, thus,

$$\frac{\partial L}{\partial \mu} < 0. \tag{16.5}$$

Therefore,

$$q_{t+1} = (1-\theta)[\bar{x} - q_t]. \tag{16.6}$$

Lemma 1 A representative trader will not dispose or destroy any part of the food grain (contrary to Basu 2011). That is, the often-heard argument that traders purposely waste/dispose a part of their procurement and allow it to rot, in order to reap higher price per unit creating artificial shortage, is not found to be true in this model. The intuition is that, since fresh stock appears at $t+2$, and grains are perishable beyond two periods, it is profitable for the traders to sell the entire amount of effective buffer stock.

Substituting Eq. (16.4) in Eq. (16.3) and solving for q_t, B_t, and q_{t+1}, we get

$$q_t = \frac{p_t + 2c\bar{x} - (1-\theta)p^e_{t+1}}{2c}. \tag{16.7}$$

Now, it might be feasible that $q_t > \bar{x}$. As a result, the entire amount of output will be sold in period t itself. We neglect this kind of solution, as the main aim of the study is to capture speculative activity of the traders which requires intertemporal storage and sale.

Now,

$$B_t = [\bar{x} - q_t] \text{ or, } B_t = [\frac{(1-\theta)p^e_{t+1} - p_t}{2c}]. \tag{16.8}$$

This result marks a departure from the result derived in Basu (2011) where hoarding had nothing to do with speculation. Clearly, if agents expect the future price to increase, their buffering increases. In his paper, hoarding is simply the gap between procurement and sale. As a result, the profit-maximizing output in Cournot competition is an outcome of the profit-maximizing exercise, sans any speculation.

Now,

$$q_{t+1} = (1-\theta)\frac{(1-\theta)p^e_{t+1} - p_t}{2c}. \tag{16.9}$$

2.1 Determination of Equilibrium Prices

Let there be n number of traders. Thus, total supply of food grain at period t and $t+1$ denoted by AS_t *and* AS_{t+1} is n times the individual supply of grain given in Eqs. (16.7) and (16.9), respectively. Now, let the demand curve be given by $AD_t = a - p_t$ and $AD_{\mathrm{t+1}} = a - p_{t+1}$ for the two periods, respectively. Market equilibrium is given by $AS_t = AD_t$ and $AS_{t+1} = AD_{t+1}$, respectively.

$$\text{That is, } n\left[\frac{p_t + 2c\bar{x} - (1-\theta)p^e_{t+1}}{2c}\right] = a - p_t$$

or

$$p^*_t = \frac{2ac + n(1-\theta)p^e_{t+1} - 2nc\bar{x}}{n+2c}. \tag{16.10}$$

Similarly, we have

$$p^*_{t+1} = \frac{2ac - n(1-\theta)^2 p^e_{t+1} - n^2\bar{x}(1-\theta) + an(2-\theta)}{n+2c}. \tag{16.11}$$

Therefore, equilibrium output levels and amount of food buffered are given by

$$q^*_t = \frac{a - (1-\theta)p^e_{t+1} + 2c\bar{x}}{n+2c} \tag{16.12}$$

$$q^*_{t+1} = \frac{(1-\theta)n\bar{x} - a(1-\theta) + (1-\theta)^2 p^e_{t+1}}{n+2c} \tag{16.13}$$

$$B^*_t = \frac{n\bar{x} - a + (1-\theta)p^e_{t+1}}{n+2c}. \tag{16.14}$$

We certainly need to assume that these variables take positive values. We need another additional assumption here for this system of equations given by Eqs. (16.10)–(16.14) to be positive and meaningful. Since the equilibrium values of prices and quantities are functions of n, the total number of middlemen, we must make n satisfy two conditions: First, n must be sufficiently large enough so that none of the traders are able to exercise any influence on price, that is, agents are price takers. Second, n must not be so large that all short-run profits are eliminated. In other words, n must be such that there is some positive profit.

2.2 Comparative Static Result: Negative Supply Shock

Let us now assume that there is a negative supply shock in output. Our main purpose is to see how speculative buffering adds to price inflation. Considering exogenous expectation, we assume that as there is a negative supply shock, traders expect future

$(t+1)$ prices to increase. Since the traders learn about supply shock at the beginning of period t, they can at best assume that future prices will increase, that is,

$$\frac{\partial p_{t+1}^e}{\partial \bar{x}} < 0 \tag{16.15}$$

Now, differentiating Eqs. (16.10)–(16.14) with respect to $\bar{x}$, and using Eq. (16.15), we get

$$\frac{\partial p_t^*}{\partial \bar{x}} = \frac{n(1-\theta)\dfrac{\partial p_{t+1}^e}{\partial \bar{x}} - 2nc}{n+2c} < 0 \tag{16.16}$$

$$\frac{\partial p_t^*}{\partial \bar{x}} = \frac{-(1-\theta)\dfrac{\partial p_{t+1}^e}{\partial \bar{x}} + 2c}{n+2c} < 0 \tag{16.17}$$

$$\frac{\partial B_t^*}{\partial \bar{x}} = \frac{(1-\theta)\dfrac{\partial p_{t+1}^e}{\partial \bar{x}} + n}{n+2c} \geq < 0 \tag{16.18}$$

$$\frac{\partial p_{t+1}^*}{\partial \bar{x}} = \frac{-n(1-\theta)^2\dfrac{\partial p_{t+1}^e}{\partial \bar{x}} - n^2(1-\theta)}{n+2c} \geq < 0 \tag{16.19}$$

$$\frac{\partial q_{t+1}^*}{\partial \bar{x}} = \frac{(1-\theta)^2\dfrac{\partial p_{t+1}^e}{\partial \bar{x}} + n(1-\theta)}{n+2c} \geq < 0 \tag{16.20}$$

and

$$\frac{\partial q_{t+1}^*}{\partial \bar{x}} \geq < 0 \leftrightarrow \frac{\partial p_{t+1}^*}{\partial \bar{x}} \leq > 0. \tag{16.21}$$

Therefore, as and when there is a crunch is food supply, price in period t necessarily increases and output supplied in period t necessarily decreases. What is noteworthy is that had traders had no expectations about a hike in period $t+1$, that is, had $\frac{\partial p_{t+1}^e}{\partial \bar{x}} = 0$, the increase in price would have been unambiguously less. Buffering (hence, output sold in $t+1$) can either increase or decrease and, accordingly, prices in period $(t+1)$ would decrease or increase post negative supply shock.

Lemma 2 Contrary to the general perception that middlemen will increase buffering and undersupply the market, we discern that buffering may actually decrease during a supply shock, that is, we can have $\frac{\partial B_t^*}{\partial \bar{x}} = \frac{(1-\theta)\frac{\partial p_{t+1}^e}{\partial \bar{x}} + n}{n+2c} > 0$. This result supports the positive-feedback mechanism as formalized in Heemeijer et al. (2007).

Lemma 3 When buffering actually decreases post supply shock, the price in period $(t+1)$ automatically increases, whereas whenever buffering increases due to supply shock, the price in period $(t+1)$ decreases, contrary to the general perception that speculative behavior is always inflationary.

Now, for speculative behavior to inflationary in $(t+1)$ as well, we need $\frac{\partial p^*_{t+1}}{\partial \bar{x}} = \frac{-n(1-\theta)^2 \frac{\partial p^e_{t+1}}{\partial \bar{x}} - n^2(1-\theta)}{n+2c} < 0$. This will be the case when $\left|\frac{\partial p^e_{t+1}}{\partial \bar{x}}\right| < \frac{n}{1-\theta}$. This is the same condition for buffering to decrease due to a supply shock. The intuition for this is that when agents expect future prices to rise less in response to a supply shock, they tend to buffer less and adding to future inflation. Since all traders are identical, this behavior leads to relatively lower supply in period $t+1$. Thus, for larger value of $\frac{\partial p^e_{t+1}}{\partial \bar{x}}$, p^*_t increases and p^*_{t+1} decreases. In other words, when $\frac{\partial p^e_{t+1}}{\partial \bar{x}}$ takes a lower value (lower than $\frac{n}{1-\theta}$), agents expect future prices to rise less and, thus, they supply relatively more in period t and relatively less in period $(t+1)$. Thus, due to lower supply in period $(t+1)$, the food price increases in this period.

Lemma 4 Though future price expectation by traders increase inflation in period t, it actually helps to lower inflation in period $(t+1)$.

The higher the value of $\frac{\partial p^e_{t+1}}{\partial \bar{x}}$, the higher the absolute value of $\frac{\partial p^*_t}{\partial \bar{x}}$ and lower the absolute value of $\frac{\partial p^*_{t+1}}{\partial \bar{x}}$. The reason is the same as for Lemma 3.

What happens to the average price?

We define the average price by

$$AP = \frac{p^*_t + p^*_{t+1}}{2} = \frac{a\{4c+n+n(1-\theta)\} + n\theta(1-\theta)\partial p^e_{t+1} - n\bar{x}\{2c+n(1-\theta)\}}{2(n+2c)}. \tag{16.22}$$

Now, carrying out the usual comparative static result of a negative supply shock, we see that

$$\frac{\partial AP}{\partial \bar{x}} = \frac{n\theta(1-\theta)\frac{\partial p^e_{t+1}}{\partial \bar{x}} - 2nc - n^2(1-\theta)}{2(n+2c)} < 0. \tag{16.23}$$

Thus, average price unambiguously increases due to a supply shock. The second and the last term in the numerator gives the pure marginal impact on the average price of a negative supply shock, while the first term denotes the further increase in the average due to expectation formation by the traders. Thus, it is clear that the effect of a supply shock is further aggravated by the speculative behavior of the middlemen. What is more interesting is that price expectation has a detrimental effect on average price. Thus, even though there might be food-price deflation in the second period, average price will always increase. In fact, the greater the amount of deflation in period $(t + 1)$, the greater is the overall food-price inflation. The reason is clear. Since higher inflation due to speculative behavior by agents leads to greater amount of buffering and, thus, lower supply in period t causing food-price inflation, the opposite happens in period $t + 1$.

Therefore, it is in the interest of the consumers and policymakers to keep the value of $\frac{\partial p^e_{t+1}}{\partial \bar{x}}$ to a minimum level, if not at zero. This brings us to the policy intervention by the government so that this objective can be realized.

3 Role of Middlemen in a Perfectly Competitive Setup Under Exogenous Expectation and with Government Intervention

Let us now look at the case of food grains where public procurement takes place. Throughout the analysis, we assume that there is Universal PDS and not Targeted PDS. For simplicity, we also assume that interyear storage by the government is zero. Hence, the government disburses the entire amount of grain procured. As a result, we use the words government procurement and disbursement synonymously. The first and the foremost alteration that has to be made in the model is that middlemen can no longer buy grains from the peasants at subsistence price, $\bar{w}$. On the contrary, they will have to pay a much larger price, p_c, which is the procurement price of the government. Since peasants can now sell as much as marketable surplus to the government at the procurement price, the middlemen have to pay a price at least as large as p_c, in order to purchase food grains from the cultivators. If the per-unit price that middlemen offer is larger than p_c, then no single farmer will sell to the government. However, this would increase the fixed cost of these traders. Thus, in equilibrium, we must have $\bar{w} = p_c$. Obviously, we assume that the unit cost of selling grains to the government and to the middlemen is the same for all cultivators. However, if this unit cost is different among farmers, then some will sell to the government and some will sell to the middlemen in equilibrium. But the point is $\bar{w}$ can no longer be very less than p_c. Rather, the gap between them will be lesser, easier is the access to the FCI warehouses for the farmers, and greater is the number of middlemen playing in the market. In the case of segregated markets, this gap will be large. Given our assumption of perfect competition, we expect $\bar{w} \to p_c$. The second modification that is required in this model is in the demand function. Since a part of total food demand from the consumers is met through PDS, open-market sale of food grains will have to take care of this. The greater the amount of PDS disbursement, the lower the demand in the open market from the final consumers. Similarly, if the price paid by the beneficiaries of PDS is high, more and more final consumers will migrate towards open market for their purchase. Assuming a linear demand function, the open-market demand function is given by $AD_t = a - p_t + p_r - R$, where p_r and R are the per-unit PDS price and R is the per-period amount PDS offtake. Similarly, for period $(t + 1)$, the aggregate open-market demand is given by $AD_{t+1} = a - p_{t+1} + p_r - R(1 - \theta)$ (since θ fraction of the output procured by the government gets perished). Here, we assume that PDS price and offtake are the same in both the periods. Since the total amount of procurement by the government is $2R$, the following relation holds: $2R + \bar{y} = \bar{x}$, where $\bar{y}$, $\bar{x}$ are the amount of output procured or purchased by the middlemen and marketable surplus, respectively.

Thus, each middleman will maximize his profit as before subject to the same constraints as given in Eqs. (16.1) and (16.2) along with the constraint that $\bar{w} = p_c$. The expressions for q_t, B_t, and q_{t+1} are given by Eqs. (16.7), (16.8), and (16.9), respectively. Market equilibrium conditions are as follows: $AD_t = a - p_t + p_r - R = AS_t = n\left[\frac{p_t + 2cy - (1-\theta)\partial p^e_{t+1}}{2c}\right]$ and $AD_{t+1} = a - p_{t+1} + p_r - R(1-\theta) = AS_{t+1} = n\left[(1-\theta)\frac{(1-\theta)\partial p^e_{t+1} - p_t}{2c}\right]$. Solving for equilibrium values of prices and outputs, we get

$$p^*_t = \frac{2c(a + p_r - R) + n(1-\theta)p^e_{t+1}2nc\bar{y}}{n + 2c} \tag{16.24}$$

$$p^*_{t+1} = \frac{(2c + n + n\theta)(a + p_r - R) - n(1-\theta)^2 p^e_{t+1} - n^2\bar{y}(1-\theta) + R\theta(n + 2c)}{n + 2c} \tag{16.25}$$

$$q^*_t = \frac{a + p_r - R(1-\theta)p^e_{t+1}2c\bar{y}}{n + 2c} \tag{16.26}$$

$$B^*_t = \frac{n\bar{y} - a - p_r + R + (1-\theta)p^e_{t+1}}{n + 2c} \tag{16.27}$$

$$q^*_{t+1} = \frac{(1-\theta)n\bar{y} - (a + p_r - R)(1-\theta) + (1-\theta)^2 p^e_{t+1}}{n + 2c}. \tag{16.28}$$

Once again, we need another additional assumption here for this system of equations given by Eqs. (16.24)–(16.28) to be positive and meaningful. Since the equilibrium values of prices and quantities are functions of n, the total number of middlemen, we must make n satisfy two conditions: First, n must be sufficiently large enough so that none of the traders are able to exercise any influence on price. Second, n must not be too large to eliminate all short-run profits. In other words, n must be such that there is some positive profit for a representative trader.

3.1 Comparison of Prices and Output Levels Between the Cases of No Public Procurement and Public Procurement

We are now in a position to compare the prices and open-market output levels between the aforementioned two cases.

Comparing prices of the first period between the cases of public procurement and no public procurement:

Open-market price level for the first period in the case of public procurement is given by Eq. (16.24), while first-period price for the case of no public procurement is given by Eq. (16.10). Taking their difference, we get

$$\Delta p_t = \frac{2cp_r - 2R(c - 2n)}{n + 2c} \geq< 0, \tag{16.29}$$

where Δ denotes the difference between the first-period price level between the two cases. Now, we will have

$$\Delta p_t < 0 \quad \text{when} \quad R > p_r \frac{c}{c - 2n}. \tag{16.30}$$

In a country like India, where the food market is so large, we expect n, the total number of middlemen to take value in few lakhs or crores. On the other hand, surely the value of c cannot be as high as n, least to say as high as $2n$. Therefore, in all circumstances, the denominator of the left-hand side (LHS) of Eq. (16.30) is going to be negative. As a result, relation Eq. (16.30) will always hold, since by definition, R is positive. Hence, we can conclude that government intervention will reduce open-market price in the first period compared to a state of no intervention.

Lemma 5 When $c < 2n$, open-market price in the first period will be less if the government procures some crop from the farmers and sells the same through PDS. Hence, government intervention will reduce price in the first period.

Comparing the output levels in the two situations, open-market output sold by the traders in the case of public procurement is given by Eq. (16.26) and the case of no public procurement is given by Eq. (16.12). Their difference is, therefore,

$$\Delta q_t = \frac{(p_r - R) - 4Rc}{n + 2c} < 0. \tag{16.31}$$

Since the first term in the numerator of the LHS is the difference between per-unit offtake price of the food item and total amount of procurement/sale by the government in the first period, their difference is always negative. Hence, unambiguously, the difference in output levels is negative. The intuition is that since the government buys a part of the total marketable surplus, output supplied in the open market during the first period is necessarily less in the case of public procurement than under no public procurement.

Comparing the price differences in period $(t + 1)$, the price difference in the case of public procurement and no public procurement is given by

$$\Delta p_{t+1} = \frac{p_r(2n + 2c - n\theta) - R[(n + 2c)(1 - n\theta) + n(1 - \theta)(1 - 2n)]}{n + 2c} \geq< 0. \tag{16.32}$$

Here also, we conjecture that since n ranges in a few lakhs or crores, the terms in the third bracket in the numerator of the LHS is going to be negative. As a result, the LHS will be positive. Hence, we conjecture that

$$\Delta p_{t+1} = \frac{p_r(2n + 2c - n\theta) - R[(n + 2c)(1 - n\theta) + n(1 - \theta)(1 - 2n)]}{n + 2c} > 0. \tag{16.33}$$

Lemma 6 When $n\theta > 1$, open-market price in the case of public procurement will greater than the open-market price in case there is no public procurement.

Analogously, comparing the output levels in the two cases, we get

$$\Delta q_{t+1} = (1-\theta)\frac{p_r + R(1-2n)}{n+2c} \geq< 0. \tag{16.34}$$

Now, this sign is going to negative when we have

$$R > \frac{p_r}{2n-1}. \tag{16.35}$$

Normally, expect the ratio $\dfrac{p_r}{2n-1} << 1$, as n is a very large value and definitely greater than p_r. Thus, output supplied by the traders in the open market will be less in the case of public procurement than in the case of no public procurement. The reason is the same as before, since a part of the marketable surplus is siphoned off due to PDS, less is procured by the middlemen and thus, they supply less in the open market compared to a situation of no PDS.

Next, we analyze the behavior of the average price in the two cases. Now, the average price in the case of public procurement is given by

$$\begin{aligned} AP &= \frac{p_t + p_{t+1}}{2} \\ &= \frac{(a+p_r-R)\{4c+n+n(1-\theta)\}+n\theta(1-\theta)\partial p^e_{t+1} - n\bar{y}\{2c+n(1-\theta)\}+(n+2c)\theta}{2(n+2c)}. \end{aligned} \tag{16.36}$$

Therefore, the difference between the average prices in the two cases of public procurement and no public procurement is given by the difference between Eqs. (16.36) and (16.22) and is illustrated as follows:

$$\Delta AP = \frac{(p_r - R)(4c+n+n(1-\theta)) - n(2c+n(1-\theta))2R + R(\theta)(n+2c)}{2(n+2c)} \geq< 0. \tag{16.37}$$

Now, the expression in Eq. (16.37) will be negative when the following relation holds:

$$R > \frac{p_r\{4c+n+n(1-\theta)\}}{4c(1-n) - \theta(n+2c) + n(1-\theta)(1-2n) + n}. \tag{16.38}$$

Larger the number of players in the market, and higher the marginal cost, more negative will be the denominator of Eq. (16.38). Since the value of n is very large, the denominator of Eq. (16.38) is most likely to be negative. Thus, the relation in Eq. (16.38) will hold true. As a result, the average price in the case of public procurement will be less than under no public procurement. This implies that government procurement will cool down average price.

Lemma 7 Since $4c(1-n)-\theta(n+2c)+n(1-\theta)(1-2n)+n<0$, average price in the case of PDS will be lower than the average price in the case of no PDS.

If we assume that in response to a negative supply shock, when none of the parameters change, and use the relation $2R+\bar{y}=\bar{x}$, then the comparative static result of a supply shock using the same set of assumptions is given by Eqs. (16.16)–(16.21).

The effect on average due a negative supply shock is the same as before (see Eq. (16.23)).

From the above analysis, we can conclude that if the government intervenes in the functioning of the food market by conducting a universal PDS, then this will have a favorable impact on the open-market price. Regarding the amount of total consumption (amount of PDS offtake and open-market purchase) it is the same as the supply of the total amount of food sold by traders in the open market when there is no PDS. Hence, government intervention necessarily makes an individual better off, since the average price a consumer has to pay is smaller and the total amount of consumption is the same.

4 Policy Intervention

In this section, we look at the various possible ways in which government intervention can salvage the economy from the pangs of food-price inflation as well as affect the behavior of middlemen so that so that their expectation of a future price rise during a supply shock can be kept to a minimum. It must be noted in this context that it is the expectation of a future price rise due to a supply shock that adds to the increase in average price. If, somehow, the expectation of a future price rise due to a supply shock can be moderated by the government's policies, then the inflationary impact on food price will be less. This is evident from the expression given in Eq. (16.23). The first term in the numerator gives the aggravation of inflation due to agents' expectation of a future price rise due to the supply shock. The second and the third term are induced by the actual supply shock. Therefore, the extent of food-price inflation can be somewhat reduced if the increase in p^e_{t+1} can be moderated. We must also note that in a Marshallian framework, price will certainly increase if supply is reduced. In this section, we try to discern how the behavior of the middlemen can be controlled so that the extent of price rise is not worsened.

a. *Reliance on import and ban on export*: This has been the conventional measure that the government has resorted to various times in the event of a supply shock. Though there have been cases in which the country has exported food items despite low production because global price is higher than domestic price, the government must have a clear-cut policy that will determine whether export should be carried out or not in such events. Since the data on acreage of crops, weather conditions, pests, etc. are available to the government prior to harvest, this gives leverage to the government to decide on whether to export a particular crop, and if yes, then how much.

   Moreover, in the wake of a negative supply shock, the government should rely on imports. With a clear-cut import policy of the government, the traders will know that if there is a supply shock, then import channels will tend to operate and, hence, final supply in the market will increase to the extent that prices are stabilized. What must be borne in mind is that small retailers must be able to purchase the imported grain at a price less than equal to what they have to pay to the big retailers or the middlemen. In other words, the government might need to give a subsidy so that the cost to the retailers does not increase. With a policy of this sort, the value of $\frac{\partial p_{t+1}^{e}}{\partial \bar{x}}$ will tend to be small, if not zero. This will happen because traders will know that markets will be flooded with imported goods that will drive down the final price. This will put a rein on speculative buffering. Needless to say, this policy will increase the fiscal deficit.

The effect of export ban and greater import will be felt on the domestic currency. Since exports imply supply of foreign currency and imports imply demand for foreign currency, export ban, and greater import will put downward pressure on the domestic currency. Under the managed float exchange rate regime that we have, the Reserve Bank of India (RBI) will intervene in the foreign currency market by going for a monetary contraction, if the exchange rate shoots the comfortable ceilings. As a result, two opposite effects will occur. The depreciation will boost exports of the nonagricultural sector, and hence output and employment; monetary contraction will reduce this initial expansion somewhat by crowding out investment and reversing the improvement of the trade balance by offsetting the initial depreciation of the currency. The actual direction of employment and output of the nonagricultural sector will depend on the relative strength of the monetary contraction vis-à-vis the exchange rate depreciation. If the monetary contraction is such that the exchange rate settles back at its initial level, unambiguously the nonagricultural sector will shrink. However, if the exchange rate intervention is small, there will be some expansion in the external sector but decline in investment will tend to lower final output. Therefore, whenever the country has to resort to imports of food, it is best for the economy if the RBI does not intervene in the foreign exchange market and allows the currency to depreciate. But, in this case, the expansion of the nonagricultural sector will lead to an increase in the indirect demand for agricultural products, which will contribute to further food inflation.

Another problem that might arise in this case is that import price may even be higher than the open-market price. In that case, a subsidy will be required for the retailers/final consumers who will be able to purchase the grain from the government at a price lower than equal to the open-market price. The higher the amount of import, the lower the open-market price will be, and thus the lower the future price expected by the middlemen will be. Now, the problem with subsidy is that it hampers development and poverty alleviation programs of the government. But it must be borne in mind that this type of subsidy is only of a transient nature. Moreover, if the RBI does not intervene in the foreign exchange market so that expansion occurs in the nonagricultural sector, the government will earn greater revenue which will take care of a part of the increase in the fiscal deficit due to the subsidy.

b. *Changing offtake price and offtake amount*: The previous paragraph points to the fact that in the case of grains where there is public procurement, the government cannot buy grains by increasing procurement price and then completely offload the procurement in the open market. If it chooses to do so, then the open-market price may be driven below the procurement price, and, thus, many economic players will reap this arbitrage opportunity. They will buy grains from the open market and sell it back to the FCI at procurement price. As a result, this system will impose a fiscal burden on the government.
The *way in which the government can intervene* in the case of grains where there is already public procurement is by increasing the offtake amount and decreasing the offtake price. The average price for these types of food items (given the exogenous supply of grains, procurement price, PDS offtake, and the number of traders) is given by Eq. (16.36) and is rewritten as follows:

$$AP = \frac{p_t + p_{t+1}}{2}$$
$$= \frac{(a+p_r-R)\{4c+n+n(1-\theta)\}+n\theta(1-\theta)\partial p^e_{t+1}-n\bar{y}\{2c+n(1-\theta)\}+(n+2c)\theta}{2(n+2c)}$$

From the above equation, it can be shown that when a negative supply shock occurs, if the government does not alter the values of offtake price and amount of PDS offtake, then the expression for the change in average price is the same as that given in Eq. (16.23). However, if the government decreases the offtake price and simultaneously increases the disbursement through PDS, then the inflationary impact of a supply shock, which is further enhanced by price expectations by the traders, can be mitigated to some extent. The expression for a change in average price when the government alters the offtake price and disbursement amount is given by

$$\frac{\partial AP}{\partial \bar{x}} = \frac{[4c+n+n(1-\theta)]\dfrac{\partial p_r}{\partial \bar{x}} + n\theta(1-\theta)\dfrac{\partial p^e_{t+1}}{\partial \bar{x}} - n[2c+n(1-\theta)] - \dfrac{\partial R}{\partial \bar{x}}[4c+n+n(1-\theta)+2nc+n^2(1-\theta)]}{2(n+2c)}$$

$$\text{Or, } \frac{\partial AP}{\partial \bar{x}} = \frac{n\theta(1-\theta)\dfrac{\partial p^e_{t+1}}{\partial \bar{x}} - n[2c+n(1-\theta)]}{2(n+2c)} + \frac{[4c+n+n(1-\theta)]\dfrac{\partial p_r}{\partial \bar{x}} - \dfrac{\partial R}{\partial \bar{x}}[4c+n+n(1-\theta)+2nc+n^2(1-\theta)]}{2(n+2c)}. \tag{16.39}$$

As $\frac{\partial p^e_{t+1}}{\partial \bar{x}} < 0$, the term $[n\theta(1-\theta)\frac{\partial p^e_{t+1}}{\partial \bar{x}} - n[2c+n(1-\theta)]]$ is negative. Now, the increase in average price due to a supply shock can be moderated (which is further aggravated by the first term in the expression contained within brackets) if there is a simultaneous increase in PDS offtake and lowering of disbursement price. This is because the term $[\{4c + n + n(1-\theta)]\frac{\partial p_r}{\partial \bar{x}} - \frac{\partial R}{\partial \bar{x}}[4c + n + n(1-\theta) + 2nc + n^2(1-\theta)\}]$ is positive and, thus, mitigates some of the increase in average price due to increase in future expected price. The reason for this result is that these policy measures will divert more consumers to PDS and, thus, demand for grains in the open market falls, which ultimately reduces open-market price. One important change that the PDS system must have is to allow the poorer sections of the society to take their respective quotas in installment, since they may not have enough cash at every point in time.

Let us now look at the feasibility of this government action in the present Indian context. The oft-cited theoretical implication of this type of policy intervention is that it will increase the fiscal deficit. A point to be noted is that the deterioration in the fiscal deficit will be only partial if only the amount of disbursement is increased and the disbursement price remains unaltered. Since procurement price does not alter in this case, the amount of procurement by the government will not increase. Moreover, this will also not increase the fixed cost of the traders. On the other hand, if the government hikes the procurement price, a trader has to pay more to the cultivators and this will increase the fixed cost and, hence, lower the profitability.

One important aspect of the country's PDS is that over time while procurement has increased, disbursement has decreased. While the reasons for a decline in the offtake are many, ranging from poorer coverage, black marketing to lower quality of grains, the reason why procurement has increased is clearly attributable to the continuous increase in MSP and procurement price. Buffering has been above the stipulated norm. The country has registered an acute shortage of storage facilities; as a result, tonnes of grains get wasted. Moreover, if the buffering amount increases over time and never get utilized, it is a leakage from the circular flow of income. It is like money in the pocket which is never used. Given these scenarios in reality, the PDS must, on an urgent basis, device measures on the proper and efficient management of food grains along the lines this chapter suggests. Also, the procurement system of the country should be such that procurement by the FCI must be high in times of negative supply shock and low in times of bumper crops. In the latter case, this will increase the marketable surplus and lead to lower price in the open market, and in the former case, it will reduce the demand for grain in the open market by increasing the amount of procurement. Needless to say, this will increase the fiscal deficit of the government in both the times. But once again, this worsening is only of a transient nature.

Given these measures, this will further help to reduce the expectation of a future increase in open-market price by the traders. The stabilizing mechanism opted by the government will provide a clear-cut signal to the traders that in future, the extent of shock will be less, due to decline in demand for food in the open market. The other effects of expansion in the nonagricultural sector as described will continue to have the same impact.

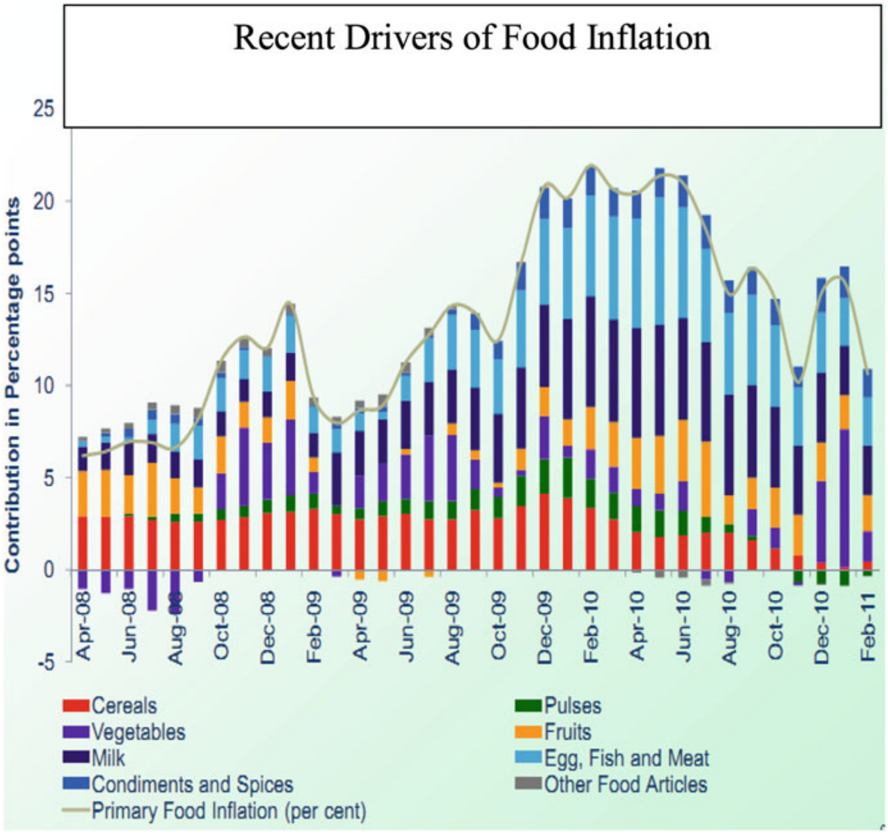
5 Conclusion

This chapter does not deal in the exact expectation formation mechanism of the traders but just assumes that they expect a future price rise when they see a supply shock. In fact, the essence of the model lies in the exogenous expectation regime. Since the agriculture market is very large, it is rational to assume the absence of rational expectation. The absence of rational expectation is further proved in Ramaswamy (2002).

This chapter provides a theoretical foundation to the behavior of big retailers. It shows how big retailers decide on the amount of hoarding on the basis of their future price expectation. It further shows that agents do not voluntarily destroy any part of output with the hope of reaping higher revenue in the future, when they see that the present price has increased. Since food inflation has become a major problem in India during the past 3 years, it is high time that the government takes some concrete short-term as well as long-term steps to put the reins on soaring prices. This chapter spells out some plausible short-run steps that the government can take, without imposing much fiscal costs.

The contribution of this chapter is primarily on the stabilizing effects of price intervention and output intervention by the government. Government intervention not only affects directly the open-market price in this model but also indirectly affects the future expectations of prices formed by traders. A clear-cut policy will carry the signal to the traders about the possible direction of government intervention in the event of a supply shock, and, thus, would insulate the price of output from the destabilizing speculation of the agents. When the agents know that the government intervention will moderate the extent of the shock, their expectation of a serious price rise will also be moderated. As a result, the food inflation will be purely due to the extent of shock and not due to the wrong hoarding actions of the middlemen. This is essentially how the government can control the behavior of the middlemen.

Appendix 16.1



References

Basu K (2011) India's foodgrain policy: an economic theory perspective. EPW xlvi(5):37–45

Chand R (1996) Diversification of Punjab agriculture: need and implications for food security. Agricu Econ Res Rev 9(2):156–163

Chand R (2012) Development policies and agricultural markets. Econ Polit Wkly XLVII(52) (December 29, 2012)

Chand R, Gulati A, Shinoj P, Ganguly K (2011) Managing Food Inflation in India: reforms and Policy Options. Policy Brief 35. National Centre for Agricultural Economics and Policy Research, New Delhi

Chavas JP (1999) On the economic rationality of market participants: the case of expectations in the U.S. pork market. J Agr Resour Econ 24(1):19–37

Deaton A, Guy L (1992) On the behaviour of commodity prices. Rev Econ Stud 59(1):1–23

Gillespie J, Schupp A (2002) The role of speculation and information in the early evolution of the United States ostrich industry: an industry case study. Rev Agric Econ 24(1):278–292

Heemeijer P, et al (2009) Price stability and volatility in markets with positive and negative expectations feedback: an experimental investigation. J Econ Dyn Control 33(5):1052–1072. doi:10.1016/j.jedc.2008.09.009

Holt MT, McKenzie AM (2003) Quasi-rational and ex ante price expectations in commodity supply models: an empirical analysis of the US broiler market. J Appl Econom 18(4)(Jul.–Aug): 407–426

Mitra S, Jean-Marc B (2011) Storage and the volatility of agricultural prices: a model of endogenous fluctuations. http://www.fordham.edu/economics/

Munir Q, Charles B (2000) Speculative price bubbles in the rice market and the 1974 Bangladesh famine. J Econ Develop 25(2):155–175

Ramaswami B, Balakrishnan P (2002) Food prices and the efficiency of public intervention: the case of the public distribution system in India. Food Policy 27(5–6):419–436, Elsevier

Virmani A, Rajeev PV (2001) Excess food stocks, PDS and procurement policy. Planning Commission Working Paper No. 5/2002–PC

Chapter 17
Interstate Variations in Levels and Growth of Industry: Trends During the Last Three Decades

T. S. Papola

1 Introduction

Inter-regional disparity in the levels of economic development and per capita income has been a major issue in development debate and policy in India. There are large variations in the different indicators of development among the states which finally get reflected in the differences in per capita incomes and levels of living. There have, of course, been changes in the extent of disparities and in the relative positions of different states over the years. Some decline in overall inequalities in per capita income among states was observed in the initial two to three decades after Independence, but there has been an increasing trend since then. The Gini coefficient of interstate inequality in per capita income was 0.152 in 1981 and increased to 0.225 by 1997–1998 (Ahluwalia 2000). In the post-2000 period, some of the poorer states have registered faster than average growth in gross state domestic product (GSDP) and growth of some of the developed states has slowed down. As a result, the Gini coefficient has remained at around 0.24 during 2001–2009 (Ahluwalia 2011).

It is generally argued that it is primarily the level of industrialization and growth of industry that determine the relative levels of economic development of different regions, for development of agriculture is primarily dependent on the quantity and quality of land which is more or less given, and growth of services mostly follows the growth of agriculture and industry. It is for this reason that most policy instruments for balanced regional development such as investment licences and fiscal and financial incentives that have been adopted in India have been directed towards industry, with the overall objective of 'industrial development of backward areas'.

Most of these policy measures have been discontinued since the introduction of economic reforms in the early 1990s. At the same time, the rate of economic growth has significantly accelerated, in which industry has played a role, even though not the major one. How has the industrial growth in post-reform period been distributed

T. S. Papola (✉)
Institute of Studies in Industrial Development (ISID),
Institutional Area Phase II Vasant Kunj, P.B. No. 75134, 110070 New Delhi, India
e-mail: trilokp41@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_17,

across states? Expectations were rather conflicting. On the one hand, discontinuation of policies favouring industrially backward areas could discourage industrial investment in less-industrialized states and, thus, increase disparities. On the other hand, deregulation permitting free flow of goods and services, internally and externally, would encourage poor states to better utilise their comparative advantage, thus leading to a decline in disparities in industrial development. It is, therefore, interesting to study the pattern of industrial growth in the post-reform period when most of the 'interventionist' measures have been removed in comparison with the pre-reform period when they were in place.

It is in this context that this chapter looks at the changes in the levels of industrialization, rates of industrial growth and shares of different states in all-India industrial output and employment. In the process, it also examines whether rates of industrial growth and changes in the levels of industrialization have gone together with GSDP growth rates of different states. The chapter also makes an attempt to examine the factors that have led to differences in the rates of industrial growth, particularly, in the more recent period. It may be noted that 'industry' is confined to 'manufacturing', in this chapter.

2 The Extent of Industrialization: Trends towards Convergence or Divergence among States?

Differences in the extent of industrialization are one of the most glaring aspects of the variations in the levels and structure of state economies. The share of manufacturing in the GSDP varies very widely among the Indian states. In terms of this indicator, Gujarat with about 30 % share of manufacturing in GSDP was the most industrialized state among the major states of India in 2008–2009 (Table 17.1). Other major states which had a higher than the national figure of 17 % were Maharashtra (23.46 %), Tamil Nadu (23.32 %), Haryana (20.0 %), Karnataka (19.85 %) and Orissa (17.04 %). Kerala had the lowest 9.96 % of its state domestic product (SDP) originating in manufacturing. Andhra Pradesh followed by Bihar and Uttar Pradesh were other states with a low level of industrialization with only 12–14 % of their SDP originating in manufacturing.

Among the three new states—Chhattisgarh, Jharkhand and Uttarakhand—Chhattisgarh and Jharkhand feature as relatively better-industrialized states with 21.94 and 32.02 % share of manufacturing in their SDP. Uttarakhand with 14.12 % of its SDP from manufacturing is among the states with a low level of industrialization. All states in the northeastern region except Assam (10.74 %) had less than 10 % of their SDP from the manufacturing industry. Among union territories (UTs) and other states, Pondicherry (65.49 %) and Goa (30.08 %) showed a relatively high degree of industrialization. The share of industry in gross domestic product (GDP) ranged between 9.96 % in Kerala, the least industrialized state, and 29.94 % in Gujarat, the most industrialized state, in 2008–2009. The range of variation seems to have marginally declined from 1980 to 1981, when the least industrialized state

Table 17.1 Share of manufacturing in total GSDP (%) at 1993–1994 prices. (Source: www.mospi.gov.in)

		1980–1981	1990–1991	2000–2001	2008–2009
Major States					
1	Andhra Pradesh	13.86	15.32	13.69	12.05
2	Bihar (+)	9.92	12.56	9.17 (3.73)	13.27 (2.50)
3	Gujarat[a]	18.92	26.14	30.41	29.94
4	Haryana[b]	13.65	19.10	20.59	20.00
5	Karnataka	15.25	18.63	17.26	19.85
6	Kerala[a]	9.52	11.11	11.68	9.96
7	Madhya Pradesh (+)	11.11	15.50	16.46 (15.08)	15.35 (12.73)
8	Maharashtra[a]	24.92	26.08	23.93	23.46
9	Orissa	9.08	11.29	12.13	17.04
10	Punjab	9.21	13.61	15.96	16.05
11	Rajasthan	12.43	12.36	16.50	15.63
12	Tamil Nadu	31.47	28.54	24.36	23.32
13	Uttar Pradesh (+)	9.01	13.87	13.85 (14.00)	14.02 (14.01)
14	West Bengal[a]	20.31	17.80	17.28	16.37
New States					
15	Chhattisgarh	–	–	18.50	21.94
16	Jharkhand	–	–	19.17	32.02
17	Uttarakhand	–	–	11.74	14.12
North Eastern States					
18	Arunachal Pradesh[a]	3.80	2.60	3.43	2.03
19	Assam	9.55	9.17	7.67	10.74
20	Manipur	6.41	13.53	7.93	7.48
21	Meghalaya	1.80	2.42	2.07	8.49
22	Mizoram	1.49	2.87	1.73	2.13
23	Nagaland[b]	5.09	3.65	1.12	1.40
24	Sikkim	0.00	0.00	4.13	3.48
25	Tripura[a]	3.44	2.78	4.85	2.82
Union Territories and Other States					
26	A&N Islands[a]	7.27	6.39	4.80	3.35
27	Chandigarh	N.A.	N.A.	15.63	12.72
28	Delhi	8.25	8.94	11.49	8.80
29	Dadra and Nagar Haveli	N.A.	N.A.	N.A.	N.A.
30	Daman and Diu	N.A.	N.A.	N.A.	N.A.
31	Lakshadweep	N.A.	N.A.	N.A.	N.A.
32	Pondicherry	20.39	28.74	49.10	65.49
33	Goa[a]	24.24	22.29	33.26	30.08
34	Himachal Pradesh[a]	3.01	7.32	15.02	13.64
35	Jammu & Kashmir[a]	N.A.	N.A.	5.86	8.10
	India	*13.80*	*16.60*	*17.20*	*17.00*
	SD	*6.78*	*5.82*	*5.74*	*5.29*
	CV	*45.52*	*33.70*	*33.06*	*30.08*

Figure in parentheses against Bihar, Madhya Pradesh and Uttar Pradesh are for the territory after division while those outside include newly formed Jharkhand, Chhattisgarh and Uttarakhand, respectively, in this as well as other tables

Estimates of standard deviation (*SD*) and coefficient of variation (*CV*) are based on 14 major states

N.A. not available

[a] Latest available data are for the year 2007–2008

[b] Latest data available are for the year 2006–2007

(Kerala) had 9.52 % of its SDP originating from manufacturing, while in the most industrialized state (Tamil Nadu) manufacturing contributed 31.47 %. But the states in the most industrialized category have changed their relative positions. In fact, West Bengal which held the second position in 1980–1981 has gone out of the group of the top five to the seventh position. Haryana which was below the national average has acquired the fourth position. Tamil Nadu yielded its first position in 1980–1981 to Gujarat in 2008–2009; the latter held the fourth position in 1980–1981. Orissa which had a much lower than the national extent of industrialization rose to the national average in 2008–2009. Other states which have experienced relatively rapid industrialization during the 28-year period in terms of a significant increase in the share of manufacturing in GSDP are Karnataka, Punjab, Madhya Pradesh, Rajasthan and Uttar Pradesh. Gujarat, of course, had the fastest advance in industrialization, raising its manufacturing share in SDP from 19 % in 1980–1981 to 30 % in 2008–2009. Among smaller states and UTs, Himachal Pradesh (from 3.01 % in 1980–1981 to 13.64 % in 2008–2009) and Pondicherry (from 20.39 % in 1980–1981 to 65.49 % in 2008–2009) made rapid advance in industrialization.

West Bengal saw a ‘deindustrialization’ insofar as manufacturing contributes now only 16.4 % in SDP as compared to 20.3 % 28 years back. Maharashtra and Andhra Pradesh also experienced some decline in the share of manufacturing in their SDP from 25 to 24 % and from 14 to 12 %, respectively. Northeastern states in which some such decline has taken place are Arunachal Pradesh (3.80–2.03 %), Nagaland (5.09–1.40 %) and Tripura (3.44–2.82 %). Andaman and Nicobar Islands also saw a significant decline in the share of manufacturing SDP from 7.27 to 3.35 %.

It is noteworthy that most states have seen either a decline or virtual stagnation in the extent of industrialization, in the post-reform period. Only three among the major states, Orissa, Punjab and Rajasthan, have experienced an increase in the share of manufacturing in their GSDP since 1990–1991. Among smaller states, Himachal Pradesh features in this category .

Amidst changes in different directions and of varying extent, the overall disparity in the *degree of industrialization* seems to have declined. Both standard deviation (SD) and coefficient of variation (CV) have declined from one decade to another since 1980–1981. SD declined from 6.78 in 1980–1981 to 5.82 in 1990–1991 and further to 5.29 in 2008–2009 and CV from 45.52 % in 1980–1981 to 33.70 % in 1990–1991 and to 30.08 % in 2008–2009 (Table 17.1).

3 Industrialization, SDP Growth Rate and Structural Transformation

Has a faster pace of industrialization been accompanied also by a larger transformation of state economies from agricultural to non-agricultural? Is there a direct relationship between the increase in the share of manufacturing and decline in that of agriculture, as has been conventionally presumed? In this connection, it needs to be noted that over the years 1980–1981 to 2008–2009, the share of agriculture in the national GDP declined from 39.70 to 16.20 % (Table 17.2). This decline has,

Table 17.2 Share of agriculture in total GSDP (%) at 1993–1994 prices

		1980–81	1990–91	2000–01	2008–09
Major States					
1	Andhra Pradesh	38.66	33.31	28.61	22.23
2	Bihar(+)	52.45	43.84	38.43 (46.56)	25.74 (31.62)
3	Gujarat[a]	38.21	27.02	15.19	16.00
4	Haryana[b]	49.09	42.94	32.07	23.10
5	Karnataka	43.56	33.45	26.37	13.83
6	Kerala[a]	41.70	31.16	23.64	15.68
7	Madhya Pradesh (+)	47.30	38.01	24.03 (25.87)	23.99 (26.23)
8	Maharashtra[a]	25.53	20.73	15.49	13.35
9	Orissa	54.59	38.69	28.22	19.24
10	Punjab	46.41	46.02	39.21	32.55
11	Rajasthan	43.80	41.11	26.73	24.00
12	Tamil Nadu	25.25	22.75	17.62	10.99
13	Uttar Pradesh (+)	48.05	39.27	35.60 (35.65)	27.72 (28.37)
14	West Bengal[a]	31.94	30.95	26.06	20.70
New States					
15	Chhattisgarh	–	–	18.25	18.33
16	Jharkhand	–	–	23.49	15.48
17	Uttarakhand	–	–	34.88	28.37
North Eastern States					
18	Arrunachal Pradesh[a]	44.96	31.79	28.99	16.31
19	Assam	49.21	41.48	34.02	23.93
20	Manipur	28.76	35.44	32.89	26.36
21	Meghalaya	41.75	29.45	25.06	21.03
22	Mizoram	26.96	21.14	19.67	15.38
23	Nagaland[b]	27.57	24.70	33.94	35.51
24	Sikkim	41.08	34.75	21.86	16.66
25	Tripura[a]	56.00	42.09	32.05	28.59
Union Territories and Other States					
26	A&N Islands[a]	43.69	47.39	29.32	11.90
27	Chandigarh	N.A.	N.A.	1.10	0.53
28	Delhi	4.28	2.98	1.31	0.63
29	Dadra and Nagar Haveli	N.A.	N.A.	N.A.	N.A.
30	Daman and Diu	N.A.	N.A.	N.A.	N.A.
31	Lakshadweep	N.A.	N.A.	N.A.	N.A.
32	Pondicherry	29.08	18.90	6.95	3.52
33	Goa[a]	20.55	14.53	8.44	4.46
34	Himachal Pradesh[a]	44.21	35.51	23.41	18.99
35	Jammu & Kashmir[a]	N.A.	N.A.	32.17	28.57
	India	*39.70*	*32.20*	*23.90*	*16.20*

Figure in parentheses against Bihar, Madhya Pradesh and Uttar Pradesh are for the territory after division while those outside include newly formed Jharkhand, Chhattisgarh and Uttarakhand, respectively, in this as well as other tables

N.A. not available

[a] Latest available data are for the year 2007–2008

[b] Latest data available are for the year 2006–2007

however, not meant a corresponding gain in the share of manufacturing which has increased at a much smaller pace, from 13.80 to 17.00 %. Major gain in the share has been for the services which rose from 36.60 % in 1980–1981 to 57.30 % in 2008–2009.

The phenomenon of a shift mainly from agriculture to services is observed in the case of most of the major states. Yet in some cases, particularly where industrialization has been rapid, decline in agriculture has been accompanied, to a large extent, by an increase in industry. Thus, in the case of Gujarat, share of agriculture declined from 38 to 16 %, that is, by 22 percentage points; it was accompanied by an equal increase in the share of both manufacturing and services, by 11 percentage points each (Tables 17.1 and 17.3). Similarly, in Orissa, a decline in the share of agriculture was accompanied by an increase not only in the share of services but also in manufacturing to a significant extent. On the other hand, in Kerala and Karnataka, services have taken the major share of the loss in the share of agriculture. In Punjab, agriculture has seen a relatively smaller decline in its share: It is the only state in which it still contributed almost one-third (32.6 %) of GSDP. The decline in the share of agriculture has, however, benefitted industry more than services. West Bengal is another stand-alone case with everything happening rather slowly: Agricultural GDP has declined by 11 percentage points only (against 24 % at the national level), industry share has significantly declined and that of services increased much less than the national average. Tamil Nadu is yet another exceptional case, where share of agriculture has sharply declined—it is now at the lowest (11 %) in any state—and share of manufacturing has also significantly declined, and all the gains have gone to services sector only. Among smaller states and UTs, a very sharp shift from agriculture to non-agricultural sector is observed in the case of Goa and Pondicherry. In the case of Goa, share of agriculture declined from 21 to 4 %, which was mostly compensated by an increase in the share of services from 40 to 56 %, Pondicherry saw a decline in the share of agriculture from 29 to 4 %; manufacturing increased its share by 45 percentage points from 20 to 65 %.

There are two questions that are of significant interest with regard to the relationship between growth and structural changes. One, has growth rate and structural transformation (shift from agriculture to non-agriculture) gone together? And two, which type of structural transformation, one characterized by shift to manufacturing or to services, has been more growth augmenting? Gujarat has been the fastest growing state during the entire period 1980–1981/2008–2009 and in both the sub-periods since 1991, having recorded a GSDP growth rate of 9.48 % during 1991–2001 and 11.71 % during 2001–2009 (Appendix A). It also has undergone a large transformation with share of agriculture in GSDP declining from 38 % in 1980–1981 to 16 % in 2008–2009. The largest transformation, has, however, been experienced by Karnataka reducing share of agriculture in its GSDP from 44 to 14 % during 1981–2009. Its rate of growth has also been quite high in recent years. Orissa has experienced the second highest growth after Gujarat during 2001–2009, and it has also seen rapid transformation in its economy: Share of agriculture in its SDP declined from 55 % in 1980–1981 to 28 % in 2000–2001 and to 19 % in 2008–2009. Kerala is another state where both growth rate and structural transformation have been fast. Slowest

Table 17.3 Share of services in total GSDP (%) at 1993–1994 prices

		1980–1981	1990–1991	2000–2001	2008–2009
Major States					
1	Andhra Pradesh	39.26	41.71	46.54	51.25
2	Bihar (+)	28.02	31.95	39.76 (43.39)	45.41 (51.28)
3	Gujarat[a]	33.22	37.34	44.18	44.38
4	Haryana[b]	25.39	29.81	40.18	46.43
5	Karnataka	31.59	39.17	46.13	54.53
6	Kerala[a]	40.92	50.35	56.09	60.73
7	Madhya Pradesh (+)	27.99	33.36	39.82 (40.55)	38.22 (39.71)
8	Maharashtra[a]	39.94	43.86	53.36	57.20
9	Orissa	27.16	34.76	43.38	45.07
10	Punjab	36.18	33.48	36.92	41.27
11	Rajasthan	33.94	35.12	41.15	41.90
12	Tamil Nadu	36.73	39.98	47.93	57.10
13	Uttar Pradesh (+)	33.94	37.90	40.30 (40.34)	42.00 (42.44)
14	West Bengal[a]	40.38	43.34	49.35	53.50
New States					
15	Chhattisgarh	–	–	37.55	34.44
16	Jharkhand	–	–	33.09	35.17
17	Uttarakhand	–	–	39.81	37.07
North Eastern States					
18	Arunachal Pradesh[a]	29.04	23.08	34.24	23.31
19	Assam	31.57	35.34	44.58	51.05
20	Manipur	23.13	41.59	46.24	41.03
21	Meghalaya	42.46	49.88	53.45	50.79
22	Mizoram	59.10	46.15	64.42	62.46
23	Nagaland[b]	52.78	59.14	53.46	48.70
24	Sikkim	41.63	51.34	52.91	50.00
25	Tripura[a]	39.37	49.84	59.23	58.42
Union Territories and Other States					
26	A&N Islands[a]	34.16	29.64	50.31	34.39
27	Chandigarh	N.A.	N.A.	72.74	72.20
28	Delhi	82.32	83.06	78.72	81.88
29	Dadra and Nagar Haveli	N.A.	N.A.	N.A.	N.A.
30	Daman and Diu	N.A.	N.A.	N.A.	N.A.
31	Lakshadweep	N.A.	N.A.	N.A.	N.A.
32	Pondicherry	34.56	37.44	40.77	29.38
33	Goa[a]	39.53	50.61	47.94	55.88
34	Himachal Pradesh[a]	33.65	38.69	41.57	40.95
35	Jammu & Kashmir[a]	N.A.	N.A.	51.44	48.76
	India	*36.60*	*40.60*	*46.90*	*57.30*

Figure in parentheses against Bihar, Madhya Pradesh and Uttar Pradesh are for the territory after division while those outside include newly formed Jharkhand, Chhattisgarh and Uttarakhand, respectively, in this as well as other tables

N.A. not available

[a] Latest available data are for the year 2007–2008

[b] Latest data available are for the year 2006–2007

Table 17.4 Relationship between structural change and its components and rate of growth of GSDP (correlation coefficients)

	1980–1981/ 1990–1991	1990–1991/ 2000–2001	2000–2001/ 2008–2009	1980–1981/ 2008–2009
Correlation between growth of GSDP & % change in the share of agriculture during 1980–1981/2008–2009	0.275	– 0.176	– 0.676[a]	– 0.181
Correlation between growth of GSDP & % change in the share of manufacturing during 1980–1981/2008–2009	0.078	0.038	0.029	0.056
Correlation between growth of GSDP & % change in the share of services during 1980–1981/2008–2009	0.010	– 0.040	0.429	0.082

[a] Significant at 0.01 level

transformation is observed in Punjab and West Bengal; both have also had slow growth of GSDP. Madhya Pradesh and Uttar Pradesh are also in the same category. Andhra Pradesh, Haryana and Rajasthan have grown relatively faster though the process of transformation has been rather slow in these states. Maharashtra already had a relatively low share of agriculture initially, seeing a significant decline in it and a reasonably high growth rate.

Among the northeastern states, Mizoram, Nagaland and Sikkim are the fastest growing states, having recorded a GSDP growth rate of 10 % per annum during 1981–2009. Mizoram and Sikkim have also undergone a large transformation with share of agriculture in GSDP declining during 1981–2009, from 27 to 15 % and from 41 to 17 %, respectively. Nagaland, however, seems to have experienced an increase in the share of agriculture from 28 to 36 %. Andaman and Nicobar Islands, Pondicherry and Goa also have very large transformation from agriculture to non-agriculture and a very high growth particularly during 2001–2009.

Insofar as decline in the share of agriculture is taken as a measure of structural transformation, its relation with growth of GSDP has been rather weak ($r = -0.181$) if we take the long period 1981–2009. Yet the two have been significantly related in the shorter period, 2001–2009, where $r = -0.676$. States with faster decline in the share of agriculture also seem to have recorded faster growth of GSDP, during this period. Changes in the share of manufacturing or services, either in the short or long term, do not seem to have any significant relation with GSDP growth rates in states (Table 17.4).

Punjab has seen the slowest transformation in its economy: Over a period of almost 30 years, the contribution of non-agricultural sectors has increased from 54 to 66 % only. It still derives about one-third of its SDP from agriculture, the highest in any state. Its growth rate has been one of the lowest around 5 %, against the national average of 7 %, during 1980–1981/2008–2009. During 2000–2001/2008–2009 when the national economy grew at 8.3 % per annum, the Punjab economy grew at 5.4 %. Strangely enough, Tamil Nadu, the state with the largest structural transformation of the economy, with the lowest, 11 %, share of agriculture in SDP, has also not done very well in terms of the growth of its GSDP. The state experienced an average growth rate of 6.5 % over the period 1980–1981/2008–2009, though it has accelerated to 7.6 % during 2000–2001/2008–2009.

Did structural transformation in favour of manufacturing help in accelerating growth of a state? Here again, Gujarat provides strong positive evidence: It increased share of manufacturing in its GSDP from 19 % in 1980–1981 to 30 % in 2008–2009 and experienced the fastest economic growth overall. Orissa and Haryana are other states with significantly large increase in the share of manufacturing and both of them have grown reasonably fast. Bihar, Karnataka, Madhya Pradesh, Rajasthan and Uttar Pradesh have moderate increase in the share of manufacturing and relatively low GSDP growth. Punjab with significantly large increase in manufacturing share experienced low growth. Maharashtra and West Bengal both saw a decline in manufacturing share; while the former grew reasonably well, the latter grew at a relatively slow rate. On the whole, there appears to be a positive relation between the increase in the extent of industrialization and the rate of economic growth. This relation that holds in the case of most of the 14 major states is also observed in the case of Assam, Meghalaya, Pondicherry and Goa which have experienced a large increase in the share of manufacturing along with high growth rates. Himachal Pradesh, with significantly large increase in manufacturing share, on the other hand, experienced low growth.

There are few major states where the services sector has played a more important role in economic growth. Kerala, which now has the highest share (60.7 %) of services in its GSDP, rising from 41 % in 1980–1981 while the share of manufacturing remaining constant at around 10 % (Table 17.3), registered a reasonably high growth. So did Haryana with services share rising from 25 to 47 % and Karnataka from 32 to 55 %. The services sector has played an important role in economic growth in most of the northeastern states, Goa and Himachal Pradesh. Tamil Nadu and West Bengal did not see a large increase in the share of services nor did they experience very high growth rates. It appears that unlike in the country as a whole, services did not make a major contribution to growth in most states in recent years. It is only a few states which had a high weight of services and experienced high growth in that sector that seems to have been reflected in what is called a 'service-led growth' nationally. In most states' industry, particularly manufacturing seems to have made a more significant contribution to the growth of GSDP. In other words, a structural change in favour of manufacturing is more often accompanied by a higher GSDP growth than a change in favour of services. The relationship, however, does not turn up to be consistent once all states are taken together for comparison, as some have had manufacturing, while others have had services, pushing the GSDP growth. As a result, the coefficient of correlation between growth rates and change in the share of manufacturing and of services are not significant in the shorter or longer periods as noted earlier. It appears that faster growth of non-agricultural sectors as a whole, irrespective of whether it is derived from manufacturing or services, leads to high growth of GSDP.

4 Rates of Industrial Growth

How have different states performed in terms of the growth of manufacturing SDP over the longer period 1980–1981 to 2008–2009 and in the post-reform period, particularly during 2001–2009 when national aggregate growth rate has been relatively high. Gujarat is the only major state which has maintained high and accelerating growth rates over the years: Its manufacturing sector grew at more than 8 % during 1981–1991, at 9.5 % during 1991–2001 and a much higher rate of 11.7 % during 2001–2009 (Table 17.5). Among other better-industrialized states, Maharashtra maintained a moderate growth rate of 6–8.5 %. Tamil Nadu had a much lower average growth rate of about 6 %; only during 2001–2009, it attained a growth of 7.7 % per annum, and West Bengal's manufacturing sector grew at a still lower rate, averaging about 5 % over the entire period and slightly more than 6 % during the post-reform period.

Some of the less-industrialized states have shown spectacular growth of manufacturing during 2001–2009. Orissa registered a manufacturing growth of 15.6 % and Bihar 13.9 % during this period. Karnataka has also recorded a manufacturing growth of 10.5 %. Haryana and Punjab had a significantly high growth of this sector during 1981–1991, but it decelerated in the following two decades, especially in Punjab, where it has been only 6 % as against the national average of more than 10 %. Similar is the case with Uttar Pradesh. Andhra Pradesh and Kerala have maintained a relatively low growth over the whole period. All the three new states have registered a high growth rate in manufacturing GSDP during 2001–2009, Jharkhand having the highest, about 17 % growth rate. Among other states and UTs, Meghalaya, Pondicherry and Himachal Pradesh registered relatively high, more than 11 % rate of growth over the entire period 1981–2009.

Growth rates of manufacturing in different states seem to show a tendency towards divergence over the longer period. The CV among growth rates of different states was 33 % during 1981–1991, it declined to 28 % during 1990–1991/2000–2001, but increased to 36 % during 2000–2001/2008–2009. Also, while better-industrialized states grew slower than the less industrialized during 1981–1991, the reverse seems to have happened in recent decades. Correlation between initial level of industrialization and growth rate was negative during 1981–1991 (− 0.317); it turned positive and significant during 1991–2001 (0.484) and 2001–2009 (0.601). Thus, it appears that the trend towards a decline in differences in the level of industrialization among states observed in earlier years has been reversed in the post-reform period.

5 Shares of States Manufacturing

Maharashtra has always accounted for the largest share in manufacturing output of the country. In 2006–2007, it contributed about one fifth of the manufacturing GSDP of all the states of India. It has maintained that share all along though there is a small decline in it from that in 1980–1981 (Table 17.6). Tamil Nadu used to be the

Table 17.5 Growth rate of manufacturing GSDP (at 1993–1994 prices). (Source: same as Table 17.1)

		1980–1981/ 1990–1991	1990–1991/ 2000–2001	2000–2001/ 2008–2009	1980–1981/ 2008–2009
Major States					
1	Andhra Pradesh	5.36	5.20	6.92	5.10
2	Bihar (+)	6.24	3.18	13.95 (1.44)	3.94
3	Gujarat[a]	8.29	9.48	11.71	8.17
4	Haryana[b]	10.42	6.80	8.13	7.33
5	Karnataka	7.07	6.90	10.51	7.42
6	Kerala[a]	3.26	5.92	6.19	5.12
7	Madhya Pradesh (+)	6.52	6.58	5.44 (2.26)	5.82
8	Maharashtra[a]	6.79	6.27	8.64	6.29
9	Orissa	8.78	4.17	15.60	6.68
10	Punjab	8.98	6.43	6.18	6.49
11	Rajasthan	6.66	9.37	7.84	6.96
12	Tamil Nadu	4.06	5.06	7.70	4.56
13	Uttar Pradesh (+)	9.53	4.80	6.26 (5.85)	5.65
14	West Bengal[a]	3.32	6.36	6.07	5.21
New States					
15	Chhattisgarh	–	–	11.66	–
16	Jharkhand	–	–	16.88	–
17	Uttarakhand	–	–	12.15	–
North Eastern States					
18	Arunachal Pradesha	8.14	7.10	2.85	6.56
19	Assam	2.96	1.87	8.86	3.91
20	Manipur	7.81	3.37	5.19	4.46
21	Meghalaya	7.50	7.74	14.85	11.22
22	Mizoram	9.85	5.42	9.27	7.81
23	Nagaland[b]	11.73	−0.55	8.38	6.11
24	Sikkim	N.E.	N.E.	6.55	N.E.
25	Tripura[a]	3.05	12.82	4.52	8.44
Union Territories and Other States					
26	A&N Islands[a]	2.63	3.87	7.56	2.80
27	Chandigarh	N.E.	N.E.	9.20	N.E.
28	Delhi	8.04	3.35	5.83	5.47
29	Dadra and Nagar Haveli	N.E.	N.E.	N.E.	N.E.
30	Daman and Diu	N.E.	N.E.	N.E.	N.E.
31	Lakshadweep	N.E.	N.E.	N.E.	N.E.
32	Pondicherry	7.44	19.53	14.02	13.05
33	Goa[a]	0.71	10.68	8.68	8.08
34	Himachal Pradesh[a]	14.52	14.90	6.65	12.46
35	Jammu & Kashmir[a]	N.E.	N.E.	11.03	N.E.
	India	*7.44*	*7.02*	*8.20*	*6.77*
	SD	*2.26*	*1.74*	*3.15*	*1.20*
	CV	*33.15*	*28.21*	*36.38*	*19.79*

Table 17.5 (continued)

	1980–1981/ 1990–1991	1990–1991/ 2000–2001	2000–2001/ 2008–2009	1980–1981/ 2008–2009
Correlation between growth of manufacturing GSDP 1980–1981/ 2008–2009 and initial share of manufacturing GSDP	– 0.208	0.484c	0.601c	0.285

Figure in parentheses against Bihar, Madhya Pradesh and Uttar Pradesh are for the territory after division while those outside include newly formed Jharkhand, Chhattisgarh and Uttarakhand, respectively, in this as well as other tables

Estimates of standard deviation (SD) and coefficient of variation (CV) are based on 14 major states

N.E. not estimated

Latest available data are for the year 2007–2008

[a] Latest data available are for the year 2006–2007

[b] Correlation is significant at 0.01 level

second largest contributor to the national manufacturing GSDP until 1990–1991 but has now given way to Gujarat: The former accounted for 14 % and latter 8 % of national manufacturing GDP in 1980–1981; their shares in 2006–2007 are 11 and 14 %, respectively. West Bengal has been a major loser with a share of 10 % in 1980–1981 and only 7 % in 2006–2007. Other losers are: Andhra Pradesh (from 7.3 to 6.1 %), Madhya Pradesh (from 5.7 to 4.7 %), Assam (from 1.42 to 0.90 %) and Delhi (from 1.95 to 1.87 %). Gainers include Karnataka, Haryana, Goa and Pondicherry. Uttar Pradesh, a significant contributor with about 8 %, has maintained its share. This pattern of changes in the GSDP shares seems to be in line with the changes in investment shares reported in an earlier study covering the immediate pre-reform and post-reform periods (Chakravorty and Lall 2007).

The four most industrialized states, viz. Maharashtra, Tamil Nadu, West Bengal and Gujarat, accounted for 53 % of the total manufacturing GDP of 14 major states of India in 1980–1981; their share is lower at 51 % in 2006–2007. West Bengal continues to be part of this group in 2006–2007, only because Uttar Pradesh has lost part of its territory to Uttarakhand, which otherwise would have had a higher share than that of West Bengal. Among the states with relatively small (1–3 %) contribution to national manufacturing GSDP in 1980–1981, Haryana, Orissa, Punjab and Himachal Pradesh have improved their shares while Kerala has a lower share in 2006–2007 than in 1980–1981. Among other major states, Andhra Pradesh and Madhya Pradesh (even including Chhattisgarh) and Bihar (even including Jharkhand) have lost, while Karnataka and Rajasthan have gained. On the whole, the relative position of different states has not changed much, except a 6 percentage point rise in the share of Gujarat, a 4 percentage point decline in the share of Tamil Nadu and 3 percentage point decline in that of West Bengal. Among the new states, only Chhattisgarh and Jharkhand each have a significant (about 2 %) share of manufacturing GDP of the country and both, especially Jharkhand, have increased their shares since their formation in 2000. Among other states and UTs, only Delhi contributes more than 1 % of manufacturing GSDP and it has maintained its share of around 2 %.

Table 17.6 State-wise distribution of manufacturing GSDP (%) at 1993–1994 prices. (Source: same as Table 17.1)

		1980–1981	1990–1991	2000–2001	2006–2007
Major States					
1	Andhra Pradesh	7.33	6.80	6.14	6.12
2	Bihar (+)	4.17	4.51	2.54 (0.67)	3.62 (0.41)
3	Gujarat	7.98	9.58	11.72	13.70
4	Haryana	2.54	3.40	3.63	3.69
5	Karnataka	5.21	5.38	5.86	6.77
6	Kerala	2.71	2.15	2.32	1.98
7	Madhya Pradesh (+)	5.71	6.31	5.70 (4.15)	4.71 (2.85)
8	Maharashtra	20.51	20.34	19.89	19.70
9	Orissa	1.79	1.55	1.49	2.21
10	Punjab	2.41	3.09	3.46	2.92
11	Rajasthan	3.25	3.47	4.46	3.99
12	Tamil Nadu	14.81	12.12	11.37	10.58
13	Uttar Pradesh (+)	7.38	9.68	8.35 (7.88)	7.39 (6.82)
14	West Bengal	9.70	6.91	7.54	7.02
New States					
15	Chhattisgarh	N.A.	N.A.	1.54	1.86
16	Jharkhand	N.A.	N.A.	1.87	3.21
17	Uttarakhand	N.A.	N.A.	0.47	0.57
North Eastern States					
18	Arunachal Pradesh	0.02	0.02	0.03	0.02
19	Assam	1.42	1.08	0.70	0.90
20	Manipur	0.12	0.14	0.11	0.09
21	Meghalaya	0.02	0.03	0.04	0.09
22	Mizoram	0.002	0.01	0.01	0.01
23	Nagaland	0.01	0.03	0.01	0.01
24	Sikkim	0.00	0.00	0.01	0.01
25	Tripura	0.05	0.04	0.11	0.06
Union Territories and Other States					
26	A&N Islands	0.03	0.02	0.02	0.01
27	Chandigarh	N.E.	N.E.	0.23	0.22
28	Delhi	1.95	2.47	2.14	1.87
29	Dadra and Nagar Haveli	N.E.	N.E.	N.E.	N.E.
30	Daman and Diu	N.E.	N.E.	N.E.	N.E.
31	Lakshadweep	N.E.	N.E.	N.E.	N.E.
32	Pondicherry	0.19	0.21	0.61	0.77
33	Goa	0.55	0.40	0.67	0.69
34	Himachal Pradesh	0.13	0.27	0.60	0.54
35	Jammu & Kashmir	N.E.	N.E.	0.26	0.30
	India	*100.00*	*100.00*	*100.00*	*100.00*
	SD	*5.30*	*4.97*	*4.92*	*4.97*
	CV	*77.67*	*72.97*	*72.96*	*73.65*

Figure in parentheses against Bihar, Madhya Pradesh and Uttar Pradesh are for the territory after division while those outside include newly formed Jharkhand, Chhattisgarh and Uttarakhand, respectively, in this as well as other tables

Estimates of standard deviation (SD) and coefficient of variation (CV) are based on 14 major states

N.A. not available, N.E. not estimated

In terms of employment, however, Uttar Pradesh accounts for the largest share of manufacturing (Table 17.7). In 2004–2005 (the latest year for which data are available), it accounted for 15 % of the manufacturing employment of the country. Tamil Nadu, West Bengal and Maharashtra employed about 11 % each, Andhra Pradesh 8 % and Gujarat 7 % of all manufacturing workers in the country. Karnataka and Madhya Pradesh contributed more than 5 % each. Employment shares of different states have not significantly changed over the years, except some decline in the case of Bihar (even including Jharkhand) and increase in the case of Gujarat. Except the 14 major states and Chhattisgarh, Jharkhand and Delhi, all other 18 states/UTs contributed less than 1 % each of the countrywide manufacturing employment in 2004–2005.

There are large differences between the employment and GSDP shares of individual states. Maharashtra with more than 21 % of GSDP contributed only 11 % of employment among the 14 major states. Uttar Pradesh with 16 % employment has much less, about 8 % share in GSDP, and Gujarat with 14 % SDP had only 7 % share in employment. This is a reflection of large variations in the industrial structure and productivity among states.

6 Conclusions: What Explains Variations?

Description and analysis of various aspects of industrial development in different states presented in the preceding sections, even though not showing any clear pattern, reveal the following interesting trends:

1. As indicated by the share of manufacturing in GSDP, Tamil Nadu, Maharashtra, West Bengal and Gujarat were the most industrialized states in that order in 1980–1981. In 2008–2009, the four most industrialized states were Gujarat, Maharashtra, Tamil Nadu and Haryana, in that order. Gujarat is at the top with 30 % of its GSDP originating from manufacturing. Gujarat has also seen the fastest pace of industrialization, followed by Haryana, Punjab and Himachal Pradesh, while West Bengal, Andhra Pradesh and Tamil Nadu experienced a decline in the share of manufacturing in their respective GSDP. Disparities in the extent of industrialization have somewhat declined during 1981–2009.
2. Most states have experienced a significant shift from agriculture to other sectors; the shift has been the largest in Orissa, Karnataka, Gujarat and Kerala and relatively small in Punjab and West Bengal. A major shift has been in favour of manufacturing particularly in Gujarat, Rajasthan and Orissa. Larger structural changes have generally been accompanied by faster GSDP growth and shift to manufacturing more often than shift to services has contributed to faster growth.
3. Growth rates of manufacturing GSDP have been quite divergent throughout 1981–2009, but especially since 2001. Rates of growth have, however, not necessarily been higher in states with an initially high level of industrialization, except during the period 2001–2009. Thus, industrial growth in recent years has led to increasing divergence.

Table 17.7 State-wise distribution of manufacturing employment Usual Principal and Subsidiary Status (UPSS) (%). (Source: NSS Report on Employment and Unemployment (various rounds))

		1983	1993–1994	1999–2000	2004–2005
Major States					
1	Andhra Pradesh	9.08	8.53	7.73	8.17
2	Bihar (+)	5.71	3.07	5.61	4.65 (2.84)
3	Gujarat	6.42	8.25	5.61	7.25
4	Haryana	1.91	1.71	1.72	2.32
5	Karnataka	6.05	6.13	5.44	4.98
6	Kerala	4.46	3.93	3.94	3.6
7	Madhya Pradesh (+)	5.51	4.36	5.3	5.29 (4.24)
8	Maharashtra	9.85	10.26	10.36	10.5
9	Orissa	3.67	2.94	3.53	3.8
10	Punjab	2.35	1.87	2.32	2.6
11	Rajasthan	3.86	3.07	3.51	4.54
12	Tamil Nadu	12.8	14.86	12.7	11.09
13	Uttar Pradesh (+)	13.26	12.55	15.32	15.80 (15.44)
14	West Bengal	10.87	14.38	12.12	10.74
New States					
15	Chhattisgarh	–	–	–	1.05
16	Jharkhand	–	–	–	1.81
17	Uttarakhand	–	–	–	0.36
North Eastern States					
18	Arunachal Pradesh	0.01	0.02	0.02	0
19	Assam	0.73	0.81	0.92	0.73
20	Manipur	0.13	0.23	0.13	0.16
21	Meghalaya	0.06	0.03	0.02	0.08
22	Mizoram	0.01	0.01	0.01	0.02
23	Nagaland	0.01	0.01	0.01	0.02
24	Sikkim	0.02	0.02	0.01	0.01
25	Tripura	0.14	0.13	0.08	0.12
Union Territories and Other States					
26	A&N Islands	0.01	0.02	0.02	0.01
27	Chandigarh	0.06	0.16	0.13	0.12
28	Delhi	1.87	2.01	2.39	2.02
29	Dadra and Nagar Haveli	0.00	0.02	0.04	0.06
30	Daman and Diu	0.00	0.02	0.04	0.03
31	Lakshadweep	0.00	0.00	0.00	0.00
32	Pondicherry	0.12	0.13	0.20	0.14
33	Goa	0.23	0.10	0.14	0.07
34	Himachal Pradesh	0.23	0.25	0.30	0.37
35	Jammu & Kashmir	0.57	0.12	0.35	0.69
	India	*100*	*100*	*100*	*100*

4. The four states with the largest share in national manufacturing GDP, namely, Maharashtra, Tamil Nadu, West Bengal and Gujarat, have continued to account for more than half the national gross value added (GVA) in manufacturing, Maharashtra remaining at the top, Gujarat replacing Tamil Nadu in the second position and West Bengal receding from the third to fourth position. The overall disparity in the shares of different states has slightly declined in 2007–2008 from 1980–1981.

5. In employment terms, Uttar Pradesh replaces Gujarat among the top four states, which account for 48 % in 2004–2005; Uttar Pradesh alone accounts for 16 % of employment, the other three, namely, Maharashtra, Tamil Nadu and West Bengal, 11 % each. There has been little change in the employment shares of different states.

As revealed by the findings as noted above, it is quite clear that states have performed differently from each other in terms of growth of manufacturing industries. What factors account for such differential performance? It may not be difficult and may even not be very useful to try to explain the differences in the levels of industrial development that have historically existed. What may be more interesting and also useful is to attempt an explanation of the changes that have taken place in the period of the past two to three decades, especially since the introduction of economic reforms which removed government regulations on investment and industrial location which, on the one hand, gave freedom and opportunity to states to base their industrial development on specialisation, and on the other, did away with the central government's use of its control and instrumentality to influence investment and industrial location in favour of industrially less-advanced states and regions.

Various factors that could have influenced the differential performance of states in industrial growth during the post-reform period can broadly be divided into the following four broad heads: capital investment, human resources, regulatory framework and infrastructure. A study (Chakravorty and Lall 2007, p. 99–102) looking at the trends in industrial investment of different states over a 7-year period immediately following the introduction of the economic reforms in 1991 found that the process of cumulative causation was in operation insofar as the existing level of industrial investment and activity attracted the new investment. Continuity and clustering were, thus, found to lead to increasing divergence. This observation is supported by findings of our study especially for the period 2001–2009.

That, however, does not mean that other factors may have had no influence on the growth of industrial activity in different states particularly if there was differential progress, in respect of them among states. Let us look at changes in human resource development and regulatory and promotional framework and see if there have been significant differences in terms of changes in them. Going by the Human Development Index (HDI) as the summary indicator of development of human resources, there is a general trend towards an improvement: HDI for country as a whole was estimated to be 0.387 in 1999–2000 and is found to have improved to 0.467 in 2007–2008 (IAMR 2011, p. 24). Similar improvements have taken place in all the states, so much so that eight states have retained the same ranking in 2007–2008, as in 1999–2000, 11 states have changed ranks but only by one or two positions. Only Rajasthan has lost by three positions and Jharkhand and the Northeast (excluding Assam) have gained by four and three positions, respectively. Similarly, there has been a general trend towards easing of regulations and promotion of investment-friendly climate in all the states. Various exercises by the World Bank and industry organisations have attempted measurement of the ease and difficulty of 'Doing Business' in different states and have found significant differences among states. It is, however, not clear whether the degree of 'ease' has changed at different speeds in the post-reform period. In general, states have competed among themselves in projecting an investment-friendly image and it appears that it has been a zero-sum game rather than any advantage of one state over the others. Gujarat and Maharashtra have, no doubt, offered the 'best' and Uttar Pradesh and West Bengal 'poor' investment climate (World Bank 2004). But that is true of both the pre- and post-reform periods. In fact, some other states like Andhra Pradesh and Karnataka have improved their investment-friendly image. Karnataka has also experienced faster industrial growth, but Andhra Pradesh has not.

One aspect of regulatory framework that has been studied most is labour regulation. A number of studies (e.g. Besley and Burgess 2004; Hasan et al. 2003, Goldar 2011) conclude that states with 'flexible' labour regions, especially those having amended laws and rules to give greater freedom to employers in modes of use of labour, have performed better with respect to industrial growth than others. Several other studies, however, argue that most of these studies are methodologically faulted insofar as they are often based on single legislation and changes in it or on answer to a leading question of impact of labour laws to the complete neglect of other factors such as infrastructure, market, credit, etc. (Bhattacharjea 2006, Reddy 2008, Nagaraj 2011). It appears that better industrial relations climate, no doubt, helped some states (e.g. Gujarat, Andhra Pradesh and Karnataka) to perform better, but the significance of this factor was far overshadowed by other factors, particularly infrastructure. There is, however, no doubt that the labour market and industrial relations regulation were a part of the overall governance and regulatory system which, as a whole, was an important factor in encouraging or stifling industrial growth.

Infrastructure is most widely accepted as the reason for the differential status and growth of manufacturing industry among the states. Analysis has often been attempted to explain such a difference in terms of a single infrastructure item such as banking facilities (Burgess and Pande 2003) and power (Adil 2010). Some other studies have taken several items of infrastructure as independent variables to explain variations in some indicator (e.g. total factor productivity (TFP) in Mitra et al, 2002) of industrial performance and found some of them more important than others. For example, the study mentioned above found investment in primary education, financial mobilisation as reflected in deposits and credit disbursal and power production capacity as the factors significantly influencing industrial productivity. Paul (2011) looked at the impact of banking outreach, physical infrastructure and labour market flexibility on the growth of manufacturing industries across 14 major states of India in the post-liberalisation period (1991–1992/2002–2003) and found that while the first two influenced industrial growth significantly the last has no significant impact.

Often infrastructure items, including physical, economic and social items (like road length and railway length per unit of geographical area, energy consumption, educational facilities, hospitals, banking facilities, post and telecommunications), have been clubbed together to construct an overall 'infrastructure index'. Utilising one such index (constructed by the Centre for Monitoring Indian Economy, CMIE) to examine the relationship between infrastructure and the extent of industrialization (share of manufacturing in the GSDP), it is observed that there is a fairly significant relation between the two. The rank correlation coefficient between the two was 0.36 for the year 1980–1981. It was stronger in 1990–1991 at 0.42 but grew weaker at 0.33 in 2000–2001 (Table 17.8). Yet, it was statistically significant in all three years.

Table 17.8 Infrastructure and level of industrialization. (Source: CMIE and ASI)

States	1980–1981		1990–1991		2000–2001	
	Rank		Rank		Rank	
	Infrastructure Development Index	% share of manufacturing in GSDP	Infrastructure Development Index	% share of manufacturing in GSDP	Infrastructure Development Index	% share of manufacturing in GSDP
Andhra Pradesh	8	6	8	8	12	12
Assam	15	11	13	15	11	16
Bihar	12	10	15	11	17	15
Gujarat	5	4	5	2	6	1
Haryana	4	7	4	4	5	4
Himachal Pradesh	13	16	10	16	10	10
Jammu and Kashmir	11	17	14	17	16	17
Karnataka	10	5	9	5	9	6
Kerala	3	12	2	14	3	14
Madhya Pradesh	17	9	17	7	20	7
Maharashtra	6	2	6	3	8	3
Orissa	14	14	12	13	14	13
Punjab	1	13	1	10	1	9
Rajasthan	16	8	16	12	19	8
Tamil Nadu	2	1	3	1	4	2
Uttar Pradesh	9	15	7	9	7	11
West Bengal	7	3	11	6	13	5
Rank correlation	*0.36*		*0.42*		*0.33*	

Appendix

Table 17.9 Transport and power infrastructure and level of industrialization: regression results. (Source: taken from Papola et al. 2011)

Independent variable/time period	Constant	Coefficient	t-value	p-value	R-square
Dependent variable: % share of manufacturing gsdp to total gsdp					
Railways length_1981	9.696	0.171	1.0200	0.3300	0.0690
Railways length_1991	13.264	0.117	0.7800	0.4500	0.0410
Railways length_2001	12.727	0.157	1.1000	0.2900	0.0750
Road length_1981	14.007	− 0.0003	− 0.0800	0.9360	0.0005
Road length_1991	17.282	− 0.002	− 0.4600	0.6520	0.0149
Road length_2001	16.883	− 0.001	− 0.4500	0.6570	0.0135
Power consumption_1981	7.251	0.044	2.0200	0.0630	0.2258
Power consumption_1991	10.691	0.021	2.0000	0.0660	0.2219
Power consumption_2001	8.251	0.019	3.7700	0.0020	0.4865
Power consumption_2004	9.913	0.015	3.4300	0.0040	0.4399
Dependent variable: per capita manufacturing GSDP					
Railways length_1981	401.280	18.930	1.6900	0.1120	0.1600
Railways length_1991	967.310	16.890	0.8500	0.4080	0.0490
Railways length_2001	1,297.850	27.020	1.0200	0.3230	0.0650
Road length_1981	401.280	18.930	1.6900	0.1120	0.1600
Road length_1991	1,492.590	− 0.140	− 0.3300	0.7480	0.0080
Road length_2001	2,055.700	− 0.120	− 0.3100	0.7620	0.0060
Power consumption_1981	401.280	18.930	1.6900	0.1120	0.1600
Power consumption_1991	275.280	4.260	3.9100	0.0020	0.5220
Power consumption_2001	109.550	4.560	5.4500	0.0000	0.6640
Power consumption_2004	80.470	5.080	5.5200	0.0000	0.6700

Composite indicators are good for summary description but not for identifying the relative importance of different infrastructure items. In most studies, transport and power have been identified as the most critical elements of infrastructure influencing the pace of industrial growth in a region or state. We, therefore, attempted an analysis to explain interstate variations in the level of industrialization and growth of manufacturing GSDP focusing on railways and road length per square kilometre of area as an indicator of transport infrastructure and electricity consumption per capita as an indicator of availability of power. Taking share of manufacturing in GSDP as the indicators of levels of industrialization of a state, we found that it was only the power consumption which had a positive and significant relationship with it, in all the three time points, 1981, 1991 and 2001 for which regression analysis was undertaken. The length of railway line had positive but not significant coefficients. Road length surprisingly came up with a negative coefficient in all the 3 years. Similar results were obtained when the indicator of the level of industrialization was changed to *per capita manufacturing GSDP,* except that the explanatory power of the model improved as also the value of the coefficient of power consumption and the coefficient of road length turned out to be positive in one case, that is, in 1981. (Appendix, Table 17.9). Our attempts to establish dynamic relationships between these items of infrastructure and growth of manufacturing industry in different states by estimating regression of base year infrastructure with growth over the next decade or to relate growth in infrastructure with growth in manufacturing GSDP over each of the three periods, however, yielded no significant results.

Outcomes of our statistical exercises, however, do not imply that various items of infrastructure do not influence the pace of industrial development in different states. There could be several reasons for the relationship not showing up significantly. *One,* the specification of the variables may not be the most appropriate. *Two,* the quality of data may vary among states. *Three,* some items may not have significantly large variations across states as over the years a larger degree of convergence has emerged with respect to items like facilities for human development, banking, transport and communications among the states. *Four,* where variations are significant, the relationship is also significant. Power availability is one example which is probably a good proxy for all items of infrastructure directly relevant for industry, and it could overshadow the influence of other items. *Five,* after the initial phase of industrialization, infrastructure may continue to be important but its influence is intermixed with that of agglomeration economies. In other words, new industries go where industries exist which are also the states that have better-developed infrastructure. Between states with developed infrastructure but very little industry and those with both developed infrastructure and a good industrial base, the latter attracts more industry than the former. Thus, Kerala with good infrastructure does not attract industry while Gujarat also with high level of industrialization does. Punjab with highly developed infrastructure has a relatively lower level of industrialization, but Maharashtra with relatively lower level of infrastructure development has a high level of industrialization. It appears that the pattern of location of new industrial activity is becoming increasingly complex and requires fresh approaches that go beyond the traditional theory of industrial location to explain it.

References

Adil M (2010) Manufacturing sector productivity in India: all India trends, pattern and network externalities on regional growth. Ph. D. Thesis, University of Maryland, College Park, (Unpublished)

Ahluwalia MS (2000) Economic performance of states in post-reforms period. Economic and political weekly, 6 May 2000

Ahluwalia MS (2011) Prospects and policy changes in the twelfth plan. Economic and Political Weekly, 21 May 2011

Besley T, Bargess R (2004) Can labour regulations hinder economic performance? Evidence from India. Q J Econ 119(1):91–134

Bhattacharjea A (2006) Labour market regulations and industrial performance: a critical review of empirical evidences. Indian Journal of Labour Economics 49(2)211–232

Burgesss R, Pande R (2003) Do rural banks matter? Evidence from the Indian social banking experiment. Development economics research papers, London, London School of Economics. Am Econ Rev 95(3):780–795

Chakravorty S, Lall S (2007) Made in India: the economic geography and political economy of industrialization. Oxford University Press, New Delhi

Goldar B (2011) Growth in organised manufacturing employment in recent years. Economic and political weekly, 12 February 2011

Hasan R, Mitra D, Ramaswamy KV (2003) Trade reforms, labour regulations and labour demand elasticities, empirical evidence from India. Working paper no. 9879, National Bureau of Economic Research, vol 89. MIT, Cambridge, pp 466–481

IAMR (2011) India. Human Development Report 2011. Institute of Applied Manpower Research and Oxford University Press, New Delhi

Mitra A, Varoudakis S, Veganzones-Varoundakis MA (2002) Productivity and technical efficiency in Indian states: the role of infrastructure. Econ Dev Cult Change 501(2)95–426

Nagaraj R (2011) Growth in organised manufacturing employment: a comment. Economic and political weekly, vol 46, No. 12, 17–24 March 2011

Papola TS, Maurya N, Jena N (2011) Inter-regional disparities in industrial growth and structure. ISID working paper series, 2011/04. Institute for Studies in Industrial Development, New Delhi

Paul R (2011) The relative impact of banking infrastructure and labour on industrial growth: evidence from Indian states. Macroeconomics and Finance in Emerging Market Economies 4(1):101–124

Reddy DN (2008) Labour regulation in Indian industry: a case study of Andhra Pradesh. Bookwell and ISID, New Delhi

World Bank (2004) India: investment climate and manufacturing industry. Finance and private sector development unit, South Asia region. World Bank, New Delhi

Chapter 18
Unit Root and Structural Break: Experience from the Indian Service Sector

Purba Roy Choudhury

1 Introduction

Generally, the survey of literature on India's growth trajectory has been preoccupied with the issue of turning points in growth rates rather than structural changes accompanying them. A parallel discourse, however, exists on the phenomenon of services rather than industry accounting for an extraordinary large share of the expansion of non-agricultural output in India. However, two major turning points of growth rates have been referred to and debated by various economists over time. The first one is associated with Independence and the transition from the colonial era to the 'Hindu rate of growth'. The slow Indian growth rate is better attributable to the Government of India's protectionist and interventionist policies, and it is from this perspective, we often describe modern economic growth in India. The second turning point is 1980, after which the Indian economy appears to have moved to a higher trend growth of 5.5–6 % per annum. We can also add that the early years of this century brought with it the third turning point with growth rates ranging from 8 to 9 %. In this chapter, we focus on these turning points of growth, suggesting that these growth patterns were different resulting from the pattern of structural change in output in these periods.

When the role of services in Indian growth became quite huge, it was described as 'disproportionality' or 'excess growth of services'. Later, the term coined was 'services revolution'. The phenomenon provoked a lot of debate regarding the determinants, explaining them and their long-term sustainability. All these resulted to the question: 'Is India pioneering a new pattern of growth where services can play the role "engine of growth," just like the same role played by industry for other countries in the past?' There are other measurement issues relating to this growth as to how much is real and how much of it is just statistical. With all these controversies and preferred points of view, there is little doubt that this exceptional growth of services

P. Roy Choudhury (✉)
Department of Economics, The Bhawanipur Education Society College,
Kolkata 700 020, India
e-mail: purbarc@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_18, © Springer India 2014

makes the structural change in India different and unique, an exception to the universal rule. Two features bring about this uniqueness in the services growth in India. The first is the premature nature of the transition to a services-dominated economy at an exceptionally low level of per capita income and without attaining a higher level of industrialization. The second one is that a corresponding share of employment has not matched the large share of services in output.

This chapter tries to analyse the trends in the service sector growth in India from 1950 to 2010. This chapter is organized in the following way. Section 1 undertakes a selective survey of literature on the growth of services and employment and its role in the process of economic development with reference to India. Section 2 tries to analyse the relative share of agriculture, industry and services in the gross domestic product (GDP) of the Indian economy as a whole, along with a decomposition of the subsectors of the service sector in India. Section 3 also takes into account the econometric methodology of the unit root properties of time series data trying to define the different tests of unit root with or without break. The breakpoints are estimated with the help of the Zivot–Andrews test. Section 4 also takes into account the empirical results and their implications. Section 5 concludes the study.

2 Conceptualization of Services and Economic Growth: A Survey of Literature

There exists a long inconclusive controversy regarding the conceptualization and role of services in the development process of a country. The debate ranges back to the classical economists from Adam Smith to Karl Marx, who were interested in services as distinct from goods for the purpose of defining productive labour. Adam Smith was of the view that the services 'perish in the very instant of their performance and seldom leave any trace of value behind them'. The classical writers discussed services from a perspective of a distinction between productive and unproductive labour. Thus, according to classical economists, services are considered as unproductive activity. Until recently, services used to be treated accordingly, by socialist economies as 'non-material production' as against the productive 'material production' of goods in their national income accounting statistics. But with the notion of services being unproductive, it has also been emphasized by development economists like Fischer (1935), Clark (1940), Kuznets (1955) and Fuchs (1980) that with the growth of the service sector, structural changes take place in the economy.

The earliest attempt to define services was made by Hill (1977) who argues that 'goods and services belong to different logical categories'. He focused on the fact that producers cannot accumulate a stock or inventory of services, stressing that services must be consumed as they are produced and cannot be stored. Thus, Hill (1977) emphasized the non-storability of services, which requires that the services must be consumed the moment they are produced. Thus, it makes it essential for the user and the provider of the service to interact so that there is no time gap in production and consumption of the service. But this conceptualization of service by Hill (1977) was put to criticism by many other economists. Melvin (1995) points out that Hill's definition only relates to contact services and there exists a range of services which do not permit a separation of the location of production and consumption in space and time, so that services trade may take place at the factor or the product level.

Griliches (1992) defines services as anything that is a result of labour that does not produce a tangible commodity. Lack of tangibility is fundamentally what leads to non-storability and to non-transferability. But this definition was also scrutinized based on the fact that software programs and digital electronic content have limited tangibility but are storable and transferable.

The marketing literature pays considerable attention to the nature of services, and in this literature, the dimensions that distinguish services from products are as follows: (1) intangibility, (2) heterogeneity, (3) simultaneity in production and consumption and (4) perishability. For example, in this definition scheme, intangibility, perishability and simultaneity all work against piling of inventories. Heterogeneity and simultaneity (where the consumer is involved in creating the service) make it more difficult to achieve scale economies. All of these features make display and consumer assessment of characteristics harder to achieve.

The use of the earlier mentioned characteristics to distinguish services from goods is subject to several outcomes. First, the differences are mostly of degree, especially in comparing services to certain kinds of products. Second and more fundamentally, the distinction between product and services is often based on the contractual nature of the market transaction. More fundamentally, every product either produces services directly or produces services in combination with other products to generate services. Thus, an automobile is purchased as a product, but the value is based on the stream of services that it will provide over time. Thus, all durable goods may be leased or rented, in which case, the service is more explicit than in the case of product transaction. The more expensive the product is (e.g. airplane vs automobile), the less possible or economically viable it is for an individual to access the service through a purchase. In fact, contracting from services from specialized producers is often a way to achieve economies of scale in producing services that would not be possible otherwise. The general point is that the boundary between product and services is often a function of market and economic conditions, rather than the intrinsic properties of what is being exchanged.

Besides the marketing literature, one area where the conceptualization of distinction between products and services has been explored in detail is that of international trade, where categorizations are important for shaping international trade policies and agreements. Locational characteristics are central to discussion of services in international trade. Bhagwati (1984) argues that services can be divided into two categories: (1) first, those that necessarily require the proximity of the user and the provider and (2) second, those that do not essentially require this but may be useful. Services that require that essential proximity have been further categorized into three groups that are:

a. Mobile provider and immobile user (e.g. shifting labour to the construction site in other country)
b. Mobile user and immobile provider (e.g. hospital services)
c. Mobile user and mobile providers (e.g. lectures, legal advice, haircuts, etc.)

However, it has been argued that the services for which physical proximity is inessential, i.e. the long distance services, are on the rise due to technical progress that makes it possible to provide services without physical proximity (Bhagwati 1984). Services like banking and insurance fall under this category. But it has also been argued that physical proximity between the producer and user of service does lead to greater efficiency. However factor mobility and trade in services are two integral aspects of service transaction (Bhagwati 1984).

But Stern and Hoekman (1988) point out that services can be (1) complementary to trade in goods; (2) a substitute of trade in goods; and (3) unrelated to goods. All these characteristics have implications on how trade can occur. Their intangibility and non-storability imply that in order to become tradable, services have to be applied to, or embodied in objects, or information flow or persons. This makes international transaction in services more complex conceptually than international transaction in goods.

Most of the economists have categorized international transaction in services into three groups:

a. Cross border or separated trade analogous to trade in goods
b. Transaction that requires the movement of the producer to the location of the demander (demander located services)
c. Transactions that imply the movement of the demander to the location of the provider (provider located services; Sampson and Snape 1985, Stern and Hoekman 1987).

The earlier mentioned classification provided by Sampson and Snape (1985) was modified by Sapir and Winter (1994) for the purpose of classifying the international transaction in services. This most commonly used classification is based on the constraints on the physical location of producer and consumer in realizing the transaction:

1. Services transactions without movement of both the receiver and the provider of service
2. Services transactions for which the consumer travels across borders to the immobile producer
3. Service transactions that are accomplished by the temporary movement of factors of production across national borders, while receiver of the services does not move
4. Service transaction by means of permanent local establishment via a foreign affiliate of a firm originating from a different country

This classification has been adopted by the World Trade Organization (WTO) established April 1994, under the General Agreement of Trade in Services (GATS). This agreement applies to four modes of supply:

Mode 1: cross-border supply of service (not requiring physical movement of the supplier and the customer)

Mode 2: provision implying the movement of the customer to the location of the supplier

Mode 3: services sold in the territory of a member by (legal) entities that have established a presence there but originate in the territory of another Member

Mode 4: provision of services requiring temporary movement of national persons

Apart from the locational classification of services, to facilitate international transactions, services have been defined and categorized using different characteristics and uses of services. Again, to arrive at a functional definition of services, one has to know the reasons behind the demand for services, for instance. Service is acquired to serve the following goals:

a. Instant benefits (e.g. travel, entertainment, haircuts, etc.)
b. Enhancement of user's consumption benefit capacity by reducing the cost–benefit ratio per product transaction (e.g. transport, communication, financial services, insurance, etc.)
c. Enhancement of user's productive capacity by reducing the cost–benefit ratio per unit of output (e.g. transport, training, business services, medical services, etc.)

Based on this classification of services, alternative groupings of services have been done, these are:

1. 'Intermediate' Versus 'Final Demand' Services: This classification scheme comes from the input–output matrix structure, where intermediate production refers to output sold to other domestic firms or agencies.
2. 'Producers' Versus 'Consumers' Services: This classification scheme is based on the end consumer and was originally developed to underscore that at least half of all service sector production is sold to other firms (i.e. producer services).
3. 'Market' Versus 'Non-Market' Services: This classification scheme differentiates between services paid for directly by a customer (industry or private households) and those paid indirectly through taxes.

Alternatively, many studies adopt a broader and simpler definition that helps in distinguishing services from products. One such broad definition of services is: 'services form a diverse group of economic activities not directly associated with the manufacture of goods, mining and agriculture. These typically involve in provision of human value added in the form of labour, managerial skill, entertainment, training and the like'.

Thus, studies have put forward alternative definitions and classification schemes. However, the basic characteristics of services on which most of the classifications are based are: non-transferability and non-storability. Other associated characteristics of services that need to be noted as services are heterogeneous and flexible in production, and imperfect competition is highly relevant for services. This implies that consumer's preference for services can be easily met, also because of simultaneity in production and absorption; services can be regarded as heterogeneous products.

Conceptualization of services therefore depends on the nature and purpose of study. It is found that categorization of services broadly used in the current literature on trade in services primarily follows GATS classification scheme, while the literature related to the role in services in economic growth and productivity in services follows use-based categorization of services. Alternatively, United Nations has also developed the Central Product Classification (CPC) for identification of services. The CPC is now used as a reference for identification of services under GATS and also to describe service components in the balance of payments as recommended in the Balance of Payments Manual of International Monetary Fund (IMF) (1993).

From an analytical point of view, the tertiary sector that contributes productively to the overall economic growth has been criticized by numerous economists but there were several others who have emphasized its importance, particularly keeping in view the recent trends in the tertiary sector's growth and the expansion specially in the post-globalization period in India. A recent study done by the World Bank (2004) by Gordon and Gupta supports that the enormous growth in India's GDP, in the past 10 years, has been largely substantiated by the growth of the service sector. The studies were pursued in the Indian context started with Bhattacharya and Mitra 1989, 1990, 1991 and 1997, and Datta 1989. Some of the subsectors within the tertiary sector, which are crucial for the growth of industry and the rest of the economy, like transport, storage and communication and financial and business services, have been expanding during this period. Hence, the question arises whether services can play the role of an engine of growth?

In the process of economic growth, Kaldor (1967) suggested that it is the manufacturing sector, which plays the role of engine of growth, as the potential for productivity growth is highest in this sector. He, in fact, provided the theoretical rationale for the patterns of structural change that Kuznets (1955) had observed in the case of advanced countries during the process of their economic development (Dasgupta and Singh 2006). Kuznets (1966) also suggested on the basis of the empirical evidence of the developed countries that the tertiary sector expands in relative terms only after the secondary sector has already acquired dominance in terms of both value added and work force in the process of rapid industrialization. When the relative size of industry predominates that of the other sectors, the tertiary sector then acquires significance in value added and work force composition. This is because of the fact that after a considerable rise in per capita income originating from the commodity-producing sector, the demand for services increases as the consumption demand for commodities gets saturated. But in the context of the developing countries, the phenomenon of a relatively large tertiary sector could be evident even much before the secondary sector could acquire a reasonable size of at least one-third in terms of value added or work force (Gemmell 1986).

Though it is comparatively easy to analyse the shift in favour of the tertiary sector in the context of the developed countries as a standard transition of development theory (because following the rapid progress in industrialization the demand for several services grows faster, which in turn reduces the share of the secondary sector in the total product of the economy), but in the case of the developing countries, the dominance of the tertiary sector before the secondary sector's relative size could outweigh that of other sectors did give rise to several concerns regarding this phenomena. Bhattacharya and Mitra (1989) stated that higher the discrepancy between the industry and agriculture growth is, the higher the growth of services across Indian states is, implying that higher levels of per capita income originating from industrialization lead to higher demand for services. In a later work, Bhattacharya and Mitra (1990) argued that a wide disparity arising between the growth of income from services and commodity-producing sector tends to result in inflation. This is particularly so if the tertiary sector value added expands because of rising income of those who are already employed and not due to income accruing to the new additions to the tertiary sector work force. In other words, if expansion in both value added and employment generation takes place simultaneously within the tertiary sector, there will be a commensurate increase in demand for food and other essential goods produced in the manufacturing sector. However, if the expansion of the tertiary sector results only from the rise in income of those who are already employed in this sector, the additional income would create demand for luxury goods and other imported goods since the demand for food and other essential items has already been met (Bhattacharya and Mitra 1989, 1990).

Again, factors like increasing role of the government in implementing the objectives of growth and employment generation, expansion of defence and public administration, the historical role of the urban middle class in wholesale trade and distribution and demonstration effects in developing countries creating demand patterns similar to those of high-income countries have been highlighted to offer a rationale for the expansion of the tertiary sector (Panchamukhi et. al. 1986). Sub-sectors like transport, communication and banking do contribute significantly to the overall economic growth as they constitute the basic physical and financial infrastructure. Especially the role of information technology (IT) and business process outsourcing services (BPOS) in enhancing the economic growth is said to be significant (World Bank 2004). In addition, the new growth theorists indicate that skill-intensive activities exert positive externalities on the rest of the economy, thereby raising productivity and growth (Romer 1990).

As far as service sector employment is concerned, a huge amount of workforce in developing countries engaged in this sector is the result of a lack of employment opportunities in manufacturing sector, resulting from the adoption of labour-saving technological change, factor market imperfections and rapid increases in the labour force (Meier 1970). It is also occasionally argued that only a small proportion of tertiary employment in the developing countries is a function of the income elasticity of demand for services, and a large majority of it is believed to be a manifestation of excess supplies of labour relative to demand. Udall (1976) pointed out that the demand for service employment is usually taken in the literature to be relatively elastic (with respect to price). He lends support to Kuznets (1966) in suggesting a shift in the demand for labour in the long run towards the tertiary sector. Kuznets (1966) urged that the share of tertiary employment in the labour force increases mainly because of slow growth of technical progress in services, a high-income elasticity of demand for some of the tertiary activities and increasing urbanisation, resulting in the rise in demand for services like transport and distribution. Galenson (1963) viewed that an increase in manufacturing activity leads to a rise in tertiary employment as income growth originating from the expansion of manufacturing activity raises the consumption of services and also the demand for service inputs into manufacturing. Therefore, the growth of employment in service activities is viewed either as purely a supply-push phenomenon or as rationalising its growth in terms of ‘demand induced’ hypothesis.

It is quite evident from the conceptualization of services literature that the tertiary sector comprises highly heterogeneous jobs, which respond differentially to demand and supply factors. Greenfield (1966), for example, by dividing the services into consumer and producer categories, noted that producer services grow as industrial corporations in order to reduce their costs and use the knowledge of the experts to shift some of the tasks previously performed by them to the producer service firms. Thus, in a growing economy, with increasing specialisation and capital accumulation, the demand for producer services is expected to rise. Similarly, with a shift to a predominantly service economy, the service organisations in various countries have become large users of IT, and this has given rise to a large demand for service functions allied to the operation of the computer hardware (Elfring 1989).

On the other hand, rising female labour participation rate is expected to have a positive effect on tertiary sector employment since women workers prefer tertiary-type employment, or in other words, this sector is more conducive to absorbing more female labour entering the job market. In particular, as Fuchs (1980) and Grubel (1987) argued, with rising female labour force participation rate, the demand for personal services grows, since employed women spend a higher proportion of their income on services which they themselves would have rendered within the household had they not been employed. Similarly, with certain demographic changes, like population ageing, the purchase of certain personal services shows an increasing tendency (Silver 1987). All this tends to suggest that different components of the tertiary sector draw their growth stimuli from different sets of factors, and it would be quite inappropriate to merge all the components in one single category. Realising the importance of this very fact, Elfring (1989) studied, in detail, the service sector employment in seven Organisation for Economic Co-operation and Development (OECD) countries under four broad categories: (a) the producer services, (b) the distributive services, (c) the personal services and the (d) social services. Bhattacharya and Mitra (1997) also classified the services sector into four categories—bureaucratic services, distributive services, consumer services and producer services. Based on cross-country analysis, their findings suggest that the impact of per capita income on the percentage share of tertiary sector in total work force is positive, though it tends to stabilise at higher stages of development. Banga and Goldar (2004), in the Indian context, noted that the importance of services as an input to production in the manufacturing sector increased considerably in the 1990s compared to the 1980s. As the authors pointed out, real value of services used in manufacturing grew at the rate of 0.4 % per annum in the 1980s, and the growth rate increased sharply to around 16 % per annum in the 1990s. Economic policy changes in the 1990s, particularly the trade reforms, created a condition favourable for increased use of services in manufacturing.

The work by Saith (2006) begins with a detailed background to the tertiary sector's growth and its varying interpretation by researchers. Historically, with the expansion of markets, there took place integration of economies of scale, and technological change created necessary conditions for the emergence of service-related activities and occupations within the manufacturing enterprises. The process of vertical integration facilitated the integration of different service-related activities within the manufacturing enterprises. They consisted of large armies of technocrats and administrators, who were required to co-ordinate activities, provide specialized services and manage the enterprises. Starting from this point, the work by Saith (2006) goes on to the relatively recent developments in this field, that is, how human capital formation and high quality labour in the tertiary sector can contribute to output growth in the economy at a faster rate by enhancing total factor productivity growth and technical efficiency.

After the 1980s, following the stagflation experienced by most of the market economies of the developed world, some of the occupations shifted to new independent standing organizations. Similarly, new industries, including IT, ushered in new dimensions to the growth process of several activities within the services sector even in developing countries, and as a result, the contribution of the services sector to the overall growth of the economy is expected to be commendable. Sarkar and Mehta (2006) point out that the growth of the *information and communications technology* (ICT) sector has led to the emergence of a 'New Economy' in India, which has been a generator of new jobs for technical persons and has been helping to earn foreign exchange through exports and attracting foreign investment. Though it is predominantly an urban activity in rural areas, the telecommunication segment of ICT does provide employment. Further, the wage level of the ICT workers is higher than non-ICT workers.

The other area, which is also of great importance, is trade in services, and this is expected to raise the rate of growth of the economy. Historically, there was an illusion among economists regarding non-tradability of services, though presently services are being traded globally in different modes. Two sets of reasons can be identified in suggesting that at least some of the components of services, which were earlier non-tradable, are becoming tradable. They are: (a) technological: computing power and the Internet and (b) liberalization of trade and investment policies that allow service providers to cross national boundaries. The measurement of barriers to trade in services began with frequency-based measures and scholars assigned exogenously determined weights to different types of barriers. But these weights now can be estimated endogenously with the help of partial and general equilibrium models (Saith 2006). Bhattacharya and Mitra (1997), based on their cross-country analysis, noted that trade openness did not have any significant effect on the relative size of the tertiary sector in total employment, though at the disaggregate level of the tertiary sector it showed some positive effect. Presumably, increasing international trade shifted workers from activities with a high incidence of low-productivity component to more specialized commercial activities.

3 Macroeconomic View of India's Service Sector

Before going into the details of analysis in this chapter, it becomes pertinent to state that the data sources used are based solely from National Accounts Statistics (NAS) of Central Statistical Organization (CSO), 2004–2005 base year series, NAS 2011, NAS 2008 and 2009 and the NAS 2004–2005 base year back series, between the entire period from 1950 to 1951 and 2009–2010. We have also taken into account the data given by Economic and Political Weekly (EPW) Research Foundation (2004).

Another important problem that comes before going into a formidable analysis is that the analysis of structural change in output is based on the division of economy in to agriculture–industry–services or primary–secondary–tertiary sectors. The demarcation of the industrial sector from the services sector has again led to certain debates. Thus, while Kuznets used transport and communication in industry, Clark put even construction in services. The general practice, however, is to include construction in industry, along with mining and quarrying and manufacturing and all other non-agricultural activities, including transport and communication in services. The choice of classification scheme is important because it can affect the conclusions one draws about the pattern of structural change accompanying growth. However, without going into much of debate, we go by the definition drawn by the CSO as a standard practice.

3.1 Macroeconomic View

A striking feature of India's growth performance over the past decade has been the strength of the service sector. The preponderance of services over industry is not a recent phenomenon for the Indian economy, but has been in place since the beginning of the 1950s. With the decline of the primary sector (i.e. mostly the agricultural sector), keeping in mind the conventional wisdom of development, the predominance of the services ahead of industry stands as a departure from development theory.

An overall macroeconomic view of India's GDP at the broad level reveals that of the share of agriculture, industry and services as a share of total GDP, the share of the service sector has shown a considerable and persistent increase in India since independence.

(1) Share of services in GDP at factor cost and the annual growth of GDP and Services at factor cost at 2004–2005 prices in India. An overall pictorial view of the broad sector of agriculture, industry and services and GDP from 1950 to 2010 shows that the rise in GDP especially after the economic reforms in 1991 is brought about by the service sector.

In Fig. 18.1, the vertical axis measures agriculture, industry and services in GDP at factor cost at 2004–2005 prices in India in Rs. Crores. The line diagram reveals that there has been significant increase in services than agriculture and industry from 1950–1951 to 2009–2010, leading to the massive increase of the GDP at factor cost over time.

The analysis of the sectoral composition of GDP for the period 1950–2010 brings out the fact that 'tertiarization' of the structure of production in India has taken place. During the process of growth over the years from 1950–1951 to 2009–2010, the Indian economy has experienced a change in production structure with a shift away

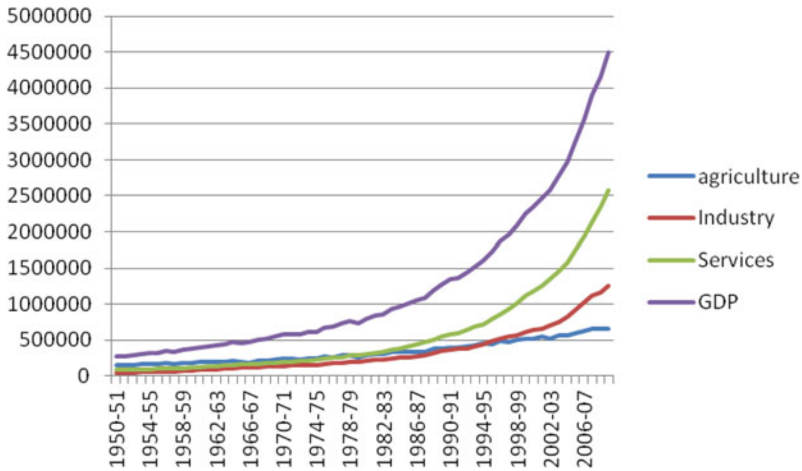


Fig. 18.1 Contribution of agriculture, industry and services to gross domestic product (GDP) at factor cost at 2004–2005 prices. (Source: RBI, Handbook of Statistics; CSO, EPW Foundation (2004), Own calculations)

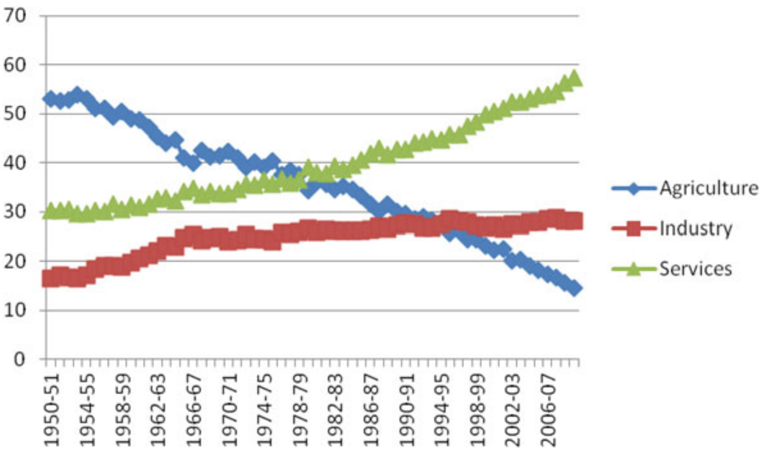


Fig. 18.2 Contribution of percentage share of agriculture, industry and services as a proportion of gross domestic product (GDP) at factor cost at 2004–2005 prices. (Source: RBI Handbook of Statistics; CSO, EPW Foundation (2004), Own calculations)

from agriculture towards industry and tertiary sector. On the other hand, Fig. 18.2 gives a clear representation of the percentage share of the contribution of agriculture, industry and services as a proportion to real GDP at 2004–2005. Agriculture has declined drastically over the years (around 55 % in 1950–1951 to 15 % in 2009–2010); industry has risen but not substantially (from around 15 % in 1950–1951 to 27 % in 2009–2010), while services contributed enormously during this period (29 % in 1950–1951 around 58 % in 2009–2010). During the 1950's, it was the primary sector which was the dominant sector of the economy and accounted for the largest share in GDP. But the whole scenario changed subsequently, and especially in the 1980's, the tertiary sector emerged as the major sector in the economy in terms of production share in the 1990s and thereafter. To analyse what are the reasons behind the enormous increases in the amount of services, it becomes necessary to identify the performance of each individual subsector of service and their contribution towards this service revolution.

When the annual growth rate of GDP at factor cost at 2004–2005 prices and that of services GDP are calculated, it is evident that both the services sector growth and GDP growth have more or less increased over time. It is clear that the services growth did not increase at a faster pace than aggregate GDP growth throughout the period from 1950–1951 to 2009–2010. But it is definitely observed that the services growth has outpaced aggregate GDP growth in all successive years from 1985 onwards. Thus, the growth of the services sector in India may be considered to have shown an enormous rise since the mid-1980s and subsequently increased by leaps and bounds thereafter in the post-globalization era. As a consequence to this phenomenon, the share of services in GDP has thus increased sufficiently during the period under consideration.

The tertiary sector emerged as the major sector of the economy in terms of both growth rates as well as its share in GDP in 1990s. It is to be noted here that while agriculture and manufacturing sectors have experienced phases of deceleration, stagnation and growth, the tertiary sector has shown a uniform growth trend during the period 1950–1951 to 1999–2010.

This sort of a transition from an agrarian economy to a service-oriented economy without achieving a high level of industrialization can be seen a sectoral jump, with that of an insignificant intermediate industrial sector deviating from the usual process of economic development.

(2) Share of the subsectors of services in GDP at factor cost at 2004–05 prices in India The emergence of services as the most dynamic sector in the Indian economy has in many ways been a revolution. The analysis of the sectoral composition of GDP for the period 1950–2010 brings out the fact that 'tertiarization' of the structure of production in India has taken place. During the process of growth over the years from 1950–1951 to 1999–2010, the Indian economy has experienced a change in production structure with a shift away from agriculture towards industry and tertiary sector. The various subsectors that comprise of the services sector, their respective share in services GDP and their average annual growth rates give us a picture that shows which subsector of services is growing fast and which is not. In India, the national income classification given by CSO is followed. In the National Income Accounting in India, service sector includes the following:

1. Trade, hotels and restaurants (THR)
    1. Trade
    2. Hotels and restaurants
2. Transport, storage and communication
    1. Railways
    2. Transport by other means

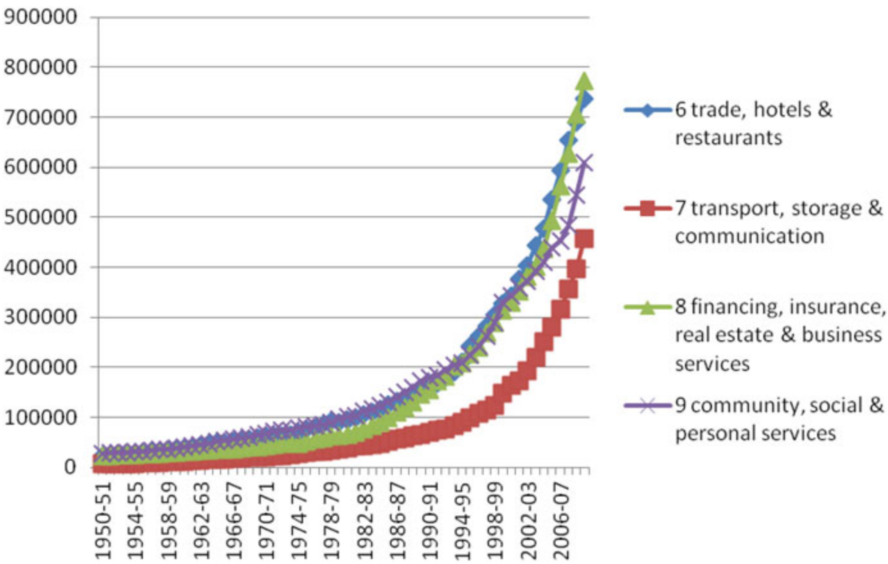


Fig. 18.3 Contribution of different services in total services and gross domestic product (GDP) at factor cost at 2004–2005 prices. (Source: CSO, EPW Foundation (2004), Own calculations)

   3. Storage
   4. Communication
3. Financing, insurance, real estate and business services
   1. Banking and insurance
   2. Real estate, ownership of dwellings and business services
4. Community, social and personal services
   1. Public administration and defense (PA & D)
   2. Other services

It is observed from Fig. 18.3 that almost all broad subsectors have grown over time, but the pick-up was the strongest in financing, insurance, real estate and trade, hotels and restaurants. There has been an increase in the transport, storage and communication, but the community, social and personal services have remained stagnant. The vertical axis in Fig. 18.3 measures GDP, services, and the subsector of services in Rs. crore. These activities account for the entire acceleration in services growth in the 1990s. The growth transport, storage, etc. in the 1990s were broadly similar to that of the previous decades.

A clearer view is available if we consider the share in services GDP at factor cost at 2004–2005 prices. Figure 18.4 shows that the share of subsectors like trade hotel and restaurants has increased substantially over the period from 1950–1951 to 2009–2010.

Share of subsectors of services in India GDP at factor cost at 2004–05 prices

Figure 18.4 shows that the share of subsectors like trade hotel and restaurants have increased substantially over the period from 1950–1951 to 2009–2010. The share of transport, storage and communication has increased many fold i.e. from 10 %

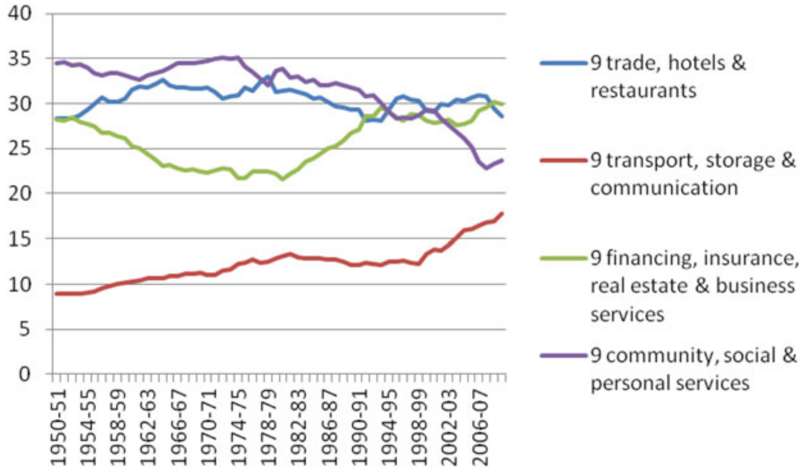


Fig. 18.4 Contribution of percentage share of different subsector of services as a proportion of services GDP at factor cost at 2004–2005 prices

in 1950–1951 to 18 % in 2009–2010. The shares of financing and real estate have increased over the period under consideration with a decrease somewhere in between, but it has captured almost 29 % share of the service sector over the years. The shares of community, social and personal services, however, have shown a considerable decrease over the concerned period. It has somewhat decreased from 35 % in 1950–1951 to 23 % in 2007–2008. This sector, which happened to be the main foundation of service sector, mainly growth services during the 1950s, happened to have lost its importance in terms of productivity especially during the post-globalization era.

The annual rate of growth of various subsectors expectedly reveals that financial services show the stable fastest rate of growth, followed rather closely by trade, hotels and restaurants. The rate of growth, in the transport, storage and communication sectors, though not very high, shows a major hike 2001–2002 onwards. Community and personal services show the lowest rate of growth among the subsectors of services.

However, these broad macroeconomic data do not allow for more than a somewhat cursory supposition of the forces at work and certainly do not make a definitive statement on the nature of services growth. Therefore, a detailed analytical research is required to examine the forces involved in such an enormous growth in the service sector in India especially in the post-liberalization era.

4 Analytical View of India's Service Sector

Given the magnitude of services growth and its interlinkages with other sectors of the economy, it is important to understand the impact of the service sector on other macroeconomic variables. In line with the above-mentioned macroeconomic view of the dominant trend in the services-led- growth, service sector in India has been growing rapidly specially in the past two decades. Its growth in fact has been higher than the growth in the other commodity-producing sectors, such as agriculture and manufacturing sectors.

For this analytical reasoning, the data on services GDP and GDP at factor cost in 2004–2005 prices as a whole from the period 1950–1951 to 2009–2010 are taken in account. A rigorous time series analyses like testing for stationary, presence of structural breaks, etc. are done with the series.

4.1 Service sector Growth in India: A time series analysis

During the past three decades, the methods of estimation of economic relationship and modelling fluctuations in economic activity have been subject to some fundamental changes. The method of estimation (ordinary least squares, (OLS)) of the standard regression model assumes that the mean and the variance of these variables are constant over time. Variables whose mean and variance change over time are known as non-stationary or unit root variables. It is now well established that different characterizations of the data-generating process of a macroeconomic time series have drastically dissimilar implications for theories and empirics in macroeconomics. For instance, traditional theories of economic fluctuations have claimed that: (1) fluctuations are mainly caused by aggregate demand shocks and (2) demand shocks have only short-term effects, and the economy reverts to the natural rate of output in the long run. Consequently, evidence of unit roots in real output time series compelled many to question the validity of these theories. In each case, the unit root properties of the variable considered are shown to have significant implications for economic theories. From an empirical perspective, the order of integration of macroeconomic variables has crucial consequences for appropriate modelling of time series data. These observations have led many economists to vigorously explore whether macroeconomic time series could be characterized as containing a unit root.

In their seminal contribution on the dynamic properties of macroeconomic time series, Nelson and Plosser (1982) found evidence in favour of the unit root hypothesis for 13 out of 14 economic and financial aggregates for the USA. Realizing the immense economic implications of this result, many economists focused their attention on the possible source of this result. In particular, Perron (1989; hereafter referred to as Perron) demonstrated that if the years of the Great Depression (1929) and the first oil crisis (1973) are treated as points of structural change in the economy and the observations corresponding to these years are removed from the noise function of the Nelson and Plosser data, then the result derived by Nelson and Ploser (1982) could be reversed for most of the variables. Based on his results, Perron asserted that Nelson and Plosser's strong evidence in support of the unit root hypothesis rested on failure to account for structural change in the data. Perron's approach consisted of incorporating an exogenous structural break in the model and then test for the presence of a unit root in the variable. Thus, dating of the potential break was assumed known a priori in Perron, and test statistics were constructed by adding dummy variables representing different intercepts and slopes, thereby extending the standard Dickey–Fuller procedure.

This approach was however questioned most notably by Banerjee et al. (1992), Christiano (1992) and Zivot and Andrews (1992), who argued that selecting the structural break a priori based on an ex post examination or knowledge of the data could lead to an over rejection of the unit root hypothesis. They pointed out that conventional critical values for test of parameter change are not valid when the breakpoint is inferred from examination of data. Additionally, Piehl et al. (1999) highlighted that the dummy variable may not actually enter at the appropriate time, due to uncertainty about the precise timing of the break, and for this reason, the estimated model may not be correct. In response, a number of studies have developed different methodologies for endogenizing the break dates in the analysis of unit root (e.g. Zivot and Andrews (1992), Lumsdaine and Papell (1997), Perron (1997) and Lee and Strazicich (2003)). This endogenization of break points had major impact on the unit root results. For instance, Zivot and Andrews (1992; hereafter referred to as Zivot and Andrews) were unable to reject the unit root hypothesis for four of the Nelson and Plosser series—-for which Perron rejected the hypothesis. Further, with finite sample critical values, they failed to reject the unit root null hypothesis for a further three series—employment, nominal wages and stock prices.

To this end, we use the procedure developed by Zivot and Andrews to test the null of unit root against the break-stationary alternative hypothesis. We also compare these results with the conventional unit root tests that do not account for any break in the data. The next section explains the econometric methodology.

4.2 Econometric Methodology

4.2.1 Unit Root Test Without Structural Break

We begin through testing for the presence of a unit root in each of the macroeconomic series using the Augmented Dicky–Fuller (1979) test. The *augmented Dickey–Fuller test* (ADF) constructs a parametric correction for higher-order correlation by assuming that the series follows an AR(k) process and adding lagged difference terms of the dependent variable to the right-hand side of the test regression:

$$\Delta y_t = c + \alpha yt - 1 + \sum_{j=1}^{k} dj \Delta yt - j + \varepsilon t$$

$$\Delta y_t = c + \alpha yt - 1 + \beta t + \sum_{j=1}^{k} dj \Delta yt - j + \varepsilon t$$

Equation (1) tests for the null of a unit root against a mean-stationary alternative in y_t, where y refers to the time series examined, and Eq.2) tests the null of a unit root against a trend-stationary alternative. The term $\Delta\ y_{t-\mathrm{j}}$ is lagged first differences to accommodate serial correlation in the errors. We select the lag length through the 't sig' approach proposed by Hall (1994). As shown by Ng and Perron (1995) the 't sig' approach produces test statistics which have better properties in terms of size and power than when lag length is selected with some information-based criteria.

As illustrated by Eqs. 1 and 2, we may elect to include a constant, *r* a constant and a linear time trend in ADF test regression. Phillips and Perron (1988) propose an alternative (nonparametric) method of controlling for serial correlation when testing for a unit root.

The Phillips and Perron (PP) method estimates the non-augmented Dickey–Fuller (DF) test equation (Eqs. 1 and 2 without $\sum_{j=1}^{k} d_j \Delta y_{t-j}$ term on right-hand side (rhs)), and modifies the t-ratio of the α coefficient, so that serial correlation does not affect the asymptotic distribution of the test statistic. For comparison purposes, we also perform the PP tests and report their results in addition to the generally favoured ADF test.

4.2.2 Unit Root Test with Single Structural Break: Zivot and Andrews Model

A problem common with the conventional unit root tests—such as the ADF, Dickey–Fuller-Generalized Least Squares (DF-GLS) and PP tests—is that they do not allow for the possibility of a structural break. Assuming the time of the break as an exogenous phenomenon, Perron showed that the power to reject a unit root decreases when the stationary alternative is true and a structural break is ignored.

Zivot and Andrews propose a variation of Perron's original test in which they assume that the exact time of the breakpoint is unknown. Instead, a data-dependent algorithm is used to proxy Perron's subjective procedure to determine the breakpoints. Following Perron's characterization of the form of structural break, Zivot and Andrews proceed with three models to test for a unit root: (1) model A, which permits a one-time change in the level of the series; (2) model B, which allows for a one-time change in the slope of the trend function; and (3) model C, which combines one-time changes in the level and the slope of the trend function of the series. Hence, to test for a unit root against the alternative of a one-time structural break, Zivot and Andrews use the following regression equations corresponding to the earlier mentioned three models.

$$\Delta y_t = c + \alpha yt - 1 + \beta t + \gamma DUt + \sum_{j=1}^{k} dj \Delta yt - j + \varepsilon t$$

$$\Delta y_t = c + \alpha yt - 1 + \beta t + \theta DTt + \sum_{j=1}^{k} dj \Delta yt - j + \varepsilon t$$

$$\Delta y_t = c + \alpha yt - 1 + \beta t + \gamma DUt + \theta DTt + \sum_{j=1}^{k} dj \Delta yt - j + \varepsilon t$$

where DU_t is an indicator dummy variable for a mean shift occurring at each possible break-date (TB) while DT_t is corresponding trend shift variable. Formally,

$$DU_t = \begin{cases} 1......if\ t > TB \\ 0.......otherwise \end{cases}$$

$$DT_t = \begin{cases} t - TB......if\ t > TB \\ 0.......otherwise \end{cases}$$

The null hypothesis in all the three models is $\alpha = 0$, which implies that the series $\{y_t\}$ contains a unit root with a drift that excludes any structural break, while the alternative hypothesis $\alpha < 0$ implies that the series is a trend-stationary process with a one-time break occurring at an unknown point in time. The Zivot and Andrews method regards every point as a potential break-date (TB) and runs a regression for every possible break-date sequentially. From amongst all possible break-points (TB), the procedure selects as its choice of break-date (*TB*) the date which minimizes the one-sided *t*-statistic for testing $\alpha\ (= \alpha - 1) = 1$. According to Zivot and Andrews, the presence of the end points causes the asymptotic distribution of the statistics to diverge towards infinity. Therefore, some region must be chosen such that the end points of the sample are not included. Zivot and Andrews suggest the 'trimming region' be specified as (0.15T, 0.85T), which we follow.

It has been proven that the Zivot–Andrews test among all, the overall $t - 2$ regressions one can choose that year as break year which gives the minimum value of the *t* statistics corresponding to the coefficient of Y_{t-1}. Further, one can choose that model as the best-fitted model which gives the minimum *t* value of the coefficient Y_{t-1}. Now after finding the best-fitted model and the breakpoint, the estimated results are compared with the critical values given by the Zivot and Andrews to determine the nature of the series.

Before going into the analysis, we will first look at the implications about the significance of the coefficients**:**

- If the coefficient of Y_{t-1} is significant, then it can be concluded that the underlying series is trend stationary (TS). This means that the growth of different sectors converges to a deterministic trend.
- If the coefficient of Y_{t-1} is not significant, then it can be concluded that the underlying series is difference stationary (DS). The implication of a DS series is that there exists a stochastic trend.
- If the coefficient of time is positive (negative) and statistically significant, then it implies that the underlying process is TS around a deterministic positive (negative) trend. Consideration of the significance of the time coefficient is meaningful only when the underlying process is TS type.
- If the series is DS type and the coefficient of time is positive (negative) and significant, then one can conclude that the degree of stability increased (deceased) over time.

Table 18.1 Results of the different unit root tests

Variables	ADF test	PP test	KPSS test (LM statistic)
Agricultural growth	− 8.475 (0.0000)	− 19.576 (0.0000)	0.085557
Manufacturing growth	− 5.714 (0.0001)	− 5.469 (0.0002)	0.131763
Industrial growth	− 5.888 (0.0000)	− 5.758 (0.0001)	0.124400
Services growth	− 5.539 (0.0000)	− 5.539 (0.0001)	0.150782
Growth of GDP	− 9.456 (0.0000)	− 9.643 (0.0000)	0.116273
Growth of trade, hotel and restaurants	− 6.048 (0.0000)	− 6.032 (0.0000)	0.155200
Growth of transport, storage and communication	− 5.168 (0.0004)	− 5.149 (0.0005)	0.225337
Growth of financial Services	− 6.271 (0.0000)	− 6.252 (0.0000)	0.086983
Growth of personal, social and community services	− 4.934 (0.0000)	− 4.934 (0.0000)	0.054035

- If DU_t is statistically significant, then one can conclude that there exists a significant change in the post-break period. Such a conclusion is meaningful only if one can find that the underlying process is of TS type.
- If DT_t is positive (negative) and statistically significant, then we can conclude that there exists positive (negative) and significant break in the growth of the series of that period.

5 Empirical Results and Economic Implications

The results of unit root tests are shown in Table 18.1.

The ADF test along with both PP and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests confirms that all these series taken into account are stationary.

The results of the detailed ADF test is given in Table 18.2.

The results of the Zivot–Andrews test can be formulated in Table 18.3.

The estimated growth equation of the Zivot–Andrews model is given in Table 18.4.

The results of both the ADF and Zivot–Andrews tests can be inferred in the following discussions:

5.1 Agricultural Growth

In case of agricultural growth, the ADF test concludes that the data are stationary at the levels. The null hypothesis of the presence of unit root is rejected at 5 % level of significance.

Table 18.2 Detailed results of the ADF (augmented Dickey–Fuller test) test

Variables/Coefficients	Coefficient of $Y_{t-1} = \alpha$	Constant c	Coefficient of trend β	Coefficient of lag variable δ
Agricultural growth	– 1.914[a] (– 8.475)	4.295[a] (2.871)	0.022 (0.538)	0.306[a] (2.323)
Manufacturing growth	– 0.745[a] (– 5.714)	3.590[a] (3.465)	0.024 (0.860))	
Industrial growth	– 0.774[a] (– 5.888)	3.590[a] (3.736)	0.026 (1.069)	
Services growth	– 0.719[a] (– 5.539)	2.153[a] (4.128)	0.069[a] (4.285)	
Growth of GDP	– 1.239[a] (– 9.456)	3.059[a] (3.736)	0.092[a] (3.869)	
Growth of trade, hotel and restaurants	– 0.799[a] (– 6.048)	3.037[a] (3.548)	0.056[a] (2.417)	
Growth of transport, storage and communication	– 0.676[a] (– 5.168)	2.459[a] (3.084)	0.078[a] (3.395)	
Growth of financial services	– 0.836[a] (– 6.271)	1.414[a] (2.178)	0.119[a] (4.544)	
Growth of personal, social and community services	– 0.652[a] (– 4.934)	2.129[a] (3.040)	0.044[a] (2.527)	

[a] Significant at the 5 % level.

Table 18.3 Zivot–Andrews test: Minimum *t* statistic and the breakpoints

Variables/Coefficients	Minimum *t* statistic	Breakpoint	Chosen model	Nature of the series
Agricultural growth	– 6.428[a]	1969	C	TS
Manufacturing growth	– 6.870[a]	1966	A	TS
Industrial growth	– 6.760[a]	1966	C	TS
Services growth	– 7.128[a]	1966	C	TS
Growth of GDP	– 6.753[a]	1966	A	TS
Growth of trade, hotel and restaurants	– 7.131[a]	1966	A	TS
Growth of transport, storage and communication	– 4.211	1992	B	DS
Growth of financial services	– 7.433[a]	1982	A	TS
Growth of personal, social and community services	– 5.290[a]	1966	C	TS

[a] The critical values of the Zivot–Andrews test at 5 % level of significance are – 4.80, – 4.42 and – 5.08, respectively.

The Zivot–Andrews test confirms the ADF test. The best-fitted model is Model C, and the series is stationary with a single break in 1969. The underlying series is time stationary, which means that the agricultural growth is converging to a negative deterministic trend. The underlying series is TS type, indicating that the variance in agricultural growth is constant and independent of time. As the coefficients of both DU_t and DT_t are positive and significant, there exists a positive trend in the series, and the growth rate has increased in the post break period.

One of the major reasons behind growth in agricultural growth may be due to implementation of Green Revolution during that time.

Table 18.4 Estimated equation of growth by Zivot–Andrews model

Variables/Coefficients	Coefficient of C	Coefficient of DU	Coefficient of Time	Coefficient of DT	Coefficient of Y_{t-1}
Agricultural growth	21.531[a] (− 6.428)	8.339[a] (2.581)	− 1.0185[a] (− 2.137)	1.074[a] (2.216)	− 5.044[a] (− 6.428)
Manufacturing growth	10.437[a] (5.733)	− 6.701[a] (− 4.057)	0.166[a] (4.141)		− 1.816[a] (− 6.870)
Industrial growth	3.669[a] (1.870)	− 4.763[a] (− 2.525)	0.279[a] (1.271)	− 0.186[a] (− 0.844)	− 0.917[a] (− 6.760)
Services growth	2.811[a] (3.271)	− 2.624[a] (− 3.237)	0.195[a] (2.048)	− 0.064[a] (− 0.699)	− 0.944[a] (− 7.128)
Growth of GDP	6.708[a] (5.192)	− 5.037[a] (− 3.844)	0.263[a] (5.863)		− 2.348[a] (− 6.753)
Growth of trade, hotel and restaurants	4.375[a] (4.785)	− 3.632[a] (− 3.013)	0.132[a] (3.972)		− 0.946[a] (− 7.131)
Growth of transport, storage and communication	9.893[a] (4.050)		− 0.024[a] (− 0.817)	0.666[a] (4.044)	− 1.601[a] (− 4.211)
Growth of financial services	2.669[a] (3.729)	3.704[a] (3.218)	0.040[a] (1.181)		− 0.959[a] (− 7.433)
Growth of personal, social and community services	1.177[a] (0.907)	− 2.317[a] (− 1.903)	0.253[a] (1.675)	− 0.816[a] (− 1.239)	0.711[a] (− 5.290)

[a] The critical values of the Zivot–Andrews test at 5 % level of significance are − 4.80, − 4.42 and − 5.08, respectively.

5.2 *Manufacturing Growth and Industrial Growth*

Here, industry includes gas electricity and water supply and construction other than manufacturing (registered and unregistered). Based on the ADF and PP tests, both these series are stationary as we reject the null hypothesis of the presence of unit root.

For manufacturing growth and industrial growth, the best-fitted model is model A and model C, respectively, though the breakpoint is 1966. In case of manufacturing growth, the coefficient of time is positive and statistically significant implying that growth of this sector is TS around a positive deterministic trend. Since the intercept dummy is statistically significant, there exists a significant change in the post-break period. In case of industrial growth, the coefficient of time is positive and statistically significant, implying that growth of this sector is TS around a positive deterministic trend. Here, both the dummies are negative and statistically significant, implying there has been a negative and significant change in the post-break period, and there has been a break in the growth of industrial growth. This can be attributable to industrial stagnation during the mid-1960s.

5.3 *Services Growth*

The ADF and PP tests indicate that the growth of services series is stationary. Though stochastic trend is absent, there is a presence of a deterministic trend.

For growth of services, the best-fitted model is model C and the breakpoint year is 1966. In case of services growth, the coefficient of time is positive and statistically significant implying that growth of this sector is TS around a positive deterministic trend. Here, both the dummies are negative and statistically significant implying there has been a negative and significant change in the post-break period, and there has been a negative break in the growth of services growth.

5.4 Growth of GDP

The ADF and PP tests indicate that the growth of GDP at factor cost is stationary. Though there is no stochastic trend, there is a presence of a deterministic trend.

The Zivot–Andrews test confirms the ADF test. For growth of services, the best-fitted model is model A and the breakpoint year is 1966. In case of services growth, the coefficient of time is positive and statistically significant implying that growth of this sector is TS around a positive deterministic trend. Here, the intercept dummy is negative and statistically significant implying there has been a negative and significant change in the post-break period.

5.5 Growth of Trade, Hotel and Restaurants

The ADF and PP tests indicate that the growth of trade, hotel and restaurants is stationary. Though there is no stochastic trend, there is a presence of a deterministic trend.

The Zivot–Andrews test confirms the ADF test. For growth of trade, hotel and restaurants, the best fitted model is model A and the breakpoint year is 1966. In case of services growth, the coefficient of time is positive and statistically significant implying that growth of this sector is TS around a positive deterministic trend. Here, the intercept dummy is negative and statistically significant implying there has been a negative and significant change in the post-break period.

5.6 Growth of Transport, Storage and Communication

The ADF and PP tests indicate that the growth of transport, storage and communication is stationary. Though there is no stochastic trend, there is a presence of a deterministic trend.

Here, the Zivot–Andrews test indicates that the growth of transport, storage and communication is DS implying that the variances in the production are not constant. The coefficient of time is negative and significant implying that the variances have decreased over time.

5.7 *Growth of Financial Services*

The ADF and PP tests indicate that the growth of financial services is stationary. Though there is no stochastic trend, there is a presence of a determined trend.

However, the Zivot–Andrews test confirms the ADF test. For growth of financial services, the best fitted model is A and the breakpoint year is 1982. In case of financial services growth, the coefficient of time is positive and statistically significant, implying that growth of this sector is TS around a positive deterministic trend. Here, the intercept dummy is negative and statistically significant implying there has been a negative and significant change in the post-break period.

5.8 *Growth of Personal, Social and Community Services*

The ADF and PP tests indicate that the growth of personal, social and community services is stationary. Though there is no stochastic trend, there is a presence of a determined trend.

However, according to the Zivot–Andrews test the best fitted model is model C and the breakpoint year is 1966. In case of personal, social and community services growth, the coefficient of time is positive and statistically significant implying that growth of this sector is TS around a positive deterministic trend. Here, both the dummies are negative and statistically significant implying there has been a negative and significant change in the post-break period, and there has been a negative break in the growth of services growth.

6 Conclusion

This chapter uses annual data to determine endogenously the most important years when structural breaks occurred and simultaneously test for the unit root hypothesis in the presence of these breaks in the growth of eight components of GDP and also GDP growth. Finally, we applied the Zivot–Andrews test. Some key conclusions follow from the results obtained from these tests. First, for all these series of growth we reject null hypothesis of unit root stating that all of them were stationary. The Zivot–Andrews tests, however, indicates that except growth of transport, storage and communication, all the growth variables have a significant structural break. Second, it is found that almost all the series exhibit structural breaks during 1960s mainly around 1966–1969. The Green Revolution and industrial stagnation in the Indian economy may be reasons for such a break. The results suggest that a one-time break in the growth of financial services at 1982. Lastly, it is important to recognize that the earlier mentioned results are derived through endogenously determining the presence of a single structural break. However, it may be argued that data may contain more than one structural break. Lee and Strazicich (2003) point out considering only one break when in fact two are present, which can result in loss of power of the test. Therefore, this analysis could be extended for the case of more than one structural break.

References

Balakrishnan P, Parameshwaran M (2007) Understanding economic growth in India: a prerequisite. Econ Polit Weekly 2915–2922
Banerjee A, Lumsdaine R, Stock J (1992) Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence. J Bus Econ Stat 10:271–287
Banga R (2005a) Critical issues in India's services led growth. ICRIER Working Paper 171
Banga R (2005b) Role of services in the growth process: a survey. ICRIER Working Paper 159
Banga R, Goldar BN (2004) Contribution of services to output growth productivity in indian manufacturing: pre and post reform. ICRIER Working Paper 139 (August)
Basu K, Maertens A (2009) The growth of industry and services in South Asia. In: Ghani E, Ahmed S (eds) Accelerating growth and job creation in South Asia. Oxford University Press, Oxford
Baumol WJ (1967) Macroeconomics of unbalanced growth: an anatomy of urban crisis. Am Econ Rev 57(3):415–426
Bhagwati J, Panagariya A, Srinivasan TN (2004) The muddles over outsourcing. J Econ Perspect 18(4):93–114
Bhagwati JN (1984) Splintering and disembodiment of services and developing nations. World Econ 7(2):133–143
Bhattacharya BB., Mitra, Arup (1989) Agriculture industry growth rates: widening disparity. Econ Polit Weekly vol. XXIV, No. 34 (August 26)
Bhattacharya BB, Mitra A (1990) Excess growth of the tertiary sector in Indian economy: issues and implications. Econ Polit Weekly 25(44):2445–2450 (November 3)
Bhattacharya BB, Mitra A (1991) Excess growth of the tertiary sector. Econ Polit Weekly, June 1–8
Bhattacharya BB, Mitra A (1997) Changing composition of employment in tertiary sector: a cross-country analysis. Econ Polit Weekly, March 15
Bhattachyara BB, Mitra A (1989) Widening disparity between industry- agriculture growth rates: an explanation of recent phenomenon. Econ Polit Weekly, August 26
Bhattacharya BB, Sakthivel S (2004) Regional growth and disparity in India. Econ Polit Weekly, March 6
Boyce J (1987) Agrarian impasse in Bengal: institutional constraints to technological change. Oxford University Press
Chen S, Datt G, Ravallion M (1994) Is poverty increasing in the developing world? Rev Income Wealth, December vol 40, No 4
Christiano L (1992) Searching for a break in GNP. J Bus Econ Stat 10:237–250
Clark C (1940) The conditions of economic progress. Macmillan and Co, Ltd., London
Dasgupta S, Singh A (2005) Will services be the new engine of indian economic growth? Dev Change 36(6):1035–1057
Datt G, Ravallion M (1998) Farm productivity and rural poverty in India. J Dev Stud, vol 34 No 4
Datt G, Ravallion M (2002) Is India's economic growth leaving the poor behind? J Econ Perspect vol 16, No 3
Datta M (1989) Tertiary sector and net material product: indian economy, 1950–51 to 1983–84. Econ Polit Weekly (September 23)
Datta M (2001) The significance and growth of the tertiary sector. Delhi: Northern Book Centre
Dicky D, Fuller WA (1979) Distribution of the estimates of autoregressive time series with a unit root. J Am Statist Assoc 74:427–431
Economic Survey (2006–2007) Government of India
Elfring T (1989) New evidence on the expansion of service employment in advanced economies. Rev Income Wealth, Series 35, No.4
EPW Research Foundation (2003) Domestic products of states of India, 1960–61 to 2001–02
Fischer AGB (1935) The clash of progress and security. Macmillan and Co, Ltd., London
Fuchs VR (1980) Economic growth and the rise of service employment. National Bureau of Economic Research Working Paper 486 (New York)

Gemmel N (1986) Structural change and economic development: the role of service sector. Macmillan Press, Hampshire
Gordon J, Gupta P (2004) Understanding India's services revolution. IMF Working Paper WP/04/171, September
Griliches Z (1992) Output measurements in the service sectors. The University of Chicago Press, Chicago
Hansda S (2001) Sustainability of services led growth: an input- output analysis of the Indian economy. RBI Occasional Working Paper (vol 22, No. 1, 2 and 3)
Hill TP (1977) On goods and services. Rev Income Wealth 23:315–338
Kaldor N (1967) Strategic factors in economic development. Cornell University Press, Ithaca
Katouzian MA (1970) The development of the service sector: a new approach. Oxford Economic Papers 22(3):362–382 (November)
Krueger AO, Lary HB, Manson T, Akrasanee N (eds), (1981) Trade and employment in developing countries: individual studies. University of Chicago Press, Chicago
Kuznets S (1955) Economic growth and income inequality. Am Econ Rev 45(1):1–28 (March)
Kuznets S (1966) Modern economic growth, rate, structure and spread. Yale University Press, New Haven
Lee J, Strazicich MC (2004) Minimum lagrange multiplier unit root test with two structural breaks. Rev Econ Stat 85(4):1082–1089
Lumsdaine R, Papell D (1997) Multiple trend breaks and the unit root hypothesis. Rev Econ Stat 79:212–218
Maddala GS, In-Moo K (1998) Unit roots, cointegration and structural change. Cambridge University Press
Melvin JR (1989) Trade in services: a theoretical analysis. Institute of Research and Policy, Halifax
Melvin JR (1995) History and measurement of the service sector: a review. Rev Income Wealth, Series 41, no. 4, 481–494 (December)
Narayan PK, Smyth R (2004) Structural breaks and unit root in Australian macroeconomic time series. Monash University Discussion Paper 18/04
National Income Accounts Statistics of India, 1950–1951 to 2002–2003. EPW Research Foundation
Nelson CR, Plosser CI (1982) Trend and random walks in macroeconomic time series. J Monetary Econ 10:139–162
Panchmukhi VR, Nambiar RG, Meheta R (1986) Structural change and economic growth in developing countries. VIII World Economic Congress of the International Economic Association, Theme 4 (December)
Papola TS (1981) Informal sector in a developing economy. Vikas Publishing House, Delhi
Papola TS (2007) Employment trend in India. In: Kaushik B (ed) The Oxford companion to economics in India. Oxford University Press, New Delhi, pp 131–136
Perron P (1997) Further evidence on breaking trend functions in macroeconomic variables. J Econom 80:355–385
Perron P (1989) The great crash, the oil price shock and the unit root hypothesis. Econometrica 57:1361–1401
Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75(2):335–346
Rakshit M (2007) Services-led growth: the Indian experience, money and finance. ICRA Bulletin (February)
Sampson GP, Snape RH (1985) Identifying the issues in trade in services. World Econ 8(2):171–182
Sapir A, Winter C (1994) Services trade. In: Greenway P, Winters A (eds) Surveys in international trade. Blackwell Publishing, Oxford
Sarkar S, Mehta BS (2005) Employment profile of ICT sector in India. Institute for Human Development, New Delhi

Sastry DVS, Singh B, Bhattacharya K, Unnikrishnan NK (2003) Sectoral linkages and growth prospects: reflections on the indian economy. Econ Polit Weekly 38:2390–2397 (June 14)
Sen A (2003) On unit root tests when the alternative is a trend break stationary process. J Bus Econ Stat 21:174–184
Singh N (2006) Services led industrialization in India: assessment and lessons. UCGC, MPRA Paper No. 1446
Singh N (2006) Services-led industrialization in India: prospects and challenges. Centre for International Development, Working Paper 290, Stanford University
Stern RM, Bernard H (1987) Analytical issues and data needs for GATT negotiation on services. The World Economy 10:39–60
Zivot E, Andrews D (1992) Further evidence of great crash, the oil price shock and unit root hypothesis. J Bus Econ Stat 10:251–270

Chapter 19
Infrastructure Development and Regional Growth in India

Prakash Singh and N. R. Bhanumurthy

1 Introduction

It is indeed an honour and privilege to contribute a chapter to a Festschrift for Prof. Biswajit Chatterjee. During his long association with Jadavpur University, Kolkata, Prof. Chatterjee, with the help of his students, has pioneered a large number of scholarly research studies in the area of international trade, labour studies, agricultural issues, poverty and inequality and other environment and development-related issues. His contributions to these research areas are widely acknowledged in the professional realm. His presidential address to the 48th Annual Conference of the Indian Econometric Society at Pondicherry University, on the issue of poverty and inequality, would be a reference piece for the enormous work that Indian scholars have done over a period of time. Here, an attempt has been made to understand the role of infrastructure development on the regional imbalances in India as this issue is close to Prof. Chatterjee's heart.

Under the planned economy model, the central focus of the successive governments in India was to attain balanced regional development, equitable distribution of growth outcomes across the regions and to bridge the gap between leading and lagging states through investment promotion and other special schemes. However, the actual experience shows the other way around.[1] There exists a vast literature on the existence of, and many argue in favour of, widening regional imbalances, particularly in the post-reform period. The research question at this juncture is: What factors contributed to such imbalances? Many studies have explored this issue based

[1] A centralised planning system was advocated to curb the bias in economic development stemming out of difference in the level of resources available to states.

P. Singh (✉)
Research Scholarat at Dept. of Business Economics, Delhi University & ICSSR,
Doctoral Fellow at Institute of Economic Growth, New Delhi, India
e-mail: prakash.archa@gmail.com

N. R. Bhanumurthy
National Institute of Public Finance and Policy, New Delhi, India
e-mail: nrbmurthy@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_19, © Springer India 2014

on various growth models ranging from neo-classical, endogenous growth models to the recent stream of new economic geography models. In the literature, factors such as timing and extent of reform initiatives, structural issues, investments, location, etc., have been identified to attribute in the persistence and widening of such divergences.[2] However, there could be other issues which arise purely from the supply side such as financial access, infrastructure, governance, etc., that have not yet received enough attention while understanding such divergences. In our recent study, we have attempted to examine the role of financial access in regional growth (see Bhanumurthy and Singh 2013); here, an attempt has been made to understand the contribution of infrastructure development in regional divergences in India.[3]

The existing literature, by and large, confirms the positive role of infrastructure, but the result of these studies cannot be taken as confirmatory as they suffer from the methodological weaknesses. Although most of the studies are based on panel data models, the results may not be robust as many of them have not addressed the time series property of panel data during estimations. Such results could be spurious and could further lead to a misleading inference (Granger and Newbold 1974). Therefore, in this study, an attempt has been made to examine the role of infrastructure on economic growth in Indian states and, at the same time, the study tries to overcome the methodological issues by considering recent advances in panel econometric models which take care of the time series property of the panel data. thus, in our view, the present study is expected to produce better results than the previous studies.

The rest of the chapter is organized as follows. The analytical framework adopted in the study would be discussed in Sect. 2. Section 3 discusses the details about the database. A brief description about the methodology adopted in the study would be presented in Sect. 4. In Sect. 5, econometric results regarding the impact of infrastructure at the state level would be dealt with. The last section would draw conclusion.

2 Analytical Framework for the Study

Infrastructure is a crucial factor of production from the supply side and is used by both households as well as by producers and is expected to contribute significantly to the overall productivity (both labour and capital) and efficiency of the production process. Following Holtz-Eakin and Schwartz (1995) who introduced public expenditure as a growth determinant in growth regression, we consider expenditure on social and economic infrastructure as a separate factor of production, as these are the crucial

[2] Ahluwalia 2000 and 2002, Nagaraj et al. 2000, Rao, Shand and Kalirajan 1999, Shand and Bhide 2000, and Aggarwal and Singh (2013) are some of the recent studies which observed the presence of increasing regional imbalance in India.

[3] The World Development Report of 1994, by pointing out that productivity growth is higher in countries with an adequate and efficient supply of infrastructure services, endorses the instrumental role of infrastructure in the economic development process. Better infrastructure availability helps in attracting foreign capital and thus further helps in increasing and improving the production process which finally culminates in welfare.

elements of governments' expenditure apart from large transfer payments in the form of subsidies. Thus, to examine the effect of social and physical infrastructure at the state level in India, the analytical framework in this study uses a production function approach, wherein infrastructure variables enter the function in the form of capital. For this purpose, the study adopts the Cobb–Douglas production function with some variation.

A standard production function with labour and capital as inputs is taken and the same is extended by including infrastructure as additional input. Endogenous growth theory extends the definition of the word capital to include other forms of capital such as human capital, social capital, infrastructure and financial capital in addition to the microeconomic definition of capital that includes machinery, i.e. physical capital and capital in the form of investments. To make things simpler, as per the requirement of the study, we have not included all the measures of capital. Instead, here the focus is on capital at the macroeconomic level, i.e. infrastructure. Broadly, infrastructure can be divided into physical and social infrastructure. Physical infrastructure includes transport facility (road, rail, sea and air), telecommunication, electricity, etc., whereas education, health and sanitation, etc., are grouped under social infrastructure.

In this framework, infrastructure affects output in two ways; one is the direct channel where infrastructure increases the output by reducing the cost of intermediate goods and helps in achieving higher investments (Bougheas et al. 2000). For example, a proposed business investment avenue at a proposed location will lose its viability because of the infrastructure unavailability (transport, telecommunication, electricity, etc.) but availability of infrastructure will make the investment viable. Thus, higher per capita availability of infrastructure capital reduces the fixed cost of production and subsequently crowds in private investments also. The other channel where infrastructure affects the output positively is through the externality effect. Better infrastructure improves return on human capital due to good quality education and health and improves efficiency of the human capital by lowering the marginal depreciation of the capital. Additionally, physical infrastructure affects the cost/output by its effect on social infrastructure. Better physical infrastructure helps in raising the human capital through increasing its efficiency, which in turn affects output and lower fixed cost of production.

Considering the arguments of the infrastructure–growth literature, the theoretical model which can be used for empirical verification of the role of infrastructure (both physical and social) in growth dynamics at the state level for India is specified below:

$Y = F$ (IP, IS and L)
where Y is output (state domestic product, SDP),
IP is *expenditure* on physical infrastructure,
IS is *expenditure* on social infrastructure and
L is *labour*.

In general, physical infrastructure can include revenue and capital expenditure on transport, energy and irrigation, and flood control whereas revenue and capital expenditure on education, health and family welfare, and water supply and sanitation is considered social infrastructure. As discussed previously, increase in the social infrastructure is expected to have a positive effect on output through its positive long-run effect on the development of human capital in both quantity and quality terms. Similarly, improvements in physical infrastructure capital will have a positive effect on state level output by reducing the fixed cost as well as its effect on human capital. Due to unavailability of data at state level, we excluded labour from the empirical model although it is an integral part of the growth model.

3 Review of Literature

Infrastructure draws its importance largely owing to the fact that it is an input in production function. Further, it plays an important role in minimizing cost and improving the quality of other inputs including human capital, making them complementary to these inputs of production.[4] Keeping the encompassing role of infrastructure in economic development, the present section tries to review some of the existing studies which have dealt with infrastructure and economic development.

It is the seminal work of Aschauer (1989a, b and c), which gave thrust to the idea of 'infrastructure as determinant of growth' in the empirical literature, that provides evidence on the high economic return associated with investments in infrastructure. In fact, these studies correlate the slowdown in the productivity of the US economy to that of decline in investment in the infrastructure sector. The results of these studies, where economic returns from infrastructure investments are as high as 60 %, invited much debate concerning the use of production function and estimation methods in investigating the effect of infrastructure on economic growth. Later, Munnell (1990); Garcia–Milà and McGuire (1992); Uchimura and Gao (1993); Canning and Fay (1993) and Easterley and Rabelo (1993) provide almost similar evidence on high return of infrastructure investments to that of Aschauer (1989c) and thus testified the conjecture. However, a host of other economists, though, questioned the results of Aschauer, mostly on the ground of unrealistically high infrastructure elasticity but they did not completely refute the role played by infrastructure (Hulten and Schwab 1991; Evans and Karras 1994; Holtz–Eakin 1994; Gramlich 1994 and Garcia–Milà et al. 1996).[5] The first wave of such studies has been criticized for the methodological shortfall (Gramlich 1994; Garcia–Milà et al. 1996; Canning and Pedroni 2004 have raised the issue of reverse causation and non-stationary series issues) and use of only developed countries data to establish the result.

Taking infrastructure as technology, which reduces cost in the production of intermediate inputs, Bougheas et al. (2000) highlight the importance of infrastructure

[4] According to the survey of firms in the World Bank investment climate assessment, infrastructure is considered as the major source of hurdle in business operation in developing and least developed countries.

[5] They argue that a positive and statistically significant coefficient for a government input in an estimated 'production function' may only indicate the degree to which increased income causes an increased level of government activities.

in the growth process. Whereas, Argy et al. (1999) view infrastructure as a catalyst, which not only enables opportunities for economic development but also creates future opportunities provided the government institutes a sound and active policy for investment in infrastructure. Although the results of Canning and Pedroni (2004) demonstrate a strong and positive inducing effect of infrastructure on economic growth, there exists vast variation in this inducing effect. The authors attribute this variation to the existence of infrastructure beyond the growth maximizing level. In other words, the study raised an issue of threshold level of infrastructure facilities. However, these findings may not be applicable to most of the developing countries as infrastructure constraints are obvious and quite high.

Following the methodology of the pioneering work of Hulten and Schwab (1991), which allows accounting for the externalities' effect of infrastructure on growth exclusively, Hulten et al. (2006) found a significant spillover effect of infrastructure on total factor productivity (TFP) in the Indian manufacturing sector, contrary to the findings of Hulten and Schwab (1991) for the US manufacturing sector. A study by O'Fallon (2003) though fails to provide a causal link between infrastructure investment and economic growth, however, provides an interesting conclusion that the impact of infrastructure on growth depends on the initial conditions of the economy. Rodriguez (2007) documents the significance of infrastructure investment in increasing or decreasing the growth rate of the economies.

Although most of the studies on the infrastructure–growth relationship have shown a positive effect of infrastructure on economic development, either indirectly through productivity or through its direct effect on output, there are some studies which have reported negative results of infrastructure on growth (Devarajan et al. 1996; Sanchez–Robles 1998 and Pritchett 1996). These studies argue that excessive amount of transportation and communication expenditures make capital expenditures unproductive, which imply that governments of developing countries have been misallocating public expenditures in favour of capital expenditures at the expense of current expenditures and Pritchett (1996) brought the issue of public investment in unproductive projects making marginal productivity of the output with respect to capital lower than the investment.

In the Indian context and particularly at the subnational level, there exist a plethora of studies examining the role of infrastructure in the economic development of subnational units based on different methodologies, indicators and time periods. In one of the earliest studies, Tewari (1984) examines the interrelationship between economic infrastructure and development and tries to identify the role of the former in the latter through analysis of state level data at two time points—1970–1971 and 1980–1981. He obtains a significantly positive relationship between infrastructure and development, and especially economic infrastructure. Similarly, Elhance and Lakshmanan (1988) exhibit that investment in infrastructure (both physical and social) facilitates reduction in the production cost in the manufacturing industry in India. Binswanger et al. (1989) using district level data for India examine the impact of physical infrastructure on agricultural output and illustrate that infrastructure helps in reducing the transaction cost, and, thus, promoting agriculture output. The combined results of these two studies indicate the importance of infrastructure for all the sectors of the economy. Ghosh and De (1998), Dutt and Ravallion (1998) show the deterministic role of infrastructure in explaining growth and poverty divergence across states. Sahoo and Saxena (1999) based on the production function approach show that infrastructure (transport, electricity, gas, water supply and communication) facilities have an increasing return to scale. Studying the link between infrastructure and development for West Bengal, Majumdar and Mukherjee (2005) confirm the existence of a long-run relationship between infrastructure and development with a strong causation from infrastructure availability on development levels. Additionally, effect of different facets of infrastructure seems to have different impacts on different dimensions of development.

Dutta et al. (2007), in a study of 14 states in India, exhibit the importance of infrastructure on macroeconomic growth, particularly the role of economic infrastructure in determining the productivity. The results of the study indicate that infrastructure plays an important role in determining the level of investment and productivity of the industrial activity. Recently, Patra and Acharya (2011) with more indicators of infrastructure examined the contour of infrastructure, regional growth and poverty dynamics. The study result corroborates the earlier finding indicating an urgent need of policy intervention in enhancing infrastructural investments to achieve the 12th Five Year Plan's inclusive growth goal. In a different study, Agarwalla (2011) endorses the role of infrastructure in productivity growth at the state level.

To sum up, the review shows ambiguous results regarding the impact of infrastructure on growth. Although most of the studies show a positive impact of infrastructure on growth, there are non-linearities as excessive investments might have negative impact. However, this situation may not be the same in developing countries like India where the infrastructure deficit has been clearly identified as one of the major factors that could hamper the country in sustaining high growth. Although we have a large number of studies examining the role of infrastructure in the economic development process at the subnational level in the Indian context that largely support the positive role of infrastructure on regional growth, many suffer either from the problem of methodological inadequacy or from coverage used in the study and do not identify the true correlation between infrastructure and development. Hence, in this study, we try to re-examine the impact of infrastructure on economic development in Indian states with the help of robust econometric methods.

4 Data Description

The study uses annual data for 15 states for the period 1985–1986 to 2007–2008. *Handbook of Statistics on Indian Economy* published by the Reserve Bank of India (RBI) is the source of SDP. For the data on infrastructure (both physical and social), the study largely relied on the RBI's various reports of 'State Finances: A Study of Budgets'.

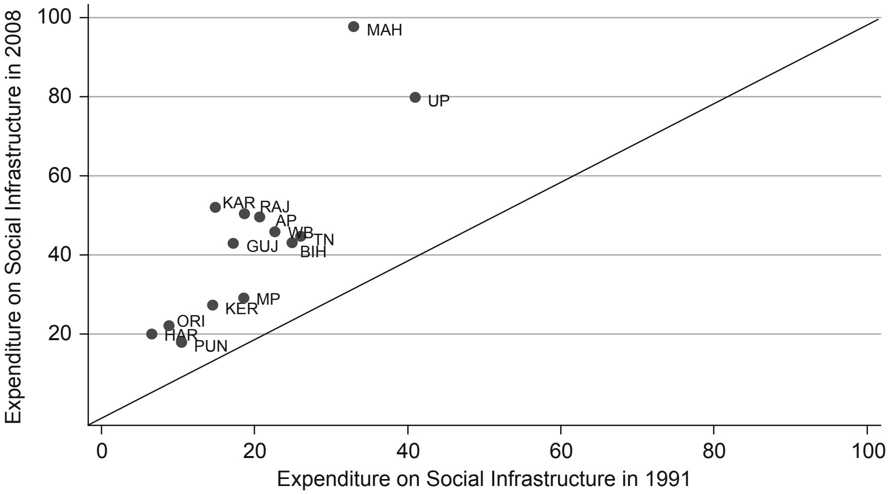


Fig. 19.1 State of social infrastructure expenditure. (Source: Authors' calculation based on data from State Finances: A Study of Budgets, RBI)

The following variables are used in the study: SDP (real net SDP at 1993–1994 prices in rupee crore), public expenditure on economic infrastructure indicators (irrigation, energy and transport in real terms in rupee crore; (RSOC)) to account for economic infrastructure and expenditure on indicators of social infrastructure—(health, education, water and sanitation in real terms in rupee crore; RECO).[6] In the case of SDP, there is a problem of different base years. As the change in base year also reflects the change in the production function, we have converted the SDP series with the 1993–1994 base by using a production function splicing (see Bhanumurthy and Singh (2013) for base year conversion details). Before we discuss the methodology and results, in the next section, a brief discussion on the trends in infrastructure indicators at the state level is presented.

5 Trends in Infrastructure Indicators

With the help of cross-sectional plots, we try to understand and investigate at the primary level the changes in infrastructure (both economic and social) at the time of reforms and the present level and its relation with the state output level. Figure 19.1 examines changes in social infrastructure in 2008–2009 compared to the situation in 1991–1992, the point when reforms took place. The plot of expenditure on social infrastructure depicts improvement in the situation of the social infrastructure in almost all the states, though the improvement in the condition is not very huge in some states. This only indicates that the expenditure on social sector development has increased over a period of time, which is in line with the government's approach

[6] To convert a nominal series into a real series, an SDP deflator has been used.

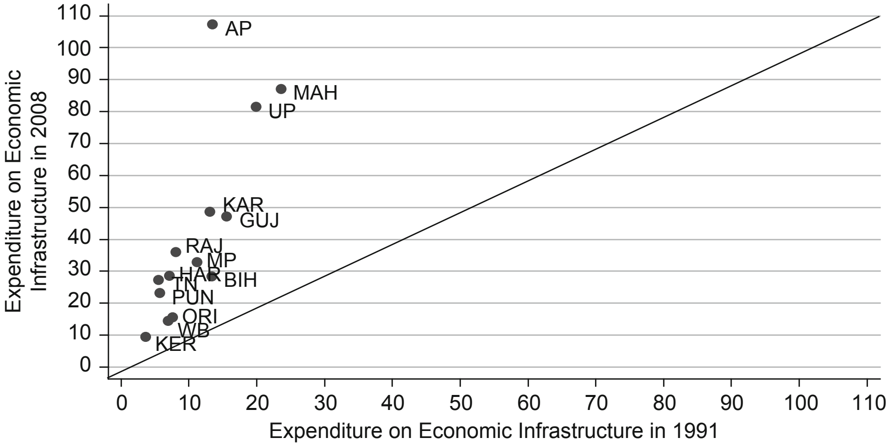


Fig. 19.2 State of economic infrastructure expenditure. (Source: Authors' calculation based on data from State Finances: A Study of Budgets, RBI)

in achieving the Millennium Development Goals (MDG) by 2015. However, one needs to recognise that initial conditions in India were quite low and, hence, there is a need for further increase in social sector expenditure.

In the case of economic infrastructure, we find similar results where the situation has improved for all the states; but, if we look at the distribution of states, majority of the states are close to the line of equality and are near the origin which clearly suggests that the situation has not improved significantly over the reform period. It is only for the states of Andhra Pradesh, Maharashtra and Uttar Pradesh that we see a substantial change in the expenditure towards economic infrastructure heads; Karnataka and Gujarat have also shown improvement but it is only to some extent (see Fig. 19.2).

It has been perceived by many that 'infrastructure deficit' would be a major deterrent for achieving high economic growth in India. However, in our view, the 'infrastructure inequality' is more important, which could be a bottleneck for balanced regional development, which is the core objective of India's federalism.

Figurs 19.3a, d and 19.4a, d draw upon the scatter plot with a simple regression line of fit for social infrastructure (RSOC) and real SDP and economic infrastructure (RECO) and real SDP, based on respective variable average values for the periods 1991–1995, 1996–2000, 2001–2005 and 2006–2008, respectively. These plots can be used to understand correlation between infrastructure and state output temporally. It is clearly visible from Fig. 19.3a, d that correlation between social infrastructure and state output has remained high for all the periods of study. Even the coefficient of the simple regression used as the line of fit in the figure connotes the same. The correlation coefficient between social infrastructure and state output varied between 87 and 94 %. The linkage between economic infrastructure and state performance on output appears to be not so robust (see R^2 values in Figs. 19.3 and 19.4). However, one

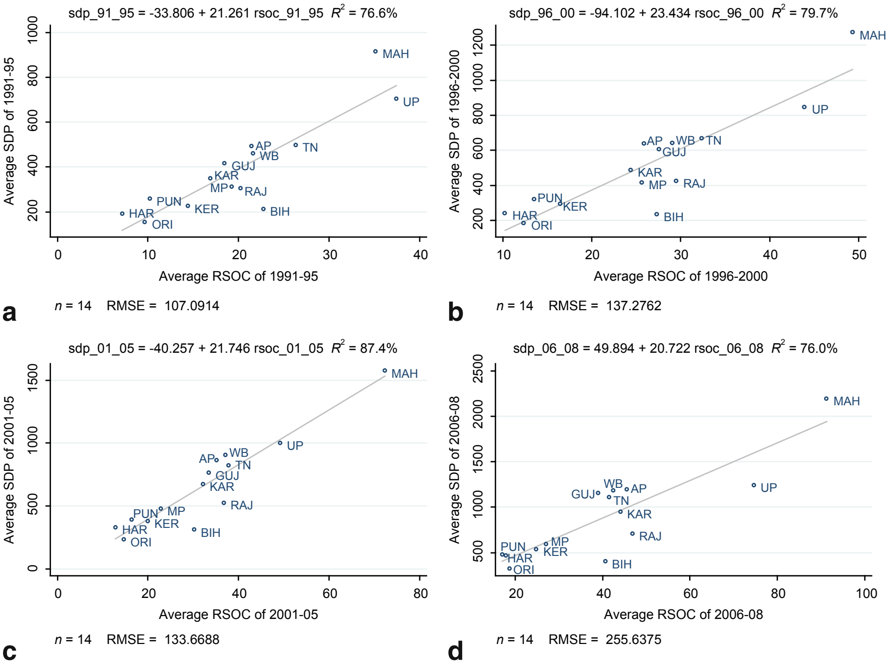


Fig. 19.3 Interaction of social infrastructure and state output over different time horizons. (Source: Authors' calculation based on State Finances: A Study of Budgets and National Income Accounts and Statistics, CSO data)

thing that is evident from these plots is that both social and economic infrastructure are highly correlated to state output.

It is clearly visible from the line plot of RSOC and real SDP that, except for Bihar and few other states, there is strong and positive link in social infrastructure (RSOC) and real SDP (see Fig. 19.5a). This linkage has even improved in the following years as evident from the plot for subsequent years (see Fig. 19.5b, d). The line plot of RECO and SDP (Fig. 19.6) indicates a similar strong comovement. Comparing the line plots for different years in Figs. 19.5 and 19.6 indicates that the effect of social infrastructure on real SDP is even higher in comparison to that of economic infrastructure. Results of correlation coefficient also reveal the importance of RSOC over economic infrastructure (the correlation coefficient for RSOC and SDP lies between 0.87 and 0.94, whereas, for RECO and SDP, it remained in the range of 0.72–0.80). It is important to note that we have exercised a similar scatter plot and line plot for RSCO, RECO and SDP with individual yearly data of 1990, 1995, 2000, 2005 and 2008 and the pattern of the results is quite similar.[7]

This preliminary analysis shows that infrastructure is crucial for the growth in the regions. In the post-reform period, there is improvement in all the indicators, but

[7] To conserve space, we have not reported the individual year plot figures.

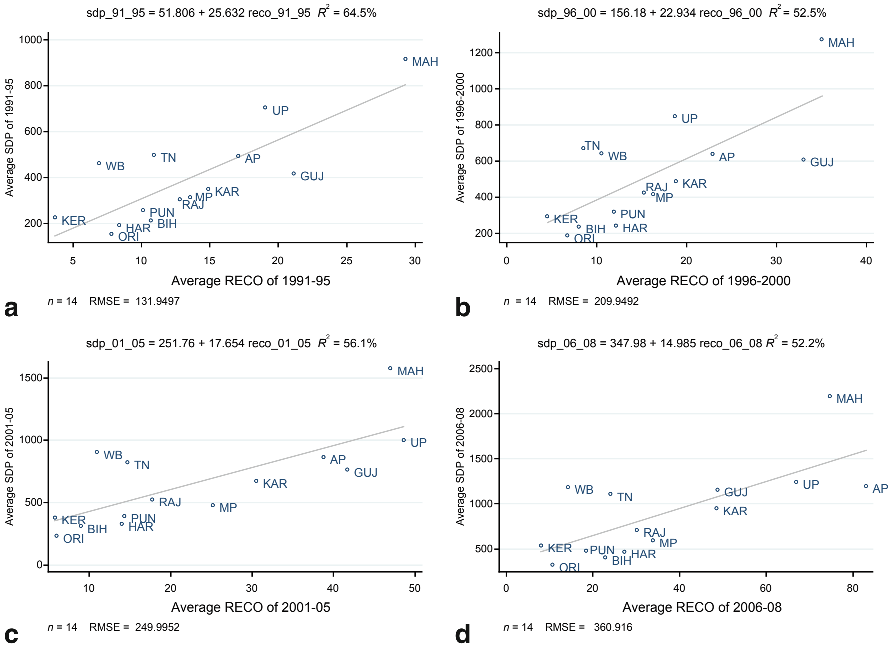


Fig. 19.4 Interaction of economic infrastructure and state output over different time horizons. (Source: Authors' calculation based on State Finances: A Study of Budgets and National Income Accounts and Statistics, CSO data)

lower coefficients and correlation values for the period of 2006–2008 also indicate that we are not isolated from the external economy effect and any policy prescription of inclusive growth should not be made in isolation to the external sector. However, for these conclusions to be robust, we undertake panel estimation procedures such as cointegration and causality exercises and they are discussed in the methodology section. In the literature, it is found that the impact of infrastructure on growth is generally examined through estimating the direction of TFP. However here, as we are focusing specifically on two inputs, we undertake the impact analysis through panel econometrics.

6 Methodology, Econometric Results and Discussion

A sufficiently large time component in panel data introduces the problem of spurious regression if the series are found to be non-stationary in nature. With recent advancement in panel data estimation methodology, it is possible to address the issue of presence of non-stationarity and further carryout the cointegration in case the series are non-stationary. Thus, as a prerequisite requirement, to test the order of integration of individual series, the Im, Pesaran and Shin (IPS) unit root test is employed (see the

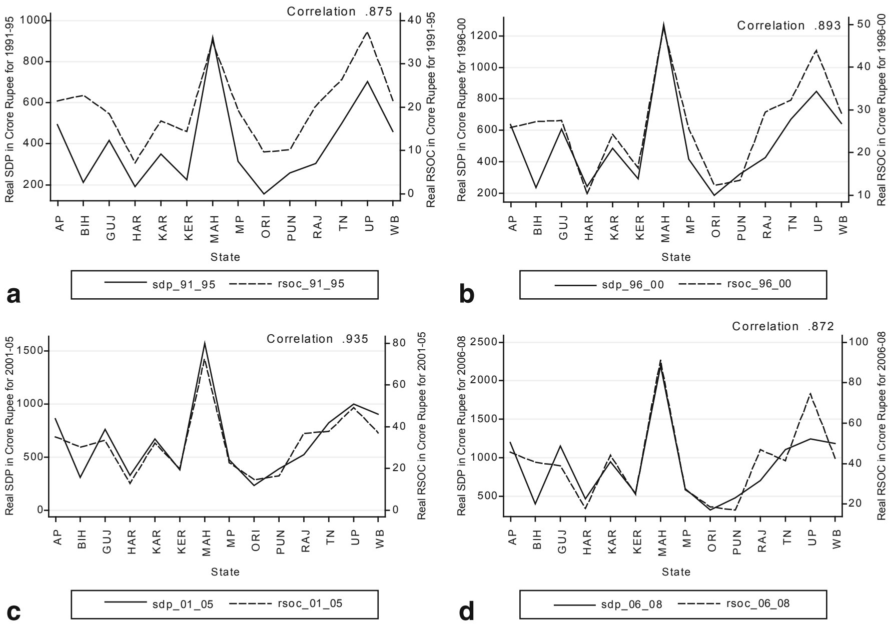


Fig. 19.5 Comovement of social infrastructure and SDP. (Source: Authors calculation based on State Finances: A Study of Budgets and National Income Accounts and Statistics, CSO data)

appendix for IPS panel unit root methodology). The basic reason to prefer IPS unit root test over Levin and Lin (1992)/Levin et al. (2002) or other panel unit root test is linked to the ability of the IPS test to allow for serial correlation and heterogeneity in the error term while accommodating an individual–specific deterministic trend but cross-sectionally independent, apart from the presence of the common time effects. The IPS procedure allows heterogeneity in the short-run dynamics, in the error structure, in the form of fixed effects and linear trend coefficients. Table 19.1 presents the results of the IPS panel unit root test. Test statistics of none of the variables at level reject the null hypothesis of unit root, suggesting the presence of non-stationarity in the individual series. Rejection of null hypothesis at the first difference series for all the variables confirms that SDP, RSOC and RECO are non-stationary at level but are difference stationary.

Once it is established that all the variables are integrated with the same order and are I(1), the appropriate way forward is to test the presence of a long-run relationship among variables using a panel cointegration test. Of the available panel cointegration tests, the Pedroni (2004) panel cointegration test is employed to test the possibility of a long-run relationship (testing null of cointegration) among the variables. Among the available tests for testing the presence of cointegration in panel data, the Pedroni panel cointegration test is the preferred choice given its ability to allow heterogeneity in intercept, trend and slope coefficient to vary across the individual series. The

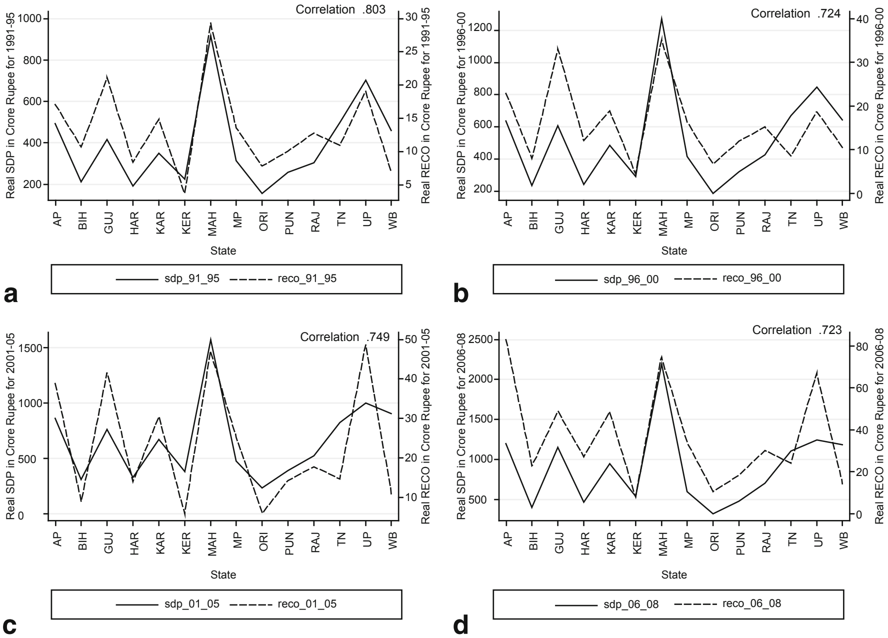


Fig. 19.6 Comovement of social infrastructure and SDP. (Source: Authors calculation based on State Finances: A Study of Budgets and National Income Accounts and Statistics, CSO data)

Table 19.1 IPS panel unit root test result

Variables	Period	Number of states	Test statistic	
			Level	1st Difference
SDP	1986–2008	14	0.520	– 3.595[a]
RSOC	1986–2008	14	– 1.509	– 3.969[a]
RECO	1986–2008	14	– 2.125	– 5.457[a]

Based on the Im–Pesaran–Shin (2003) method. At level trend and constant both are included but for the first-difference series only the constant is included while testing the unit null in the series
[a] Indicates significant at 1 % level of significance

Pedroni cointegration test also considers a provision for regressors to be fully endogenous. Since both the dynamics and the cointegrating vector itself are permitted to vary across individual members of the panel, we use four within-group tests and three between-group tests statistics. The within-dimension tests impose a common coefficient under the alternative hypothesis, whereas, between-dimension (or 'group') tests allow for heterogeneous coefficients under the alternative hypothesis. Further, three different specifications are allowed with respect to the deterministic trend and year fixed effect. These specifications are: (1) only trend is assumed in the cointegrating equation, (2) both trend and year fixed effect are present in the cointegrating equation and finally, (3) only the year fixed effect is allowed in the cointegrating equation. The

Table 19.2 Pedroni panel cointegration test result

Test	Trend	Time dummy	Trend and time dummy
Panel ν–stat	6.231[a]	0.733	4.812[a]
Panel ρ–stat	– 1.818[a]	– 1.990[a]	– 0.619
Panel PP–stat	– 6.011[a]	– 3.361[a]	– 2.855[a]
Panel ADF–stat	– 5.319[a]	– 2.044[a]	– 1.876[a]
Group ρ–stat	0.082	– 1.218	0.683
Group PP–stat	– 4.209[a]	– 3.777[a]	– 2.039[a]
Group ADF–stat	– 2.282[a]	– 2.022[a]	– 0.601

For ν–Stat 5 % critical value is 1.645 and for rest of the tests it is – 1.645

[a] Indicates significant at 5 level of significance

Table 19.3 Pedroni panel group: mean FMOLS result

VARIABLES	With time dummy Coefficient (t–statistic)	Without time dummy Coefficient (t–statistic)
RSOC	8.11[a] (15.78)	19.99[a] (32.00)
RECO	3.24[a] (5.29)	4.52[a] (10.51)

[a] Indicates significant at 1 % level of significance

Pedroni panel cointegration test results are presented in Table 19.2. The test result of panel cointegration reveals that the assumption made about presence of; trend, trend and year fixed effect, year fixed effect in the cointegrating equation adds to the sensitivity of rejecting the null of cointegration. However even then, by and large, for all the three specifications linked with presence of trend and year fixed effect significantly rejects the null of no cointegration for majority of the within-group and between-group tests statistics, thus, establishes existence of a long-run relationship among SDP, RSOC and RECO.

Given that GSDP, RECO and RSOC are cointegrated, we further employ the Pedroni fully modified ordinary least squares (FMOLS) estimator in order to estimate the long-run relationship. The between–group FMOLS–based estimator permits for a more flexible alternative hypothesis and suffers much less from small sample size distortion than the within–group estimator.[8] Moreover, the Pedroni FMOLS technique takes care of possible endogeneity and simultaneity of the estimators. We have estimated two Pedroni between–group FMOLS equations assuming the presence of time effect to take care of any time fixed effect in estimation and then without the time fixed effect. The results of the FMOLS estimation with assumption of time dummy and without time dummy are displayed in Table 19.3. The result of the FMOLS estimation reveals that the long–run coefficient of both the variables is positive and significant at the conventional level of significance. Thus, panel FMOLS results

[8] Pedroni 2001) proposes two methods to apply this fully modified method to panel cointegration regression: the pooled (or within–group) panel FMOLS estimator and the group–mean (between–group) FMOLS estimator. Pedroni (2001) using Monte Carlo simulation shows that of the between–group and within–group estimators, the between-group estimator has a much smaller size distortion, whereas both are unbiased.

Table 19.4 Panel Granger causality test results

Dependent Variable	Short run (independent variables)			Long–Run
	$\sum_{j=1}^{p} \Delta SGDP_{t-j}$	$\sum_{j=1}^{p} \Delta RECO_{t-j}$	$\sum_{j=1}^{p} \Delta RSOC_{t-j}$	Lagged ECM term
	$\sum \beta_i = 0$: F–stat (p–value)	$\sum \beta_i = 0$: F–stat (p–value)	$\sum \beta_i = 0$: F–stat (p–value)	$\Theta = 0$: t–stat (p–value)
$\Delta SGDP$	–	6.88[a] (0.00)	7.80[a] (0.00)	– 0.063[a] (0.01)
$\Delta RECO$	15.45[a] (0.01)	–	0.12	– 0.045
$\Delta RECO$	36.90[a] (0.01)	1.94	–	– 0.024

[a] Indicates significant at 10 % level of significance

invariably endorse the importance of public investment in different heads of social and physical infrastructure. Additionally, the large coefficient of social infrastructure compared to that of economic infrastructure, irrespective of the assumption about time dummy, depicts the higher importance of social infrastructure in the economic development of the states.

Given that the variables are cointegrated, we now turn to analyse the direction of causality among GSDP, RECO and RSOC. In the presence of cointegration, the Granger causality is based on the vector error correction model (VECM) procedure. Table 19.4 presents the result of the Granger causality test. The first row of the Table 19.4, with $\Delta SGDP$ as the dependent variable, tests the null hypothesis that state output is not caused by lags of $\Delta RECO$ and $\Delta RSOC$ in the short run or by the ECM term which tests long-run causality from social infrastructure and economic infrastructure to state output. It is seen from the table that the coefficients of social infrastructure, economic infrastructure and lagged ECM term are significant at 1 % level. This then rejects the null of no causality from economic and social infrastructure to state output. Similarly, the second row of the table tests causality from output and social infrastructure to economic infrastructure. As the coefficients of social infrastructure and lagged error term are not significant but the coefficient of state output is significant at 1 % level of significance, this implies it is only state output which causes economic infrastructure. Similarly, if we look at the third row of the table it is only state output which causes social infrastructure.

From the causality results, one can confirm that although there is a bidirectional causality between output and infrastructure variables, it is the infrastructure that has strong causation with output growth and not vice versa, at least in the long run. In the short run, output growth continues to be a driver of infrastructure in the Indian states.

7 Conclusions

The positive and significant role of infrastructure (both physical and social) in the process of growth has been highlighted very well in theoretical literature. Infrastructure has been considered as a strong supply-side factor that improves productivity (both labour and capital) both in the long run as well as in the short run. However, at the empirical level, the findings are mixed particularly in the case of developed economies. For developing economies, such as India, where the infrastructure deficit is quite large and they are faced with severe fiscal constraints, the role of infrastructure could be quite substantial. Further, one can hypothesize that the growth divergence in India, which is very well established at least in the post-reform period, could also be due to the variation in the level and stock of infrastructural facilities at the state level. Towards this end, in this chapter, an attempt has been made to understand the role of infrastructure in regional economic growth and divergence in India. Based on panel data cointegration models for 15 major Indian states, the study finds that there is a high positive correlation between infrastructure expenditure with that of economic activity. Of the two, social infrastructure appears to be highly correlated with growth.

The analysis suggests that there exists a long-run cointegrating relationship between level of infrastructure development and economic output. The long-run coefficients, after adjusting for time, suggest that return to expenditure on social infrastructure is higher compared to physical infrastructure expenditure. The causality results show that there exist bidirectional causality between infrastructure and economic growth. This indicates that for the development of social sector, increased economic activity is also necessary.

8 Appendix IPS Panel Unit Root Test

The IPS panel unit root test uses the principle of augmented Dickey–Fuller (ADF) unit root test:

$$\Delta Y_{it} = \alpha_i + \beta_i t + \rho_i Y_{it-1} + \varepsilon_{it} \tag{19.1}$$

where, $i = (i = 1,\ 2,\ 3 \ldots \text{n})$ represents cross-section units like country, state, or firm and t ($\text{t} = 1,\ 2,\ 3 \ldots .\text{T}$) represents the time period of the observation. The error term ε_{it} is i.i.d. α_i and β_i allow for fixed and unit specific time trends for each i. The null hypothesis of a unit root is $|\rho_i| = 0$ against the alternative of $\rho_i < 0, i = 1....N_1$ and $\rho_i = 0, i = N_1 + 1.....N$. The IPS test averages the ADF individual unit root test statistics t_p which are obtained from estimating the equation 1 for each i (allowing each series to have different lag length if necessary); the test statistics is calculated as

$$\bar{t} = \tfrac{1}{N} \sum_{i}^{N} t_\rho \tag{19.2}$$

as $T \to \infty$ (for a fixed value of N) followed by $N \to \infty$ sequentially, IPS test statistics is standard normal distribution.

9 Pedroni (residual based) Cointegration Tests

The Pedroni (2004) panel cointegration in spirit extends the Engle–Granger (1987) residual-based methodology to examine the existence of long-run (cointegrating) relationship in the panel data framework. The Pedroni cointegration tests are based on estimating the static cointegration regression given by

$$Y_{it} = \alpha_i + \beta_i t + \beta_{1i} X_{1it} + \beta_{2i} X_{2it} \ldots . \beta_{mi} X_{mit} + \varepsilon_{it} \tag{19.3}$$

Where $i = 1, \ldots, N$, $t = 1, \ldots, T$ and $m = 1, \ldots, M$. $\varepsilon i$is the autoregressive term of the form $\hat{\varepsilon}_{it} = \rho_i \hat{\varepsilon}_{it-1} + u_{it}$ and X_{mit} are the observable variables with dimension of $i{*}t$. The parameters $\alpha_i$it and $\beta_i$allow for the possibility of state-specific fixed effects and deterministic trends, respectively.

Pedroni (2004) proposes several tests for cointegration that allow for heterogeneous intercepts and trend coefficients across cross sections. Pedroni considers the following panel unit root test to test the null hypothesis of no cointegration, $\rho_i = 1$:

$$\hat{\varepsilon}_{it} = \rho_i \hat{\varepsilon}_{it-1} + u_{it}$$

The test statistics of the within-dimension approach are panel ν–statistic, panel ρ–statistic, panel PP–statistic and panel ADF–statistic. These statistics pool the autoregressive coefficients across different members for the unit root tests on the estimated residuals. The three test statistics of the between-dimension approach are group ρ–statistic, group PP–statistic[9] and group ADF–statistic. These statistics are based on estimators that simply average the individually estimated coefficients for each member. The heterogeneous panel and heterogeneous group mean panel cointegration test statistics are calculated as follows:

Panel ν –statistic:

$$Z_\nu = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11i}^{-2} \hat{\varepsilon}_{it-1}^{2} \right)^{-1}$$

Panel ρ–statistic:

$$Z_\rho = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11i}^{-2} \hat{\varepsilon}_{it-1}^{2} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11i}^{-2} \left(\hat{\varepsilon}_{it-1} \Delta \hat{\varepsilon}_{it} - \hat{\lambda}_i \right) \right)$$

Panel PP–statistic:

$$Z_t = \hat{\sigma}^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11i}^{-2} \hat{\varepsilon}_{it-1}^{2} \right)^{-1/2} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{L}_{11i}^{-2} \left(\hat{\varepsilon}_{it-1} \Delta \hat{\varepsilon}_{it} - \hat{\lambda}_i \right) \right)$$

[9] PP tests are likely to be more robust to fat tails in data.

Panel ADF–statistic:

$$Z^*{}_t = \hat{\sigma}^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\hat{\varepsilon}_{it-1}^{*2}\right)^{-1/2}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{L}_{11i}^{-2}\left(\hat{\varepsilon}_{it-1}^{*}\Delta\hat{\varepsilon}_{it}^{*}\right)\right)$$

Group ν–statistic:

$$\tilde{Z}_t = \sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{it-1}^{2}\right)^{-1/2}\left(\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it-1}\Delta\hat{\varepsilon}_{it}-\hat{\lambda}_i\right)\right).$$

Group PP–statistic:

$$\tilde{Z}_t = \sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{\varepsilon}_{it-1}^{2}\right)^{-1/2}\left(\sum_{t=1}^{T}(\hat{\varepsilon}_{it-1}\Delta\hat{\varepsilon}_{it}-\hat{\lambda}_i)\right)$$

Group ADF–statistic:

$$\tilde{Z}_t^* = \sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{s}_i^2\hat{\varepsilon}_{it-1}^{*2}\right)^{-1/2}\left(\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it-1}^{*}\Delta\hat{\varepsilon}_{it}^{*}\right)\right)$$

Here, $\hat{\varepsilon}_{it}$ is the estimated residual and $\hat{L}_{11i}^{2}$ is the estimated long-run covariance matrix for $\Delta\hat{\varepsilon}_{it}$. Similarly, $\hat{\sigma}_i$ and $\hat{s}_i(\hat{s}_i^{*2})$ are, respectively, the long-run and contemporaneous variances for individual i (cross section). Pedroni (1999) discusses these issues in detail with the appropriate lag length determined by the Newey–West method. Pedroni (1997, 1999) has shown that all seven test distributions follow standard normal asymptotically as

$$\frac{\chi_{N,T}-\mu\sqrt{N}}{\sqrt{\nu}} \rightarrow N(0,1)$$

Where $\chi_{N,T}$ is the standardised form of for each seven statistics, while μ and ν are the mean and variance of the underlying series.

The panel ν–statistic is a one-sided test where large positive values reject the null of no cointegration. The remaining statistics diverge to negative infinitely, which means that large negative values reject the null. The critical values are also tabulated by Pedroni (1999).

10 Panel FMOLS

Pedroni (2001) FMOLS is a non-parametric estimation technique which transforms the residuals from the cointegration regression and thus gets rid of serial correlation. Therefore, the problem of endogeneity of the regressors and serial correlation in the error term is avoided by using FMOLS. Group-mean FMOLS estimators have relatively minor size dissertation in small samples. Additionally, it allows for heterogeneity across the cross section. To understand the correction of endogeneity and serial correction in FMOLS, let us consider a panel model of two variables:

$$Y_{it} = \alpha_i + \beta_i X_{it} + u_{it} \tag{19.4}$$

The OLS estimate of the coefficient β_i in panel regression is given by

$$\hat{\beta}_{i,OLS} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)^2 \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)(Y_{it} - \bar{Y}_i)$$

where, $\bar{X}_i$ and $\bar{Y}$ refer to the individual means of each i cross section. This estimator is asymptotically biased and its distribution is dependent on the nuisance parameter (Pedroni 2001). To correct for endogeneity and serial collation, Pedroni (2001) has suggested for group-mean FMOLS estimator that incorporates the Phillips and Hansen (1990) semiparametric correction to the OLS estimator to eliminate the bias due to the endogeneity of the regressors. The authors also adjust for the heterogeneity that is present in the dynamics underlying X and Y. The FMOLS statistic is

$$\hat{\beta}_{i,FOLS} = N^{-1} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} (X_{it} - \bar{X}_i)^2 \right)^{-1} \left(\sum_{t=1}^{T} (X_{it} - \bar{X}_i) Y^{*}{}_{it} - T\hat{Y}_i \right)$$

where

$$Y_{it}^{*} = (X_{it} - \bar{X}_i) - \frac{\hat{\Omega}_{21i}}{\hat{\Omega}_{22i}} \Delta X_{it}$$

$$\hat{\gamma} = \hat{\Gamma}_{21i} + \hat{\Omega}_{21i}^{0} - \frac{\hat{\Omega}_{21i}}{\hat{\Omega}_{22i}} \left(\hat{\Gamma}_{22i} + \hat{\Omega}_{22i}^{0} \right)$$

where $\hat{\Omega}$ and $\hat{\Gamma}$ are respectively the covariance and sum of autocovariances obtained from the long-run covariance matrix from the model.

11 Panel Causality

To test for panel causality, the most widely used method in the literature is that proposed by Holtz–Eakin et al. (1988 and 1989). Their time-stationary vector autoregressive (VAR) model is of the form:

$$Y_{it} = \alpha_0 + \sum_{j=1}^{m} \alpha_j Y_{it-j} + \sum_{j=1}^{m} \delta_j X_{it-j} + f_{yi} + u_{it} \tag{19.5}$$

$$X_{it} = \beta_0 + \sum_{j=1}^{m} \beta_j Y_{it-j} + \sum_{j=1}^{m} \gamma_j X_{it-j} + f_{xi} + \nu_{it} \tag{19.6}$$

Where Y_{it} and X_{it} are the two cointegrated variables, $i = 1, \ldots,$ N represents cross-sectional panel members and u_{it} and v_{it} are error terms. This model differs from the standard causality model in that it adds two terms, f_{xi} and f_{yi}, which are individual fixed effects for the panel member i.

In the equations mentioned earlier, the lagged dependent variables are correlated with the error terms, including the fixed effects. Hence, OLS estimates of the earlier mentioned model will be biased. The remedy is to remove the fixed effects by differencing. The resulting model is

$$\Delta Y_{it} = \sum_{j=1}^{m} \alpha_j \Delta Y_{it-j} + \sum_{j=1}^{m} \delta_j \Delta X_{it-j} + \Delta u_{it} \tag{19.7}$$

$$\Delta X_{it} = \sum_{j=1}^{m} \beta_j \Delta Y_{it-j} + \sum_{j=1}^{m} \gamma_j \Delta X_{it-j} + \Delta v_{it}. \tag{19.8}$$

However, differencing introduces a simultaneity problem because lagged endogenous variables on the right-hand side are correlated with the new differenced error term. In addition, heteroscedasticity is expected to be present because, in the panel data, heterogeneous errors might exist with different panel members. To deal with these problems, an instrumental variable procedure is traditionally used in estimating the model, which produces consistent estimates of the parameters.

Assuming that u_{it} and v_{it} are serially uncorrelated, the second or more lagged values of Y_{it} and X_{it} may be used as instruments in the instrumental variable estimation (Easterly et al. 1997). Then, to test for the causality, the joint hypotheses $\delta_j = 0 for j = 1, \ldots.., m$ and $\beta_j = 0 for j = 1, \ldots.., m$ are simply tested.

The test statistics follow a chi–squared distribution with (k–m) degrees of freedom. The variable X is said not to Granger–cause the variable Y if all the coefficients of lagged X in the equation (19.7) are not significantly different from zero, which implies that the history of X does not improve the prediction of Y.

References

Agarwalla A (2011) Estimating the contribution of infrastructure in regional productivity growth in India. Indian institute of management, Ahmedabad, working paper series number 2011–05–01

Aggarwal A, Singh P (2013) Regional development policy in India. In: Farole T (eds) The internal geography of trade: lagging regions and global markets. The World Bank, Washington DC, pp 229–255

Ahluwalia MS (2000) economic performance of states in post–reforms period. Econ Polit Wkly 35(19):1637–1648

Ahluwalia MS (2002) State level performance under economic reforms in India. In: Krueger AO (eds) Economic policy reforms and the Indian economy. Oxford University Press, New Delhi, pp 91–125

Argy F, Lindfield M, Stimson B, Hollingsworth P (1999) Infrastructure and economic development. CEDA Background Paper

Aschauer DA (1989a) Is public expenditure productive? J Monet Econ 23(2):177–200

Aschauer DA (1989b) Public investment and productivity growth in the group of seven. Econ Perspect 13:17–25

Aschauer DA (1989c) Does public capital crowd out private capital? J Monet Econ 24(2):171–188

Bhanumurthy NR, Singh P (2013) Financial sector development and economic growth in Indian states. Inter J Econom Policy Emer Econom, (forthcoming)

Binswanger HP, Khandkur SR, Rosenzweig MR. (1989) How infrastructure and financial institutions affect agriculture output and investment in India. Policy planning and research working paper No. 163

Bougheas S, Demetriades PO, Mamuneas TP (2000) Infrastructure, specialization, and economic growth. Can J Econ 33(2):506–522

Canning D, Fay M (1993) The effect of transportation networks on economic growth. Colombia University Working Paper, New York

Canning D, Pedroni P (2004). The effect of infrastructure on long run economic growth. Mimeo

Devarajan S, Swaroop V, Zou HF (1996) The composition of public expenditure and economic growth. J Monet Econ 37(2):313–344

Dutt G, Ravallion M (1998) Why have some Indian states done better than Others at reducing rural poverty. Economica 65(257):17–38

Dutta J, Nicholas H, Nicholas V (2007). Industrial activity in Indian states: The role of infrastructure. http://mpra.ub.uni-muenchen.de/7184/

Easterly W, Rebelo S (1993) Fiscal policy and economic growth: an empirical investigation. J Monet Econ 32(3):417–458

Easterly W, Loayza N, Montiel P (1997) Has Latin America's post-reform growth been disappointing? Journal of International Economics 43(3–4): 287–311

Elhance AP, Lakshamanan TR (1988) Infrastructure–production system dynamics in national and regional systems: an economic study of the Indian economy. Reg Sci Urban Econ 18(4):511–531

Engle RF, Granger CWJ (1987) Cointegration and error correction: representation, estimation and testing. Econometrica 55(2):251–276

Evans P, Karras G (1994) Is government capital productive? evidence from a panel of seven countries. J Macroecon 16(2):271–279

Garcia-Milà T, McGuire TJ (1992) The contribution of publicly provided inputs to states' economies. Reg Sci Urban Econ 22(2):229–241

Garcia–Milà T, McGuire TJ, Porter RH (1996) The effect of public capital in state level production functions reconsidered. Rev Econ Statist 78(1):177–180

Ghosh B, De P (1998). Role of infrastructure in regional development: a study over the plan period. Economic and political weekly, 33(47/48):3039–3043 + 3045–3048

Gramlich EM (1994) Infrastructure investment: a review essay. J Econ Lit 32(3):1176–1196

Granger CWJ, Newbold P (1974) Spurious regressions in econometrics. J Econometrics 2(2): 111–120

Holtz–Eakin D (1994) Public sector capital and the productivity puzzle. Rev Econ Statist 76(1): 12–21

Holtz-Eakin D, Schwartz AE (1995) Infrastructure in a structural model of economic growth. Reg Sci Urban Econ 25(2):131–151

Holtz–Eakin D, Newey W, Harvey SR (1988) Estimating vector autoregressions with panel data. Econometrica 56(6):1371–1395

Holtz–Eakin D, Newey W, Harvey SR (1989) The revenues–expenditures nexus: evidence from local government data. Int Econ Rev 30(2):415–429

Hulten CR, Bennathan E, Srinivasan S (2006) Infrastructure, externalities, and economic development: a study of Indian manufacturing industry. The World Bank Economic Review 20(2):291–308

Hulten CR, Schwab RM (1991) Public capital formation and the growth of regional manufacturing industries. Nat Tax J 45(4):121–134

Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. J Econometrics 115(1):53–74

Levin A, Lin CF (1992). Unit root test in panel data: asymptotic and finite sample properties. University of California at San Diego, Discussion paper 92–93

Levin A, Lin CF, Chu CSJ (2002) Unit root tests in panel data: asymptotic and finite–sample properties. J Econometrics 108(1):1–24

Majumder R, Mukherjee D (2005) Infrastructure and Development Interlinkage in West Bengal: A VAR Analysis. http://mpra.ub.uni-muenchen.de/4820/

Munnell AH. (1990). Why has productivity growth declined: productivity and public investment? New England Economic Review, pp. 2–33

Nagaraj R, Varoudakis A, Véganzonès MA (2000) Long–run growth trends and convergence across Indian states. J Int Devel 12(1):45–70

O'Fallon C (2003) Linkages between infrastructure and economic growth, New Zealand Ministry of Economic Development (www.med.govt.nz)

Patra AK, Acharya A (2011) Regional disparity, infrastructure development and economic growth: An inter–State analysis. Research Practice in Social Sciences 6(2):17–30

Pedroni P (1997) Panel cointegration; asymptotic and finite sample properties of pooled time series tests, with application to the PPP hypothesis: New Results. Working Paper, Department of Economics, Indiana University

Pedroni P (1999) Critical values for cointegration tests in heterogeneous panels with multiple regressors. Oxford Bull Econ Statist 61(0):653–670

Pedroni P (2001) Fully modified OLS for heterogeneous cointegrated panels. In Badi H. Baltagi, Thomas B. Fomby, R. Carter Hill (eds) Nonstationary panels, panel cointegration, and dynamic panels, advances in econometrics, Vol. 15, Emerald Group Publishing Limited, pp. 93–130

Pedroni P (2001) Purchasing power parity tests in conintegrated panel. Review Econom Stat 83(4):727–731

Pedroni P (2004) Panel cointegration; asymptotic and finite sample properties of pooled time series tests, with an application to the PPP hypothesis. Econometric Theory 20(3):597–625

Phillips PCB, Hansen BE (1990) Statistical inference in instrumental variables regression with i(1) processes. Rev Econ Stud 57(1):99–125

Pritchett L (1996) Mind your P's and Q's. World Bank Policy Research paper, No. 1660. The World Bank, Washington DC

Rao MG, Shand RT, Kalirajan KP (1999) Convergence of incomes across Indian states: A divergent view. Economic Political Weekly 34(13):769–778

Rodríguez F (2007) Have collapses in infrastructure spending led to cross–country divergence in per capita GDP? Working papers 52,United Nations, Department of Economics and Social Affairs

Sahoo S, Saxena KK (1999) Infrastructure and economic development: some empirical evidence. The Indian Economic Journal 47(2):54–66

Sanchez–Robles B (1998) Infrastructure investment and growth: some empirical evidence. Contemporary Econ Pol 16(1):98–108

Shand RT, Bhide S (2000) Sources of economic growth—regional dimensions of reforms. Economic Political Weekly 35(42):3747–3757

Tewari RT (1984) Economic infrastructure and regional development in India. Man Development 6(4):63–80

Uchimura K, Gao H (1993) The importance of infrastructure on economic development. Latin America and Caribbean Regional Office, World Bank, Washington DC

Chapter 20
The Phenomenon of Wasted Vote in the Parliamentary Elections of India

Sanmitra Ghosh

1 Introduction

One widely known law in political science, known as Duverger's law (Duverger 1954), states that the top two parties in an electoral district should share all the votes and nothing should be left for a third party. The law, which is a statement about single-member district plurality systems, predicts that there should be no "wasted vote" even within a large electorate.

The theoretical rationale underlying Duverger's law is based on expected utility maximization. The first formalization of the model is due to Mckelvey and Ordeshook (1972). They demonstrate that in multiparty elections, a voter might be willing to vote for his/her second or even lower ranked candidates if his/her most preferred party has a poor *viability* in the district and the second- or third-ranked party is a close contender for the seat. Consequently, the supporters of all the candidates other than those of the two most viable parties should switch their support to the more preferred of the top two candidates, and the vote shares of all except these two should drop to zero .

The expected utility maximization hypothesis implicitly assumes that the voter calculates the probability of a tie between pairs of candidates for the first place, as he/she becomes *pivotal* only in the case of such a tie. The question, however, arises that in reality to what extent the voter's motivation to choose a certain candidate could be attributed to the phenomenon of pivotal voter, as in a large electorate the probability of such an event should be extremely low. True enough, in the real world, very few voters actually think of breaking a tie while casting their vote. But as Abramson et al. (1992) observe "Like all theories, the calculus of voting is a simplification of reality that seeks to capture the most salient features of the actual situations. Many voters may see some candidates as having real chances of winning and others as likely losers, and they may weigh these perceptions against the relative attractiveness of the candidates."

S. Ghosh (✉)
Department of Economics, Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: sanmitraz@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_20,

Subsequently, a number of distinct theoretical models have analyzed the consequences of strategic voting in single-member district plurality systems. Using the framework of Bayesian games, Cox (1987) demonstrates that strategic voting works to improve the winning chances of candidates who are expected to perform well. As an extension of this theory, Palfrey (1989) proves that as the size of the electorate grows, *every* voting equilibrium resembles Duvergerian equilibrium in the limit.

Myerson and Weber (1993) construct a counterexample of the Palfrey model using a somewhat different framework. The authors consider a model in which three political parties contest an election. They demonstrate that only under special conditions can positive votes be received by *all* three parties—indeed, the second and third parties must be *very* close in terms of the probabilities of winning the election. Under such circumstances, the coordination on one of these candidates becomes very difficult and, consequently, the votes might be split between the two. Myerson and Weber have shown that non-Duvergerian three-party equilibria are robust to small perturbations in the perceived ex ante winning chances of the candidates. However, Fey (1997) shows that the non-Duvergerian equilibria are dynamically unstable if learning on the part of the voters from past election data and opinion polls is incorporated into the model.

There have been several empirical studies to measure the extent of strategic voting in an electorate. The first of this kind are the studies by Cain (1978) and Black (1978). Subsequent authors like Curtice and Steed (1988), Galbraith and Rae (1989), and Johnston and Pattie (1991) have compared the level of support obtained by a party in two consecutive elections to estimate the level of strategic voting. They use either the switch in support from the incumbent to the (major) opposition party or an increment in the vote share of the latter between pairs of elections as a measure of strategic voting in an electorate. Niemi et al. (1992) and Evans and Heath (1993) have used "self-reporting intention" technique to identify the strategic voters from a sample of survey respondents. Abramson et al. (1992) and Alvarez and Nagler (2000) estimate strategic voting by measuring the objective difference between the stated vote and the preference ranking of the candidates on a feeling thermometer scale.

There has been relatively less empirical work on strategic voting and party systems in India. A notable early exception is the study by Chhibber and Kollman (1998), which makes a comparison of the party systems in India and the USA. They test the prediction of Duverger's law at the national level and find that fiscal centralization leads to a two-party system in the same through coordination among parties on policy platforms. Chhibber and Murali (2006) note that in Indian state assembly elections exceptions to Duverger's law at the constituency level are attributable to the federal structure of Indian polity, where voters often have to take both district level and national competitiveness of parties into consideration. On the other hand, in a recent work, Dunleavy and Diwakar (2011) emphatically argue that Duverger's law does not seem to hold in any of the single-member district plurality systems outside the USA, including India.

In this chapter, we test the validity of the predictions of Duverger's law at the constituency[1] level, namely that the volume of wasted votes should be zero at least in

[1] In India, electoral districts are called constituencies.

large elections. We use data from Indian parliamentary (*Loksabha*) elections between 1967 and 1998. We exploit constituency-level data on the votes received by each party in each of the 15 major states in India. Even though the volume of the wasted votes appears to be quite large, we find that much of it is *ineffective* in the sense that even when it is transferred to the second candidate en bloc, it does not alter the outcome of the election. We find mixed evidence in favor of the theory that when the elections become close, more people vote strategically and the volume of wasted votes decreases. The chapter also identifies the role of ethnic heterogeneity in explaining the volume of wasted votes.

A detailed road map for the rest of the chapter is as follows. In Sect. 2, we spend considerable time describing the data. Section 3 presents the econometric model that tests the relative strength of different strands of theory, offering explanations for wasted votes. Finally, Sect. 4 summarizes the principal findings and concludes.

2 Measuring the Wasted Votes

This section tries to measure the amount of wasted votes and looks at the variations across states and elections. It is divided into three parts. Section 2.1 studies the variation at the state level and finds that there is considerable difference in the level of wasted votes across states. Section 2.2 reestimates the wasted votes to see to what extent these are "effective" in switching the identity of the winner from the candidate with the maximum votes to the one with the second largest number of votes. Finally, Sect. 2.3 examines the overall variation of wasted votes at an all India level[2] over time. It is found that with the exception of a few elections, there is no secular trend, upward or downward.

2.1 The Crude Estimate: State-Level Scenario

The starting point of the analysis is the celebrated example of a three-party equilibrium by Myerson and Weber (1993). They demonstrate that positive votes can be received by *all* three parties only if the second and third parties are *very* close in terms of the probabilities of winning the election, making the coordination on one of these candidates difficult for the voters. The same point had been emphasized in the empirical work by Johnston and Pattie (1991), though they did not give any theoretical justification in favor of the proposition. We begin here by looking at the degree of closeness between the second and the third parties in the electoral districts and the amount of third-party support in the same. The crude measure of wasted votes used in this section consists of the sum total of the vote shares[3] of all the parties who

---

[2] We shall mean the sum total of all the 15 major states whenever we refer to the "country" or "all India" figures.

[3] The vote share is the fraction of the total number of "valid votes" obtained by a party or a candidate.

Table 20.1 Distribution of mean wasted votes according to the difference between the second and third parties

	Percentile of Difference between 2 & 3			
	0–5	5–10	90–95	95–99
All India	0.23	0.22	0.03	0.02
AP	0.22	0.22	0.02	0.02
ASM	0.14	0.22	0.00	0.01
BH	0.21	0.23	0.06	0.04
GUJ	0.22	0.16	0.03	0.01
HAR	0.29	0.31	0.06	0.02
KAR	0.23	0.25	0.03	0.02
KER	0.19	0.10	0.02	0.01
MP	0.21	0.19	0.05	0.02
MH	0.25	0.25	0.03	0.01
ORI	0.25	0.22	0.05	0.02
PNJ	0.14	0.23	0.03	0.02
RAJ	0.25	0.18	0.04	0.03
TN	0.20	0.18	0.02	0.01
UP	0.24	0.25	0.10	0.08
WB	0.23	0.20	0.02	0.01
	Max 0.29			*Min* 0.01

All the entries are proper fractions denoting the means of the wasted votes in respective categories

contested the election excepting the top two parties. We divide the constituencies according to the extent of the difference in vote shares between the second and third largest parties. The entire range of difference between the second- and third-party vote shares in each of the 15 major states as well as that for the entire country between 1967 and 1998 is divided up into 20 quantile groups. We look at the arithmetic means of the wasted votes as defined above in each of these groups. The means for the first and last two of these quantile groups are reported in Table 20.1.

Certain trends are noticeable in the data. First, consonant with the predictions of Myerson and Weber (1993), we observe that the wasted votes are declining with increasing difference between the second and third parties. However, this declining tendency is more prominent in some of the states than in the others. In Kerala, Tamil Nadu, Maharashtra, and West Bengal, for instance, the wasted votes are steadily decreasing both in the upper and lower ends of the range. In some other states like Andhra Pradesh and Rajasthan, the decline is somewhat arrested particularly at the lower end. Still others, like Assam, Madhya Pradesh, and Punjab, behave in a slightly erratic manner. While the cases of Assam and Punjab are clearly visible from Table 20.1, both Madhya Pradesh and Punjab register a steep jump in the fourth quantile (not reported in the table).

The states can also be classified into “high” or “low” wasted vote states. Bihar, Haryana, Karnataka, and Uttar Pradesh are clearly members of the first group. On the other hand, in states like Kerala and Tamil Nadu, the wasted votes seem to be low across the range. However, what is astonishing about this is that even in the last quantile group, the value of the wasted votes is at least 1 % of the total number of valid votes. It should be mentioned that the value of the difference between the second- and third-party vote shares in this group is as high as 46 % on average.

Table 20.2 Distribution of mean and standard deviation of wasted votes across states

States	Mean	Standard deviation
AP	0.11	0.08
ASM	0.13	0.10
BH	0.15	0.09
GUJ	0.07	0.06
HAR	0.15	0.11
KAR	0.12	0.09
KER	0.06	0.05
MP	0.11	0.07
MH	0.11	0.09
ORI	0.11	0.08
PNJ	0.12	0.08
RAJ	0.09	0.07
TN	0.07	0.06
UP	0.16	0.09
WB	0.10	0.08

The entries are means and standard deviations of wasted votes across all constituencies over all the elections from 1967 to 1998, calculated for each state. Unit: proper fraction

The above-mentioned classification can be elaborated further by looking at the distribution of the mean wasted votes across states as shown in Table 20.2.[4]

It can be seen that the average wasted vote is greater than 0.10 for 10 out of the 15 states. The values cluster around 0.13. For the rest, the average value is centered on 0.08. Among these, Uttar Pradesh has the highest magnitude of wasted votes, while Kerala has the lowest score.

However, the picture is more revealing than just this. The high average wasted vote is accompanied by high standard deviations as well. This implies that in states with high wasted votes, there is considerable variation in its level across constituencies and also over time. Indeed, some elections are associated with higher wasted votes across the board than the others. But this is the subject matter of Sect. 2.3, where we look at the aspect of time variation in more detail.[5]

2.2 *The "Effective" Wasted Votes: State Level Scenario*

In this section, we turn to a different, somewhat artificial, way of measuring the wasted votes. The rational expectation assumption of Bayesian Nash equilibrium makes two predictions about the election outcome in a single-member district—the first one is regarding the identity of the winner and the second one is related to the distribution of the votes among the candidates. Duverger's law makes a strong statement about the second prediction of the theory, namely that the top two parties should sweep all votes of the constituency and nothing should be left for a third

[4] Let the set of *constituency years* in state s be $c(s)$. The entries in the table denote the mean and standard deviations of wasted vote calculated over all $c \in c(s)$.

[5] See A.1 in the Appendix for a detailed election-wise break up of the state averages of wasted vote.

party. In Sect. 2.1, we have already seen that this prediction is not supported by data. This implies that a large section of the electorate casts their votes sincerely, while according to the theory, they should act strategically and choose the more preferred of the top two candidates.

In the present section, we ask the following question. Suppose the voters who have voted for the third party indeed vote strategically and shift their votes to one of the top two candidates. We then delve into the probability of the event as to what extent this rearrangement of votes is capable of switching the identity of the winner. If the identity remains unchanged, the wasted votes are "ineffective" in the sense that the outcome of the election is compatible with a strategic voting equilibrium in which candidate *one* continues to win.

In order to see the effectiveness of the wasted vote, we do the following exercise. Let the vote shares of the first two parties be v_1 and v_2. Then, the crude measure of wasted votes will be equal to $v_3 = 1 - v_1 - v_2$. Let these wasted votes be distributed between candidates *one* and *two* in the proportions $(1 - x)$ and x, such that *two* is just able to make a tie with *one*. Hence, $xv_3 + v_2 = (1 - x)v_3 + v_1$, which yields $x = (v_1 + v_3 - v_2)/2v_3$. x is the minimum proportion of the wasted votes, which, if transferred to *two* en bloc, enables *two* to turn the table. It is greater than or equal to 0.5 by construction.

Contrast the above with a situation where the wasted votes are distributed randomly to the top two candidates, so that each of them gets 0.5 proportion of it. This is similar to the hypothetical situation where the third-party voters randomly allocate their votes between the first and the second candidates in a very close election. The difference between these two measures, namely $(x - 0.5)$, gives us the required index for the effectiveness of the wasted vote. It measures the extent to which the allocation would have to be systematically biased in favor of the second candidate to make the tie possible. Consider the case when $(x - 0.5)$ is small. If all voters coordinated their votes on the first two candidates, a small amount of excess popularity of the second candidate (over the first) within the class of third-party voters would be enough to alter the outcome of the election. The wasted votes are precious under such circumstances. On the contrary, the further the value of x is from 0.5, wasted votes become less and less costly.

It must be noted that in constituencies where the difference in the vote share between the first and the second candidate (the closeness of the election) exceeds the crude measure of wasted votes, wasted votes can never make a switch in the outcome. These are uninteresting cases, since there is little incentive to vote tactically. We exclude these constituencies from the sample while calculating the effectiveness of the wasted votes. Moreover, this also implies that within the sample the value of $(x - 0.5)$ always lies between 0 and 0.5.

Table 20.3 reports the means and standard deviations of this new index for each of the 15 states.[6] The last column denotes the percentage of *constituency years* in

[6] Let the set of *constituency years* in state s be $c(s)$. The first two columns denote the mean and standard deviations of $(x - 0.5)$, calculated over those $c \in c(s)$ for which difference between vote shares of parties 1 and 2 is less than the sum total of the votes shares of all other parties.

Table 20.3 Distribution of mean and standard deviation of "effective" wasted votes across states

States	Mean	Standard deviation	Percentage of constituencies
AP	0.22	0.14	46.80
ASM	0.23	0.18	56.60
BH	0.20	0.13	47.94
GUJ	0.22	0.14	29.46
HAR	0.14	0.12	52.94
KAR	0.21	0.14	49.09
KER	0.25	0.15	55.62
MP	0.25	0.14	43.23
MH	0.23	0.14	42.59
ORI	0.22	0.15	45.00
PNJ	0.23	0.16	40.22
RAJ	0.21	0.14	39.49
TN	0.26	0.14	25.15
UP	0.18	0.14	51.95
WB	0.25	0.14	52.45

The entries are means and standard deviations of the "effective" wasted votes, i.e., $(x - 0.5)$, across those constituencies for which difference between vote shares of parties 1 and 2 is less than the sum total of the vote shares of all other parties, over all the elections from 1967 to 1998, calculated for each state. The last column gives the percentage of such constituency years in each state.

each state where the difference between the first and second candidates could be bridged by rearranging the wasted votes in the favor of the second candidate. As can be observed, the index is generally small for the high wasted vote states like Uttar Pradesh, Haryana, and Bihar and high for low wasted vote states like Gujarat, Kerala, and Tamil Nadu. This shows that, as a whole, the greater the volume of the wasted votes, the more likely it is that it will be effective in switching the election outcome.

There is one thing worth noticing in Table 20.3. If we look at the mean value of the "effectiveness" index across states, it seems that almost nowhere the wasted votes are effective enough to make much of a change in the election outcomes. For instance, the minimum value is as high as 0.14. This means that the redistribution of third-party votes would have to be at least 64 % vs. 36 % in favor of the second candidate before election outcomes can be overturned. Within the class of third-party voters, the preferences over the first two candidates are unlikely to be so skewed. We claim that much of the puzzle about the "volume" of wasted votes, as discussed in Sect. 2.1, is thus dispelled.

The standard deviations are not very telling in this case. They cluster around 0.14, indicating that the patterns of variation in the effectiveness index are more or less similar across states.[7] We shall now turn to the analysis of the time variation in the wasted votes—both in its crude form and for the effective estimate.

[7] See A.2 in the Appendix for a detailed election-wise break up of the state averages of effective wasted vote.

2.3 Time Variation: The All India Scenario

Wasted or third-party votes have varied considerably over the years from the first general election of independent India until now.[8] The political and economic situations which affect the decisions of a voter have not, of course, remained the same during this span of more than 50 years. There have been significant changes in the demographic characteristics of the Indian electorate, particularly those relating to education, health, and the caste composition. First of all, the electorate has grown in size, both in terms of the absolute number and also in terms of the proportion of the adult population who exercise their franchise. Political and economic empowerment has been extended to certain sections of the population. The Dalits and women have emerged as important demographic groups in terms of their political clout. With the advancement of technology, new modes of information dissemination have come into vogue and these have played an important role in the social, economic, and political life of individuals.

There have also been important political events, both within the country and also in the outside world, which have influenced the outcomes of certain elections. For example, the declaration of emergency in 1975 had a great impact on the outcome of the following parliamentary election of 1977. Almost every election is unique in this respect. Still, one can say that some of these have been preceded by more "normal" years than the others. This fact is borne out very well from the data.

The over-time variation in the wasted votes in India is quite exceptional in the sense that for the most part it has moved in the direction opposite to what is predicted by the theory of calculus of voting: Its level has been high in those elections where the race was close. We analyze the period between 1967 and 1998. In the earlier elections, the wasted votes turn out to be high. In our dataset, we find it to be as high as 12 % in 1967 and 10 % in 1971. The 1977 election proved to be an exception—the wasted votes fell to 5 %—due to the extraordinary political developments that preceded it. In the next election, that is the one that took place in 1980 before the usual term of the government was over, the wasted vote level shot up to 14 %. There was widespread confusion in the electorate regarding the relative chances of various parties. The Congress (I) was trying hard to fight back, while the Janata Dal had the advantage of being in office for the past 3 years. The memory of the emergency was still fresh in the minds of the people. That the race was indeed very close is proved by the relatively low differences in the vote shares between the winner and the second and also that between the second and the third parties.

An interesting, but unexplored, stylized fact emerges from Table 20.4.[9] We have calculated the arithmetic mean of wasted votes—both of the crude measure and of

---

[8] Chhibber and Kollman (1998) have found it to vary from 22 % in 1980 to 2 % in 1977. See footnote 11 of the same.

[9] Let the set of constituencies distributed over all the 15 major states in the election year t be $c(t)$. The first two columns of 20.4 denote the mean and standard deviation of crude wasted votes calculated over all $c \in c(t)$. The next two columns show the mean and standard deviation of the effectiveness index calculated over those $c \in c(t)$ in which the difference between vote shares of parties 1 and 2 is less than the sum total of the vote shares of all other parties. The last column shows proportion

Table 20.4 Distribution of means and standard deviations of crude and effective estimates of the wasted votes across elections

Election	Crude mean	Crude standard deviation	Effective mean	Effective standard deviation	Proportion of constituencies
1967	0.12	0.09	0.23	0.14	65.57
1971	0.10	0.08	0.23	0.15	33.68
1977	0.05	0.05	0.28	0.15	29.32
1980	0.14	0.09	0.22	0.15	37.17
1984	0.09	0.07	0.24	0.15	27.61
1989	0.10	0.08	0.23	0.14	33.57
1991	0.13	0.08	0.23	0.14	52.69
1996	0.16	0.08	0.20	0.13	64.30
1998	0.14	0.09	0.19	0.13	64.22

The entries are means and standard deviations of wasted votes calculated over the appropriate constituencies across all states for each election year

the effectiveness index—across all the constituencies distributed over the 15 major states for each of the elections and examined its variation over the entire period. The last three general elections, held in 1991, 1996, and 1998, witnessed very high levels of wasted votes. This happened despite the fact that the vote distances between the political parties were continuously declining.

An investigation into the "effectiveness" of the wasted votes reveals little variation in the level except in the year 1977, when the index was higher than its usual level implying that wasted votes were actually valueless. This is expected given that the election was not close at all. Again, after 1991, there is a continuous decline in the value of this index indicating that wasted votes have become more costly in recent years. This result, combined with the fact that the level of wasted votes had also been rising at the same time, is a bit counterintuitive. Figures 20.1 and 20.2 show the movements of the wasted votes and the effectiveness index. [10] It can be seen clearly that, as a whole, there is no trend in the data. However, in the last three elections, the two graphs are rising and falling, respectively.

We shall wind up the description of the data with the transition probability matrix. Specifically, we ask the following question: Suppose that the level of wasted votes in constituency i is high (low) during election t, then, what is the probability that the level of wasted votes in constituency i remains high (low) during election $(t + 1)$? Persistence in the patterns of wasted votes reveals the importance of constituency-level determinants. We proceed as follows. We divide the range of wasted votes into three quantile groups: high, medium, and low. Then label each constituency with the appropriate tag. Next, we see how the constituency has changed its status over the years. The result is summarized in Table 20.5. The values in each cell denote the probability of a transition from a certain status to the other. For example, pLL is the probability that if a constituency is a low wasted vote constituency, it will remain

of $c(t)$ in which difference between vote shares of parties 1 and 2 is less than the sum total of the vote shares of all other parties.

[10] 20.1 and 20.2 are constructed from columns 1 and 3 of Table 20.4, respectively.

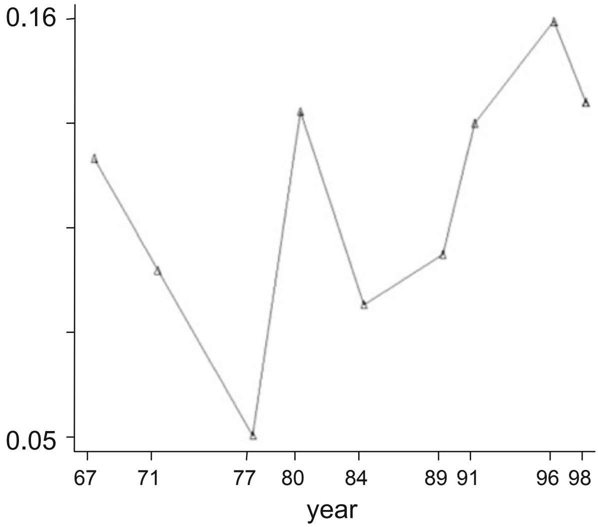


Fig. 20.1 Wasted votes

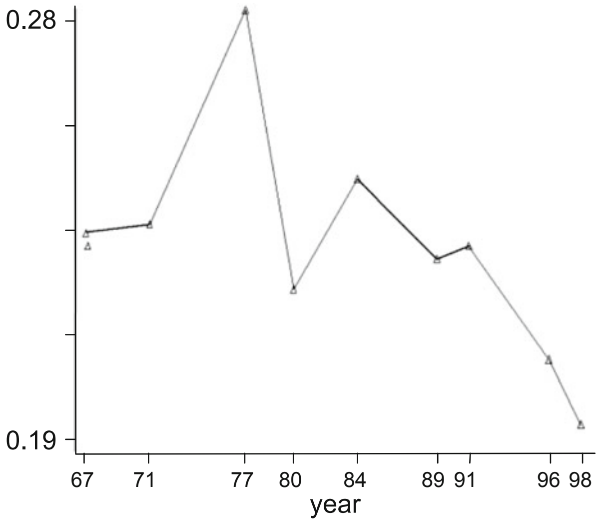


Fig. 20.2 Effectiveness

so in the next election, *p*LM is the probability that a low wasted vote constituency will shift to a medium class and so on. The values of these probabilities are shown in Table 20.5.

It can be seen that the diagonal values are larger than the off-diagonal ones. This indicates that there is indeed a lot of persistence in the level of wasted votes at a constituency level. This provides us with the motive for introducing constituency-level dummies to capture the time-invariant constituency-level characteristics when we try to estimate the influences of various regressors on the level of wasted votes. This issue will be taken up in the next section.

Table 20.5 Transition probabilities

Initial Status	Next period status		
	Low	*Medium*	*High*
Low	0.492743	0.3418	0.164006
Medium	0.344487	0.38403	0.269962
High	0.182168	0.294389	0.523444

In this section, we assess the relative importance of various factors identified in the literature as affecting the volume of the wasted votes. We use panel data techniques to estimate the model. This analysis has its shortcomings. It cannot be used, for instance, to infer individual-level behavior, such as strategic voting. The regression results presented here should be interpreted with caution. Though care has been taken to control for unobserved constituency-level characteristics, large political systems, being too complex, omitted variables bias is hard to eradicate completely. The model, therefore, is best viewed as a predictor of wasted votes without any claim on causality.

3 Modeling Wasted Votes in the Indian Electorate

3.1 The Data and the Methodology

The dependent variable used in this analysis is wasted vote at the constituency level, measured in two different ways.[11] The first one is the total number of votes received by all the parties in the constituency excepting the votes of the top two parties, divided by the total number of valid votes in the constituency. This is a proper fraction, say p, which can vary from 0 to 1. To avoid range restriction on the error term, we work with logit of p, i.e., our dependent variable is log of the ratio $p/(1 - p)$. We call this expression *wv*.

Note, first, that *diff12* is the difference in the vote shares of the top two parties in a constituency. When the election is close—i.e., *diff12* is small—the volume of wasted votes can be large. In particular, footnote 12 demonstrates that the upper bound for wasted votes is given by (1 − *diff12*)/3.[12] Rather than considering the actual volume of wasted votes, our second measure calculates the wasted vote, p, as a proportion of its theoretical upper bound, (1 − *diff12*)/3. Finally, we take the logit of [p ÷ (1 − *diff12*)/3] to ensure that the range remains $-\infty$ to ∞. We call this second measure *wv1*.

[11] A detailed description of each of the variables is provided in the Appendix.

[12] The upper bound is derived in the following manner. Let there be three parties in a constituency whose vote shares are denoted by v_1, v_2, and v_3, where $0 < v_3 < v_2 < v_1 < 1$ and $v_1 + v_2 + v_3 = 1$. Then, $v_2 > v_3 \Rightarrow 2v_2 + v_3 > 3v_3$. The left-hand side (LHS) of the last inequality can be written as $v_1 + v_2 + v_3 - (v_1 - v_2) = 1 - (v_1 - v_2)$. This implies that $v_3 < [1 - (v_1 - v_2)]/3$. Note that, *diff12* = $(v_1 - v_2)$. Hence, the upper bound of v_3 is given by [1 − *diff12*]/3.

There are three sets of regressors used in the analysis. They pertain to three different theories for explaining wasted votes. We test each of these in turn, and look at their comparative explanatory power in explaining the volume of wasted votes in the Indian elections. The first one is the well-known calculus of voting theory, which attributes the volume of wasted votes to the closeness of the election. As elections get more and more close between the top two candidates wasted votes become more and more costly. Under such situations, voting sincerely might pave the way for a low ranked candidate to win the election. Hence, voters do better by choosing the more preferred of the top two parties. The prediction, therefore, is that the less the distance between the top two parties, the less is the volume of wasted votes. An extension of this theory (Cox 1997) holds that the larger the difference between the second and the third parties, the more prominent is the identity of the loser and hence less the wasted vote volume. In the present chapter, we call the vote share differences between the top two parties and that between the second and the third parties, *diff12* and *diff23*, respectively.

The second class of theory relates the volume of wasted votes to the extent of ethnic heterogeneity within the electorate. A priori, ethnic fragmentation may affect the volume of wasted votes in two opposite ways. In a constituency with high level of ethnic fragmentation, voters' preferences regarding policy outcomes may be extremely disperse making wasted votes costly, thereby creating tendencies to vote strategically. On the other hand, ethnic identity may have certain emotional or ideological underpinning, which induces a person to vote sincerely. Consequently, with increasing fragmentation within the polity, votes may be divided up between candidates representing different ethnic interests. Ordeshook and Shvetsova (1994) find that in single-member district plurality systems, heterogeneity at the constituency level does not affect the level of the wasted vote. We measure heterogeneity as the probability that two randomly picked persons will belong to the same ethnic group.[13] We incorporate religious, linguistic, and caste heterogeneity measures (*hetrrelg, hetrlang,* and *hetrcast,* respectively) and find their effects on wasted votes.

The third and the final theory to be discussed here is one which seeks to explain wasted votes in terms of the size and activism of the central or the federal government. Chhibber and Kollman (1998) have argued that as the federal government centralizes power, "Voters develop national policy preferences and candidates associate themselves with certain national policy positions." As a result, locally competitive but nationally uncompetitive parties are abandoned by the voters. One implication of this theory is that as the size of the public good increases, the identity of the winner becomes more important and wasted votes turn out to be a costly option to the voter. We introduce four different measures of government activism—the total capital disbursement, total capital outlay, total revenue expenditure, and the revenue expenditure on development activities (*tcd, tco, tre,* and *rxdev,*[14] respectively). We do not find any systematic influence of any of these factors on the level of wasted votes.

---

[13] Exact definitions of the heterogeneity measures are provided in the Appendix.

[14] All of these are measured per capita.

We have constituency-level data for the variables related to wasted votes and closeness. The rest are computed from state-level data. We run regressions with robust standard errors on a panel using constituency and state-level dummies. The former takes the following form:

$$y_{ct} = a_c + \delta_t + x_{ct}\beta + z_{s(c)t}\chi + \varepsilon_{ct}, \tag{20.1}$$

while the latter looks like

$$y_{ct} = a_{s_{(c)t}} + \delta_t + x_{ct}\beta + z_{s(c)t}\chi + \varepsilon_{ct}, \tag{20.2}$$

where *yct* is the wasted votes in constituency *c* during the *t*'th election, αc and δt denote constituency and election-specific dummies, *xct* is the vector of closeness measures of the *t*'th election in constituency *c, s*(*c*) identifies the state in which constituency *c* is located, and *zs*(*c*)*t* is the corresponding vector of state-level regressors in the *t*'th election.

3.2 The Main Results

Table A.3 in the Appendix presents the basic results. The three sets of regressors are introduced in succession. We start by testing the effect of the closeness measures on the volume of wasted votes. The first two models, namely models 1 and 2, use the original measure of wasted votes (*wv*) and involve state- and constituency-level dummies, respectively. The same exercise is repeated, for the deflated measure, i.e., *wv1* and the results are reported in models 3 and 4. We also incorporate various state-level regressors related to newspaper circulation, literacy rate, the proportion of the rural population, and the state domestic product. These variables are not the focus of this analysis. We introduce them to control for various demographic characteristics at the constituency level.

The results are unexpected as far as the *diff12* variable is concerned. The estimated coefficient is found to be negative, statistically significant, and robust across specifications. This implies that as the election becomes closer—i.e., as the difference in vote shares between the top two parties decreases—wasted votes register a rise. This violates the prediction of the theory that, in close elections, voters should vote strategically and opt for one of the top two candidates. However, the *diff23* variable, all throughout, has the expected negative sign, which shows that as the distance between the second and the third candidate increases, the volume of wasted votes declines. This implies that voters abandon candidates who are identified as the likely losers.

Keeping the closeness variables in place, we now introduce the heterogeneity measures into the model. As stated earlier, we have measures for linguistic, religious, and caste heterogeneities in the demographic composition of the electorate. Among these, the linguistic and religious heterogeneity measures are found to have a positive effect on the volume of the wasted votes, whereas an increment in the caste heterogeneity seems to reduce the same. There is hardly any theory that tells us whether these results are to be expected. However, one can put forward informal explanations in favor of these results. The schedule caste and scheduled tribe population is extremely dynamic in India, both socially and politically. They have been the center of many political debates in the past few decades, reservation being one of them. The issue is extremely sensitive and has a symbolic value in the minds of certain sections of the population. There is little doubt that with the increasing population of these castes in the electorate, people will perceive the political competition even more sharply. Thus, with the same level of closeness, a constituency with a more heterogeneous caste composition will have less wasted votes than one where degree of such heterogeneity is small.

The other two heterogeneity measures, however, indicate just the opposite. They show that wasted votes are likely to be higher in a district where the population is multilinguistic and multireligious. This difference in the behavior of the various heterogeneity measures is hard to explain. Perhaps the clue to this problem lies in the different manners in which these various identities—linguistic, religious, and caste—enter the political life of an individual. While we have already mentioned the importance of caste in the political decision-making process, we conjecture that the role of the other two is probably not one of promoting the sense of political competition. Rather, these may have certain emotional or ideological underpinning, which induces a person to vote sincerely.

In the end, we incorporate government-spending variables in the model, representing government activism or the size of the government. These variables are not significant except in the last two models, where the revenue account spending is found to be significant. Among these, only the development expenditure seems to have a consistent negative sign, indicating that higher expenditure on development projects causes wasted votes to decline.

Although we did not put much emphasis on the demographic variables, some of them seem to have significant effect on the wasted vote. The coefficient on newspaper circulation has a negative sign, indicating wasted votes are likely to be low in a more informed constituency. However, an increase in the proportion of rural population seems to decrease the volume of wasted votes. One implication of this is that it had been relatively difficult for new parties to make a breakthrough in the rural areas, which are found to be favoring the status quo. Another interesting feature is the positive sign on the per capita state domestic product. It implies that richer constituencies tend to have more wasted votes. Given that we have already controlled for information and literacy, it probably indicates that the cost of a wasted vote is less to the more affluent people.

3.3 The Measurement Error Problem

We have already estimated the effect of the closeness of the election on the level of wasted votes. This estimate is flawed in that election closeness suffers from measurement error. Theory maintains that an ex ante prediction of election closeness affects wasted votes; on the other hand, we have taken actual (or ex post) election closeness as the regressor. Ex post election closeness is, after all, only an erroneous approximation of ex ante closeness. The measurement error biases the estimated coefficient downwards.

One way to address this problem is to find instruments for election closeness. A natural choice for the instrument is the lagged value of the variable, that is, (*diff12*)-1 and (*diff23*)-1, respectively, in our case. There is one practical problem, however, in constructing the instrumental variables from the lagged values of the vote share differences. The elections are too far apart, making the correlation between the instrument and the relevant variable extremely low. The correlation between *diff12* and its lagged value is only 0.13, while that between *diff23* and (*diff23*)-1 is about 0.23. Consequently, the lagged values provide us with very poor instruments in the present case.

We reestimated all the 12 models of Table A.3 using two-stage least squares technique. The variables turned out to be mostly insignificant. However, *diff23* continued to be robust. The only other variable which we found significant was newspaper circulation. It was negative and significant, substantiating the role of information dissemination in the reduction of wasted votes.

3.4 The Year-Wise Estimates

The aggregate models described in Eqs. (20.1) and (20.2) are too restrictive in the sense that they force the same β and γ across all the states and elections. As we have already noted in Sect. 2.3, there is considerable variation in the wasted votes across elections, and each election has its unique features. In order to see if the estimates vary significantly over time or if they remain more or less unchanged around the values estimated by the aggregate model, we have run "ordinary least square" regressions on wasted votes separately for each election year.

Since the demographic control variables in our dataset are values at the state level, multicollinearity rules out the incorporation of state and constituency dummies in the estimated regressions. Hence, we use lagged values of the dependent variable (*wv* and *wv1*) to crudely control for unobserved constituency-specific characteristics.

The estimates have deteriorated considerably, but this is partly due to a massive reduction in the number of observations. The summary of the signs on the coefficients and their significance levels is given in Table A.4 in the Appendix. The entries denote the number of times the coefficient on a particular regressor was found positively or negatively significant or insignificant.

The closeness measures are again very strongly significant with negative sign. The caste heterogeneity is mostly correctly signed. However, there has been reversal in the sign of the linguistic heterogeneity. Newspaper has also worsened considerably, but its place has been taken by literacy rate, which is a substitute to the former.

The government expenditure variables are still far from satisfactory. They are mostly insignificant. When significant, they are signed incorrectly, pointing to a direction opposite to the prediction of the theory, namely increased government activity reduces wasted votes (see *tco* and *tre*). The development expenditure variable is signed both ways. Only the capital disbursement variable seems to be moving in the correct direction, but only in a few cases. One should not read much into this, since capital account variables are very noisy.

4 Conclusion

The chapter attempts an aggregative analysis of wasted or third-party votes in parliamentary elections of India. It does not, however, measure the extent of strategic voting, which is a closely related aspect.

We measure the volume of wasted votes at the constituency level and find that it is quite large. The average across all the states and all the parliamentary elections from 1967 to 1998 is as high as 11 %. However, we have also shown that a reallocation of the wasted votes to the top two candidates is unlikely to affect the identity of the winner of the constituency in most of the cases.

The chapter also looks at the determinants of wasted votes at the constituency level. It demonstrates that as the distance between the second- and third-ranked candidates in a constituency decreases, the volume of wasted votes tends to go up. It identifies an aberration to one of the predictions of the calculus of voting theory in the context of the Indian electorate. It shows that when the elections are close between the top two candidates, the volume of wasted votes tends to be larger. We pose it as a puzzle.

At the policy level, it identifies the role of information dissemination and spread of literacy in bringing down the volume of wasted votes. However, it does not find any evidence in favor of government activism in achieving the same target.

Acknowledgments I am grateful to Sandwip Das, Sugato Dasgupta, Satish Jain, and Rajendra Prasad Kundu for their helpful comments. The usual disclaimer applies.

Appendix

Data sources and definitions of the variables

- Wasted votes

1. *wv*—Let the total number of votes received by all the parties in the constituency excepting the top two parties divided by the total number of valid votes in the same be *p*. $wv = \ln [p/(1-p)]$. This is constructed from constituency-level data.
   Unit: real number
   Source: Election Commission of India

2. *wv1*—Let the vote shares of the first and the second parties be v_1 and v_2, respectively. Let $q = p/[(1-v1 + v_2)/3]$, $wv1 = \ln[q/(1 - q)]$. Constituency-level data.
   Unit and source: same as above

- Closeness of the race
- (1) *diff12*—Let the vote shares of the first two parties as a proportion of the total number of valid votes in the district be v_1 and v_2. $diff12 = v_1 - v_2$.
- (2) *diff23*—Let the vote shares of the second and the third parties as a proportion of the total number of valid votes in the district be v_2 and v_3. $diff23 = v_2 - v_3$. Constituency-level data.
  Unit: proper fraction
  Source: Election Commission of India

- Heterogeneity indices

These variables have been constructed from the state-level population data obtained from the Census of India. The figures for the four census years, 1961, 1971, 1981, and 1991 have been interpolated to generate the numbers for the intermediate years. We have taken the 3-year moving averages of these numbers to smoothen the series. The figures for the election years have been taken from this series.

We have calculated the index as the probability that two randomly picked persons from the sample would belong to two different demographic groups. Let the proportion of people belonging to each group be p_1, p_2, etc.

Then, the index is given by $1 - \sum p_i{}^2$, $i = 1, 2 \ldots n$.

1. *hetrlang*—We have calculated this index from the number of people belonging to the 14 major language groups; namely, Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Malayalam, Marathi, Oriya, Punjabi, Sanskrit, Tamil, Telegu, and Urdu.
2. *hetrrelg*—Constructed from the eight major religious groups, classified as Hindu, Muslim, Christian, Sikh, Buddist, Jain, Others, and Not Stated.
3. *hetrcast*—Constructed from scheduled caste (SC), scheduled tribe (ST), and other population in each constituency.

Source: Census of India, various volumes

- Government spending (state-level data)

1. *tcd*—Per capita total capital disbursement
   Unit: Rupees (1960 prices)
2. *tco*—Per capita total capital outlay
   Unit: Rupees (1960 prices)
3. *tre*—Per capita total revenue expenditure
   Unit: Rupees (1960 prices)
4. *rxdev*—Per capita revenue expenditure on development activities
   Unit: Rupees (1960 prices)

Source: Reserve Bank of India Bulletin, various issues, Reserve Bank of India

- Newspaper circulation (state-level data)
- The average number of copies of newspapers, magazines, and periodicals sold or distributed freely per publishing day, in each of the 15 major states.
  Unit: thousands of copies
  Source: Annual Reports of the Registrar for Newspapers in India, various years
- Literacy rate (state-level data)
- The number of literates in each state divided by the total population of the state.
  Unit: proper fraction
  Source: Census of India, various volumes
- Proportion of the rural population (state-level data)
- The proportion of rural population in each state
  Unit: proper fraction
  Source: Census of India, various volumes
- Per capita state domestic product (*sdp*) (state-level data) Unit: Rupees (1960 prices)
  Source: Reserve Bank of India Bulletin, variousissues, Reserve Bank of India

Table A.1 Distribution of mean and standard deviation of wasted votes across state and elections

State	Election									
	1967	*1971*	*1977*	*1980*	*1984*	*1989*	*1991*	*1996*	*1998*	*State (all elections)*
AP	0.15, 0.08	0.09, 0.08	0.05, 0.06	0.14, 0.06	0.05, 0.06	0.05, 0.05	0.12, 0.07	0.17, 0.07	0.18, 0.09	0.11, 0.08
ASM	0.11, 0.12	0.14, 0.08	0.05, 0.05	0.01, NA	0.18, 0.1	NA	0.26, 0.02	0.21, 0.06	0.2, 0.06	0.13, 0.1
BH	0.19, 0.08	0.16, 0.09	0.09, 0.07	0.22, 0.05	0.13, 0.08	0.12, 0.1	0.16, 0.08	0.14, 0.08	0.18, 0.09	0.15, 0.09
GUJ	0.07, 0.07	0.06, 0.06	0.04, 0.04	0.09, 0.06	0.07, 0.06	0.05, 0.03	0.06, 0.04	0.08, 0.06	0.13, 0.08	0.07, 0.06
HAR	0.15, 0.12	0.11, 0.05	0.05, 0.05	0.27, 0.05	0.11, 0.07	0.06, 0.03	0.25, 0.05	0.28, 0.07	0.22, 0.09	0.15, 0.11
KAR	0.12, 0.1	0.03, 0.03	0.03, 0.03	0.17, 0.05	0.06, 0.04	0.16, 0.07	0.19, 0.06	0.22, 0.07	0.19, 0.07	0.12, 0.09
KER	0.1, 0.09	0.06, 0.04	0.02, 0.03	0.03, 0.02	0.07, 0.05	0.06, 0.03	0.06, 0.03	0.09, 0.03	0.09, 0.03	0.06, 0.05
MP	0.13, 0.08	0.08, 0.06	0.05, 0.06	0.16, 0.06	0.1, 0.06	0.11, 0.06	0.09, 0.07	0.16, 0.06	0.12, 0.09	0.11, 0.07
MH	0.11, 0.08	0.08, 0.06	0.04, 0.06	0.12, 0.08	0.1, 0.06	0.14, 0.09	0.16, 0.08	0.19, 0.08	0.06, 0.06	0.11, 0.09
ORI	0.11, 0.1	0.22, 0.06	0.06, 0.06	0.15, 0.03	0.07, 0.05	0.06, 0.04	0.13, 0.06	0.16, 0.08	0.1, 0.07	0.11, 0.08
PNJ	0.21, 0.08	0.13, 0.07	0.05, 0.04	0.09, 0.07	0.12, 0.08	0.14, 0.08	0.16, 0.08	0.17, 0.05	0.03, 0.02	0.12, 0.08
RAJ	0.1, 0.08	0.07, 0.05	0.04, 0.03	0.18, 0.06	0.12, 0.06	0.06, 0.04	0.11, 0.08	0.11, 0.06	0.09, 0.07	0.09, 0.07
TN	0.05, 0.06	0.04, 0.05	0.04, 0.04	0.04, 0.04	0.03, 0.03	0.08, 0.07	0.09, 0.06	0.15, 0.05	0.09, 0.06	0.07, 0.06
UP	0.2, 0.08	0.12, 0.07	0.07, 0.05	0.25, 0.06	0.16, 0.06	0.18, 0.07	0.2, 0.07	0.24, 0.05	0.23, 0.06	0.16, 0.09
WB	0.13, 0.1	0.17, 0.08	0.07, 0.07	0.09, 0.06	0.04, 0.04	0.06, 0.05	0.15, 0.05	0.11, 0.05	0.16, 0.05	0.1, 0.08
All India	0.12, 0.09	0.1, 0.08	0.05, 0.05	0.14, 0.09	0.09, 0.07	0.1, 0.08	0.13, 0.08	0.16, 0.08	0.14, 0.09	

The first and second entries in each cell denote the mean and the standard deviation of wasted votes for the particular state year, respectively

Table A.2 Distribution of mean and standard deviation of "$x - 0.5$" and percentage of constituencies with "effective" wasted votes across states and elections

State	Election									
	1967	*1971*	*1977*	*1980*	*1984*	*1989*	*1991*	*1996*	*1998*	*State (all elections)*
AP	0.24, 0.14, 53.33 %	0.19, 0.14, 25 %	0.32, 0.15, 24.39 %	0.36, 0.12, 28.57 %	0.24, 0.17, 30.95 %	0.25, 0.13, 26.19 %	0.24, 0.15, 60.98 %	0.18, 0.12, 86.49 %	0.16, 0.11, 87.5 %	0.22, 0.14, 46.80 %
ASM	0.24, 0.33, 66.67 %	0.48, NA, 14.29 %	0.23, 0.31, 50 %	NA	0.24, 0.14, 71.43 %	NA	0.19, 0.18, 100 %	0.21, 0.14, 80 %	0.18, 0.21, 57.14 %	0.23, 0.18, 56.60 %
BH	0.14, 0.11, 70.59 %	0.2, 0.14, 50 %	0.38, 0.14, 12.24 %	0.2, 0.15, 3.33 %	0.23, 0.13, 26.32 %	0.17, 0.11, 28.21 %	0.25, 0.12, 42.86 %	0.2, 0.1, 50 %	0.16, 0.1, 86.67 %	0.2, 0.13, 47.94 %
GUJ	0.25, 0.12, 65.22 %	0.27, 0.14, 33.33 %	0.31, 0.16, 34.62 %	0.21, 0.15, 15.38 %	0.15, 0.1, 23.08 %	0, NA, 3.85 %	0.28, 0.15, 20 %	0.16, 0.18, 25 %	0.19, 0.14, 50 %	0.22, 0.14, 29.46 %
HAR	0.05, 0.02, 60 %	0.17, 0.16, 44.44 %	0.18, NA, 20 %	0.08, 0.08, 71.43 %	0.36, 0.21, 30 %	0.12, 0.05, 30 %	0.15, 0.07, 100 %	0.07, 0.06, 80 %	0.11, 0.09, 100 %	0.14, 0.12, 52.94 %
KAR	0.24, 0.15, 64 %	NA, NA, 19.23 %	0.13, 0.04, 39.29 %	0.28, 0.22, 16.67 %	0.22, 0.15, 25 %	0.3, 0.09, 39.13 %	0.21, 0.14, 86.36 %	0.14, 0.12, 95.24 %	0.21, 0.12, 73.91 %	0.21, 0.14, 49.09 %
KER	0.3, 0.1, 57.89 %	0.29, 0.06, 21.05 %	0.32, 0.18, 60 %	0.23, 0.15, 35 %	0.21, 0.19, 35 %	0.25, 0.19, 65 %	0.27, 0.14, 70 %	0.29, 0.13, 80 %	0.15, 0.13, 75 %	0.25, 0.15, 55.62 %
MP	0.21, 0.17, 70 %	0.34, 0.1, 28.57 %	0.28, 0.16, 42.5 %	0.27, 0.15, 47.06 %	0.29, 0.1, 14.29 %	0.29, 0.14, 31.43 %	0.25, 0.14, 50 %	0.2, 0.12, 56.25 %	0.22, 0.13, 61.11 %	0.25, 0.14, 43.23 %
MH	0.24, 0.12, 48.65 %	0.26, 0.18, 17.95 %	0.25, 0.11, 43.75 %	0.19, 0.14, 14.29 %	0.31, 0.12, 29.73 %	0.2, 0.14, 56.52 %	0.24, 0.13, 53.66 %	0.24, 0.14, 80 %	0.2, 0.15, 37.5 %	0.23, 0.14, 42.59 %
ORI	0.29, 0.19, 68.75 %	0.18, 0.12, 63.64 %	0.15, 0.1, 61.9 %	0.05, NA, 7.14 %	0.16, 0.2, 14.29 %	0.31, 0.04, 23.81 %	0.21, 0.16, 77.78 %	0.23, 0.14, 58.82 %	0.26, 0.16, 38.1 %	0.22, 0.15, 45 %
PNJ	0.23, 0.17, 75 %	0.15, 0.2, 27.27 %	0.49, NA, 7.69 %	0.18, 0.13, 33.33 %	0.26, 0.19, 50 %	0.41, 0.04, 50 %	0.1, 0.09, 40 %	0.22, 0.12, 100 %	0.4, NA, 8.33 %	0.23, 0.16, 40.22 %
RAJ	0.36, 0.08, 72.22 %	0.22, 0.19, 33.33 %	NA, NA, 20 %	0.16, 0.1, 63.16 %	0.25, 0.1, 15.79 %	0.12, 0.16, 8 %	0.2, 0.15, 45.45 %	0.15, 0.11, 60.87 %	0.24, 0.16, 47.83 %	0.21, 0.14, 39.49 %

Table A.2 (continued)

State	Election									
	1967	*1971*	*1977*	*1980*	*1984*	*1989*	*1991*	*1996*	*1998*	*State (all elections)*
TN	0.22, 0.15, 51.28 %	0.22, 0.2, 48.72 %	0.16, NA, 12.82 %	0.34, 0.23, 12.82 %	0.34, NA, 15.38 %	0.34, 0.08, 15.79 %	0.44, 0.01, 5.13 %	0.38, 0.08, 12.5 %	0.2, 0.13, 50 %	0.26, 0.14, 25.15 %
UP	0.2, 0.14, 81.25 %	0.25, 0.15, 24.53 %	0.43, NA, 12.05 %	0.2, 0.13, 93.55 %	0.28, 0.15, 31.58 %	0.17, 0.14, 57.14 %	0.18, 0.13, 83.33 %	0.16, 0.11, 85.29 %	0.12, 0.1, 90 %	0.18, 0.14, 51.95 %
WB	0.22, 0.12, 88.89 %	0.23, 0.16, 72.73 %	0.34, 0.13, 41.46 %	0.22, 0.13, 21.43 %	0.2, 0.15, 40.48 %	0.23, 0.14, 26.19 %	0.24, 0.13, 61.9 %	0.25, 0.13, 63.41 %	0.29, 0.16, 71.79 %	0.25, 0.14, 52.45 %
All India	0.23, 0.14, 65.57 %	0.23, 0.15, 33.68 %	0.28, 0.15, 29.32 %	0.22, 0.15, 37.17 %	0.24, 0.15, 27.61 %	0.23, 0.14, 33.57 %	0.23, 0.14, 52.68 %	0.2, 0.13, 64.3 %	0.19, 0.13, 64.22 %	

The first, second, and third entries in each cell denote the mean and standard deviation of $(x-0.5)$ and the percentage of constituencies in which the difference between the top two candidates can be bridged by wasted votes, for the particular state year, respectively

Table A.3 Regression Results—Aggregate Model

Specifications	Closeness				Heterogeneity				Government acitivism			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
	Regressions with robust standard errors											
Dependent variable	wv	wv	wv1	wv1	wv	wv	wv1	wv1	wv	wv	wv1	wv1
Independent variables												
diff12	−3.677	−3.605	−4.397	−4.241	−3.702	−3.634	−4.431	−4.278	−3.699	−3.630	−4.420	−4.263
	(−50.52***)	(−47.75***)	(−39.15***)	(−35.64***)	(−50.28***)	(−47.49***)	(−39.42***)	(−35.74***)	(−49.84***)	(−46.99***)	(−39.04***)	(−35.20***)
diff23	−7.827	−7.77	−14.555	−14.371	−7.843	−7.791	−14.576	−14.398	−7.839	−7.786	−14.590	−14.415
	(−83.25***)	(−79.08***)	(−99.68***)	(−94.8***)	(−83.18***)	(−79.31***)	(−100.07***)	(−95.23***)	(−83.06***)	(−79.07***)	(−100.08***)	(−95.22***)
Hetrlang					1.056	1.084	1.535	1.641	0.850	0.866	2.170	2.291
					(3.220***)	(3.232***)	(2.782***)	(2.868***)	(2.359***)	(2.334**)	(2.945***)	(3.004***)
Hetrrelg					2.041	2.008	2.383	2.462	2.256	2.224	3.393	3.479
					(2.577***)	(2.635***)	(2.056**)	(2.152**)	(2.756***)	(2.766***)	(2.744***)	(2.811***)
Hetrcast					−1.618	−1.676	−2.293	−2.268	−1.651	−1.711	−2.253	−2.232
					(−3.213***)	(−3.278***)	(−3.066***)	(−2.985***)	(−3.278***)	(−3.346***)	(−3.022***)	(−2.942***)
Tcd									0.001	0.001	−0.003	−0.003
									(0.519)	(0.473)	(−1.345)	(−1.290)
Tco									−0.0002	−0.0003	0.000	−0.000
									(−0.148)	(−0.182)	(0.046)	(−0.079)
Tre									−0.001	−0.001	0.004	0.004
									(−0.609)	(−0.646)	(1.781*)	(1.786*)
Rxdev									−0.001	−0.001	−0.007	−0.006
									(−0.579)	(−0.512)	(−2.147**)	(−2.067**)

Table A.3 (continued)

Specifications	Closeness				Heterogeneity				Government acitivism			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
	Regressions with robust standard errors											
Dependent variable	wv	wv	wv1	wv1	wv	wv	wv1	wv1	wv	wv	wv1	wv1
Independent variables												
Newspaper	−0.00003	−0.00003	−0.00003	−0.00003	−0.00004	−0.00004	−0.00005	−0.00005	−0.00004	−0.00004	−0.00006	−0.00006
	(−4.943***)	(−4.855***)	(−3.436***)	(−3.244***)	(−6.866***)	(−6.869***)	(−4.547***)	(−4.358***)	(−6.897***)	(−6.885***)	(−4.903***)	(−4.667***)
Literacy	−0.439	−0.333	0.167	0.379	−0.353	−0.239	0.378	0.538	−0.262	−0.151	−0.119	0.021
	(−1.358)	(−0.973)	(0.298)	(0.674)	(−0.988)	(−0.637)	(0.621)	(0.882)	(−0.719)	(−0.392)	(−0.194)	(0.034)
Rural	−2.413	−2.558	−3.606	−3.793	−4.075	−4.223	−5.645	−5.875	−4.161	−4.306	−6.356	−6.553
	(−4.238***)	(−4.426***)	(−4.260***)	(−4.546***)	(−5.366***)	(−5.674***)	(−5.219***)	(−5.475***)	(−5.247***)	(−5.481***)	(−5.537***)	(−5.707***)
Sdp	0.0002	0.0002	0.0001	0.0001	0.0002	0.0002	0.00004	0.00004	0.0003	0.0003	0.0002	0.0002
	(2.485***)	(2.474***)	(0.735)	(0.654)	(2.153**)	(2.118**)	(0.335)	(0.304)	(2.684***)	(2.627***)	(1.168)	(1.149)
Election dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes		Yes		Yes		Yes		Yes		Yes	
Constituency dummies		Yes		Yes		Yes		Yes		Yes		Yes
No. of Obs	3393	3393	3393	3393	3393	3393	3393	3393	3393	3393	3393	3393
R Squared	0.8204	0.8546	0.8732	0.8956	0.8217	0.8559	0.8739	0.8964	0.8219	0.8561	0.8742	0.8966

Values in parentheses denote the *t* values. Levels of significance: * = 90 %; ** = 95 %; *** = 99 %

Table A.4 Regression results—year-wise

Total Regressions = 8	Dependent Variable : *wv*			Dependent Variable : *wv1*		
Independent Variables	Positive and significant[a]	Negative and significant	Not significant	Positive and significant	Negative and significant	Not signficant
Lagged-dependent variable	7		1	3		5
diff12		8			8	
diff23		8			8	
Hetrlang	1	3	4	1	4	3
Hetrrelg	2	1	5	2	2	4
hetrcast	2	4	2	1	3	4
Tcd		2	6	1	3	4
Tco	4		4	3	1	4
Tre	4	2	2	4	1	3
Rxdev	2	3	3	1	2	5
newspaper	2		6	1	1	6
Literacy		4	4		3	5
Rural	1	3	4		1	7
Sdp	1	1	6	1	2	5

We have run eight regressions, one for each election, starting from 1971. The entries denote the number of times the coefficient on a particular regressor was found positively or negatively significant or insignificant

[a] Coefficients are significant at least at the 90 % level

References

Abramson PR, Aldrich JH, Paolino P, Rohde DW (1992) "Sophisticated" voting in the 1988 presidential primaries. Am Polit Sci Rev 86(1):55–69

Alvarez RM, Nagler J (2000) A new approach for modelling strategic voting in multiparty elections. Brit J Polit Sci 30:57–75

Black JH (1978) The multicandidate calculus of voting: applications to canadian federal elections. Am J Polit Sci 22:609–638

Cain BE (1978) Strategic voting in Britain. Am J Polit Sci 22:639–655

Cox GW (1987) Duverger's law and strategic voting. Mimeo. University of Texas

Cox GW (1997) Making votes count: strategic coordination in the world's electoral systems. Cambridge University Press, New York

Chhibber P, Kollman K (1998) Party aggregation and the number of parties in India and the United States. Am Polit Sci Rev 92:329–342

Chhibber P, Murali G (2006) Duvergerian dynamics in the Indian states: federalism and the number of parties in the state assembly elections. Party Polit 12(1):5–34

Curtice J, Steed M (1988) Appendix 2. In: Butler D, Kavanagh D The British general election of 1987. St. Martin's Press, New York

Dunleavy P, Diwakar R (2011) Analysing multiparty competition in plurality rule election. Party Polit. doi: 10.1177/1354068811411026

Duverger M (1954) Political parties: their organization and activity in the modern state. Wiley, New York

Evans G, Heath A (1993) A tactical error in the analysis of tactical voting: a response to Niemi, Whitten and Franklin. Brit J Polit Sci 23:131–137

Fey M (1997) Stability and coordination in Duverger's law: a formal model of preelection polls and strategic voting. Am Polit Sci Rev 91(1):135–137

Galbraith J, Rae N (1989) A test of the importance of tactical voting: Great Britain, 1987. Brit J Polit Sci 19:126–136

Johnston RJ, Pattie CJ (1991) Tactical voting in Great Britain in 1983 and 1987. Brit J Polit Sci 21:95–128

Mckelvey RD, Ordeshook P (1972) A general theory of the calculus of voting. In: Herndon JF, Bernd JL Mathematical applications in political science, vol 6. University of Virginia Press, Charlottesville

Myerson RB, Weber RJ (1993) A theory of voting equilibria. Am Polit Sci Rev 87:102–114

Niemi RG, Whitten G, Franklin MN (1992) Constituency characteristics, individual characteristics and tactical voting in the 1987 British general election. Brit J Polit Sci 22:229–254

Ordeshook P, Shvetsova O (1994) Ethnic heterogeneity, district magnitude and the number of parties. Am J Polit Sci 38:100–123

Palfrey TR (1989) A mathematical proof of Duverger's law. In: Ordeshook PC Models of strategic choice in politics. University of Michigan Press, Ann Arbor, p 69–91

Part III
Macroeconomics and Finance

Chapter 21
Monetary Policy and Crisis

Chandana Ghosh and Ambar Nath Ghosh

1 Introduction

There is a debate over whether monetary policy is effective or not during periods of crisis. Keynes' and early Keynesians' perception was that in times of crises, conventional monetary policy does not work. This is because, in such periods, interest rate hits the lower bound. As a result, monetary policy fails to lower the cost of credit. In fact, in the 1930s and 1940s, interest rate in the USA was near zero and, hence, money might have been quite irrelevant at the margin. This explains early Keynesians' disregard of the monetary sector (see Krugman 1998 in this context). There is a strong view that monetary policy during the "lost decade" in Japan (1991–2000) and also in the USA in the recent financial crisis which started from 2007, may not have been effective (see in this connection Krugman (1998, 2008), Sawyer (2009) and Mishkin (2009) among others). In the context of the recent financial crisis in the USA, Krugman (2008) commented "we are already, however, well into the realm of what I call depression economics. By that I mean a state of affairs like that of the 1930s in which the usual tools of monetary policy—above all the Federal Reserve's ability to pump up the economy by cutting interest rates—have lost all traction". The Federal Open Market Committee (FOMC) in its minutes from the October 28–29, 2008, meeting (Board of Governors of the Federal Reserve System 2008) expressed serious misgivings regarding the effectiveness of policy of cuts in Federal funds rate. Apart from the conventional liquidity trap argument against monetary policy, a recession or a depression activates the mechanism of financial accelerator which raises cost of credit to firms and financial institutions even in the face of aggressive

C. Ghosh (✉)
Economic Research Unit, Indian Statistical Institute,
Kolkata 700108, West Bengal, India
e-mail: chandana@isical.ac.in

A. N. Ghosh
Department of Economics, Jadavpur University,
Kolkata 700032, West Bengal, India
e-mail: ang.juecon@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_21,

rate cuts by the central bank. This was in evidence in the financial crisis of the USA (Mishkin 2009). To know why this may happen, one can go through Bernanke and Gertler (1989) and Bernanke et al. (1996). Besides the mechanism noted above, there is another factor which seriously impedes the effectiveness of conventional monetary policy. This consists of interest inelasticity of investment brought about by the collapse of confidence in times of crisis. In such times, firms become pessimistic about profitable investment avenues and cease to undertake investment even if credit is made available at very low interest rates. In fact, as stated in Bernanke and Gertler (1995), empirical studies have not found any evidence in support of interest elasticity of investment even in normal times. In this connection, it may be instructive to quote what top officials in Fed told Fox Business network on 22 March 2012. Richard Fisher, Dallas Fed president stated, “We will not support further quantitative easing under these circumstances because there is a lot of money lying on the sideline, lying fallow. We do not need any more monetary morphine. The real problem in our country is job creation and prosperity. And we need to get better fiscal policy to complement what we at the Fed have done, because it is not working as effectively as it should.” In a similar vein, Bernenke, Fed chairman, said, “right now in terms of debt and consumption, we are still way low relative to the pattern before the crisis. . . . We lack of source of demand to keep the economy growing.” (These quotations are from Economic Times 24 March 2012.) The Fed chairman said this even when Fed cut the Federal funds rate to near zero and promised to keep it there till 2014. In the Eurozone also, recession is the major problem today even though the European Central Bank (ECB) has pumped in substantial liquidity into the banks through its long term refinance operations (LTRO) at very low interest rates. All these point to ineffectiveness of conventional monetary policy in times of crisis.

The objective of the present chapter is to point to a route different from those mentioned above, through which conventional monetary policy, instead of removing a crisis, propagates it. This mechanism operates through the adverse impact that rate cuts by central banks may have on economic agents’ expectation regarding the future course of the economy. This chapter argues that a rate cut by the central bank to counter recession gives out a strong signal that it apprehends continuation or deepening of the ongoing recession and that policy intervention is necessary to reverse the process. This may spread alarm and induce the economic agents to revise downward their expected future income. Under such circumstances, as the chapter shows, saving may vary inversely with interest rate and investment may cease to be interest sensitive or may even vary directly with interest rate. Thus, when there occurs an adverse exogenous shock that starts off a process of recession and the central bank resorts to conventional monetary policy of cutting rates to reverse the process, saving may go up but investment may go down or remain unchanged. This brings about further contraction leading to self-fulfilling expectation. This process may go on until the interest rate hits the lower bound.

2 The Model

In the current crisis that is ravaging the advanced part of the world, monetary policy, as we have stated above, has failed to bring about a turnaround. This chapter seeks to point to a mechanism other than those already postulated, which may render monetary policy counterproductive in tackling a crisis. This happens because aggregate saving may vary inversely with interest rate and investment may cease to be interest-sensitive or may even vary directly with interest rate when the central bank regulates interest rate to counter recessionary forces. The reason for this non-standard behaviour of saving and investment in times of crisis may be explained as follows. Let us focus first on saving. Saving function is given by

$$S = S\left(\underset{+}{Y}, \underset{+}{i}, \underset{+}{Y^e}\left(\underset{+}{i}\right)\right) \tag{21.1}$$

where $S \equiv$ aggregate saving, $Y \equiv$ GDP, $i \equiv$ interest rate and $Y^e \equiv$ expected future GDP.

Interest rate, i, these days is regulated by the central bank. Monetary policy of the central bank, a la Romer (2000), is given by

$$i = i\left(\underset{-}{\bar{Y} - Y}\right) \tag{21.2}$$

where $\bar{Y}$ is the natural rate of Y. Equation (21.2) says that lesser the Y relative to its natural rate, $\bar{Y}$, lower is the level at which the central bank sets its interest rate. The central bank obviously behaves in the manner shown in Eq. (21.2) when it apprehends the recession to continue or deepen, and feels that policy intervention is needed to reverse the process. Thus, a rate cut may act as a signal of the central bank's perception regarding the future performance of the economy, and this may induce a downward revision in private economic agents' expected future income. (Note that the central bank regulates interest rate to achieve both inflation and interest rate targets. However, we have omitted inflation rate from Eq. (21.2) for simplicity. Moreover, the purpose of this chapter is to examine whether monetary authority can tackle recession by cutting rates.) If there is inflation in times of recession (which can, of course, occur if there is a cost-push or structural imbalance), the monetary authority may not be able to cut rates. In such a case, the monetary authority is unable to adopt a monetary policy to tackle recession. We are not considering this scenario here. The objective here is to show that even when the central bank can cut rates to tackle recession, without being impeded by the fear of stoking inflationary forces, it may exacerbate recession. Thus, when an economy is in recession and the central bank cuts rates to reverse the situation, economic agents get the signal that the central bank is apprehensive of continuation or deepening of recession and revises downward their expected future income (all kinds of future income and not just future interest income from current saving). Of course, one may argue that a rate cut may generate the hope of a revival, but economic agents may not be convinced of such a scenario until signs of a turnaround become perceptible. The reason why people may not be willing to act on the basis of such a belief of revival is that they may be strongly averse to getting caught in an unexpected prolongation of recession with less than enough to fall back on or substantial unutilized new production capacity. If this line of thinking is true, a rate cut will induce a decline in future expected income of the economic agents. Another point to note is that the more aggressive the rate cut, the greater may be the fear generated regarding the enormity of the crisis. Thus, the more aggressive the rate cut, i.e. the greater the desperation shown by the central bank to counter a recession, the larger may be the decline in the expected future income of the economic agents.

If a rate cut, as we assume here, induces individuals to lower their expected future income, the income effect produced by a policy-induced decline in interest rate will be larger than what happens when a reduction in interest rate leaves expected future income (other than the future interest income from current saving) unaffected. The implication of the above discussion is that a decline in interest rate induced by a rate cut to tackle recession may raise current saving instead of lowering it.

For the reason stated above, a counter-cyclical policy-induced reduction in interest rate by engendering the fear of continuation or deepening of recession may fail to stimulate investment even if there is a decline in the borrowing cost. Thus, when interest rate is regulated by the central bank to counter recessions, investment function may be written as

$$I = I\left(\underset{-}{i}, \underset{+}{Y^e}\left(\underset{-}{i}\right)\right). \tag{21.3}$$

From Eq. (21.3), it is clear that a cut in the interest rate by the central bank will lower borrowing cost but it may, at the same time, worsen investors' expectation regarding future income. The former exerts a positive impact on investment, but the latter works just in the opposite direction. Thus, when interest rate is regulated by the central bank as a counter-cyclical policy measure, a rate cut may not be able to stimulate investment. It may even lower investment. For our purpose, it is sufficient to assume that investment ceases to be interest-sensitive, when the central bank regulates interest rate to counter cyclical fluctuations in gross domestic product (GDP). To demonstrate very clearly how conventional monetary policy of cutting rates for tackling recession may be counterproductive, we shall develop a simple dynamic model below. We shall take simple linear forms of the saving and investment functions. We shall, for reasons explained above, assume saving to be an increasing function of current income, but a decreasing function of interest rate. Again, for reasons explained above, we shall assume investment to be exogenously given. We shall dynamize the model by incorporating a lag in the monetary policy function, as represented by Eq. (21.2), of the central bank.

We consider an economy where prices are assumed to be fixed for simplicity, so that inflation considerations do not impede the freedom of the monetary authority in adopting counter-recessionary measures. For reasons specified earlier, we postulate the following saving function:

$$S_t = \bar{S} + sY_t\text{-}ai_t \quad \bar{S} < 0, \quad 1 > s > 0, \quad a > 0 \tag{21.4}$$

where $S_t \equiv$ aggregate saving in period t, $Y_t \equiv$ aggregate income or GDP in period t and $i_t \equiv$ interest rate in period t. The investment function, as we have discussed earlier, is given by

$$I_t = \bar{I}. \tag{21.5}$$

The goods market is in equilibrium when

$$\bar{S} + sY_t - ai_t = \bar{I}. \tag{21.6}$$

Instead of the LM, we rewrite the monetary policy function of the central bank given by Eq. (21.2) as follows:

$$\begin{aligned} &i_t = gY_{t-1} \quad g > 0 \quad \text{for} \quad Y_{\text{t-1}} > \frac{\bar{i}}{g} \\ &i_t = \bar{i} \quad \text{for} \quad Y_{t-1} \leq \frac{\bar{i}}{g}. \end{aligned} \tag{21.7}$$

The differences between Eqs. (21.2) and (21.7) are the following. We have omitted $\bar{Y}$ from Eq. (21.7) for simplicity. Our focus here is only on those values of Y_t which are less than $\bar{Y}$. We have also incorporated a one-period lag in the policy response. We have also specified a lower bound for i, denoted $\bar{i}$. Equation (21.7) enshrines the conventional monetary policy of the central bank. It states that lesser the value of Y_{t-1}, lower is the value at which the central bank fixes the interest rate in period t until interest rate hits the lower bound $\bar{i}$. Combining Eqs. (21.6) and (21.7), we can write down the equilibrium condition of the economy as follows:

$$\begin{aligned} &\bar{S} + sY_t - agY_{t-1} = \bar{I} \quad \text{for} \quad Y_{t-1} > \frac{\bar{i}}{g} \\ &\bar{S} + sY_t - a\bar{i} = \bar{I} \quad \text{for} \quad Y_{t-1} \leq \frac{\bar{i}}{g}. \end{aligned} \tag{21.8}$$

We can solve Eq. (21.8) for Y_t as a function of Y_{t-1}:

$$Y_t = \frac{\bar{I} - \bar{S}}{s} + \frac{ga}{s}Y_{t-1} \quad \text{for} \quad Y_{t-1} > \frac{\bar{i}}{g} \tag{21.9a}$$

$$Y_t = \frac{\bar{I} - \bar{S}}{s} + \frac{a}{s}\bar{i} \quad \text{for} \quad Y_{t-1} \leq \frac{\bar{i}}{g} \tag{21.9b}$$

Let us explain Eqs. (21.9a) and (21.9b). Focus on Eq. (21.9a) first. Corresponding to any given value of Y_{t-1}, the excess of investment over saving at $Y_t = 0$ is given

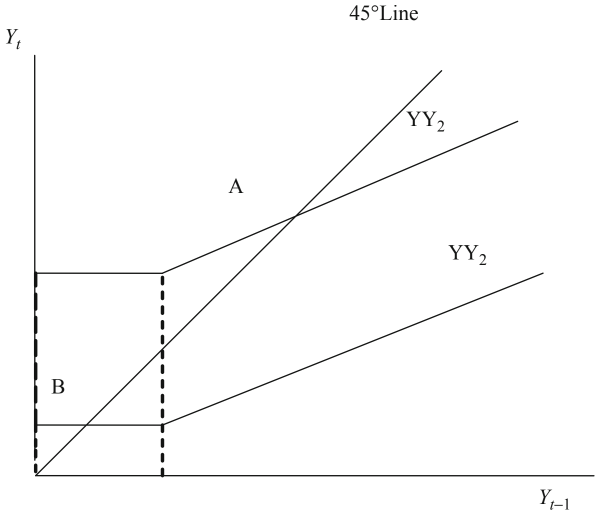


Fig. 21.1 Steady state and crisis

by $\bar{I} - (\bar{S} - gaY_{t-1})$—see Eq. (21.8). Hence, the value of Y_t that equates saving and investment is given by the right-hand side of Eq. (21.9a). Following a unit increase in Y_{t-1}, the excess of investment over saving at $Y_t = 0$ rises by ga. Hence, the value of Y_t that equates saving and investment goes up by (ga/s). For stability of the steady-state value of Y as yielded by Eq. (21.9a), we assume (ga/s) to be less than unity. Thus, $0 < (ga/s) < 1$. Let us now consider Eq. (21.9b). For $Y_{t-1} \leq (\bar{i}/g)$, interest rate equals $\bar{i}$. Hence, for any such value of Y_{t-1}, the excess of investment over saving at $Y_t = 0$ equals $\bar{I} - (\bar{S} - a\bar{i})$—see Eq. (21.8). Hence, the value of Y_t that equates saving and investment in period t is given by $\left[\left\{\bar{I} - (\bar{S} - a\bar{i})\right\}/s\right]$.

It follows from Eqs. (21.9a) and (21.9b), as we shall explain shortly, that if the value of Y_t that Eq. (21.9b) gives is less than or equal to $(\bar{i}/g)$, the steady-state value of Y will be given by Eq. (21.9b). Let us explain. If Y_{t-1} increases from $(\bar{i}/g)$, then per unit increase in Y_{t-1} as follows from Eq. (21.9a) Y_t will rise by (ga/s), which we have assumed to be less than unity. Hence, per unit increase in Y_{t-1}, the excess of Y_{t-1} over Y_t will increase by $1 - (ga/s)$. Hence, steady state cannot occur at any $Y_{t-1} > (\bar{i}/g)$. However, since the unique value of Y_t given by Eq. (21.9b) is less than or equal to $(\bar{i}/g)$ for every $Y_{t-1} \leq (\bar{i}/g)$, the steady-state value of Y will be given by Eq. (21.9b).

The situation is shown in Fig. 21.1 by the schedule YY_2 in the (Y_{t-1}, Y_t) plane. The horizontal portion of the YY_2 schedule represents Eq. (21.9b), while the upward sloping part represents Eq. (21.9a). The steady state lies on the horizontal part. It occurs exactly at that point at which the horizontal part intersects the 45° line. This point is labelled B in the figure. The situation discussed above and delineated by YY_2 obtains when

$$\frac{\bar{i}}{g} \geq \frac{\bar{I} - \bar{S}}{s} + \frac{a}{s}\bar{i}. \tag{21.10}$$

The above inequality is satisfied when, given other variables, s is sufficiently large and $\bar{I}$ is sufficiently small.

When Eq. (21.10) is not satisfied, the value of Y_t that Eq. (21.9b) gives exceeds $(\bar{i}/g)$. Obviously, in this case, the steady state cannot occur at any $Y_{t-1} \leq (\bar{i}/g)$. If Y_{t-1} rises from $(\bar{i}/g)$, then, as follows from Eq. (21.9a), per unit increase in Y_{t-1}, the value of Y_t will go up by $(ga/s) < 1$ by assumption. Therefore, the excess of Y_t over Y_{t-1} will go down by $(1 - (ga/s))$ per unit increase in Y_{t-1}. Hence, the steady state will occur at some $Y_{t-1} > (\bar{i}/g)$. This steady state will be yielded by Eq. (21.9a). The situation is captured in Fig. 21.1 by YY_1 schedule. The horizontal part of this schedule represents Eq. (21.9b), while the upward sloping part represents Eq. (21.9a). The steady state occurs on the upward sloping part of YY_1. The steady state corresponds to the point of intersection of the upward sloping part of the YY_1 schedule and the 45° line. This point is labelled A in the figure.

We are now in a position to show how conventional monetary policy aggravates crisis. Suppose the economy was at point A initially. Suppose economic agents, because of some extraneous factors, become pessimistic regarding the future so that $\bar{I}$ falls. In consequence, YY_1 shifts down to YY_2. The economy will, therefore, move from point A to point B leading to self-fulfilling pessimism. The conventional monetary policy of cutting interest rates in the face of declining output plays a crucial role in the economy's journey from A to B. With the decline in autonomous investment, Y falls from its steady-state value. To stem or reverse this decline, the central bank cuts interest rate, which worsens expectations, and thereby, leads to an increase in saving. Y, therefore, contracts more and this process goes on until interest rate hits the lower bound. Had there been no interest rate cutting, i.e. had interest rate been kept unchanged, the fall in Y would have been smaller. It would have been given by the distance between A and the point lying vertically below it on YY_2.

Let us demonstrate this result mathematically. For simplicity, suppose the lower bound of interest rate is zero. This means that the interest rate hits the lower bound at $Y_{t-1} = 0$—see Eq. (21.7). The implication of this assumption is that the evolution of Y over time is given only by Eq. (21.9a). Denoting the steady-state value of Y by $\bar{Y}$ and substituting it in Eq. (21.9a), we get

$$\bar{Y} = \frac{\bar{I} - \bar{S}}{s} + \frac{ga}{s}\bar{Y} \Rightarrow \bar{Y} = \frac{\bar{I} - \bar{S}}{s - ga}. \tag{21.11}$$

We can use Eq. (21.11) to examine how the steady-state value of Y changes when the central bank keeps the interest rate unchanged as well as when the central bank adjusts the interest rate in accordance with Eq. (21.7). For this purpose, we rewrite Eq. (21.11) as

$$\bar{S} + s\bar{Y} - ag\bar{Y} = \bar{I}. \tag{21.12}$$

First, consider the case where interest rate is kept unchanged. In this case, we can derive the change in $\bar{Y}$ by treating the third term on the left-hand side of Eq. (21.12) as constant. Thus,

$$d\bar{Y} = \frac{d\bar{I}}{s}. \tag{21.13}$$

Now focus on the second case. In this case, the third term on the right-hand side of Eq. (21.12) is no longer a constant, and from Eq. (21.12) we find

$$d\bar{Y} = \frac{d\bar{I}}{s - ga}. \tag{21.14}$$

Absolute value of Eq. (21.14) is obviously greater than that of Eq. (21.13). Let us explain the reason. Following a decline in $\bar{I}$, saving exceeds investment at the initial steady-state value of Y by the absolute value of the numerator of either Eq. (21.13) or (21.14). Steady-state Y will, therefore, contract. In the case interest rate is constant, saving declines by s per unit decrease in Y_{t-1} and Y_t from their initial steady-state value. But in the other case, per unit fall in Y_{t-1} and Y_t from their initial steady-state value saving declines by s because of the unit decrease in Y_t. But the unit decrease in Y_{t-1} induces the central bank to lower interest rate by g. This raises saving by ga. Thus, in the net per unit decline in Y_{t-1} and Y_t from their initial steady-state value saving falls by s-ga, which is less than s. Hence, the steady-state value of Y in the second case has to fall more to restore equality between saving and investment. This leads to the following proposition:

Proposition 1: The contractionary forces unleashed by an adverse exogenous shock in times of crisis may be reinforced by the central bank's conventional monetary policy of interest rate cutting in the face of declining output.

3 Conclusion

Even though the central banks in advanced countries such as the USA and Eurozone have reduced the interest rate to near zero and promised to keep the interest rate at such a level for a long period, recessions in those countries are hardly abetting. In fact, in many of these countries, economic conditions are getting worse. Monetary policy, therefore, seems to be a failure. This chapter seeks to explain why monetary policy by itself not only may fail to stem recessionary forces but also tend to reinforce them. The argument of the chapter runs as follows. If, following an adverse exogenous shock and the consequent onset of recession, the central bank cuts rates, it may produce a perverse outcome through its impact on people's expectations. A rate cut by the central bank sends out a strong signal that the central bank is apprehending either continuation or deepening of the ongoing recession and it considers policy intervention to be necessary for the reversal of the process. This may cause alarm and induce economic agents to revise downward their expected future income. Of course, the rate cut may generate a hope for revival, but people may not be willing to act on the basis of such a belief, and thereby, get caught in an unexpected prolongation of recession with less than enough to fall back on or substantial unutilized new production capacity. Hence, a rate cut is likely to induce worsening of expectations regarding future performance of the economy, and the economic agents may be prepared to reverse their expectations only when signs of a revival become perceptible. If this line of argument is true, the rate cut may raise saving and leave investment unchanged or even lower it. There will, therefore, be further contraction fulfilling people's expectations. The whole point of this chapter is to argue that monetary policy alone is weak and unreliable in a situation of recession. It should be accompanied by strong fiscal policy. Actually, a recession should be viewed as an opportunity for the government to invest aggressively in infrastructure and finance it not by borrowing from the public but by money creation. In the presence of large-scale excess capacity and unemployment, such a policy will not generate inflation but restore investors' and other economic agents' confidence by swelling the order books of firms and by easing up future supplies of crucial infrastructural inputs. It will not lead to the problem of sustainability of public debt either. Independence of the central bank from the treasury and the bogey of solvency of the central bank seriously impede the scope for fiscal policy, and are at the root of the current global crisis. These issues need a fresh look and we reserve it as our future research agenda.

References

Bernanke B, Gertler M (1989) Agency costs, net worth and business fluctuations. Am Econ Rev 79:14–31 (March)

Bernanke B, Gertler M (1995) Inside the black box: the credit channel of monetary policy transmission. J Econ Perspect 9(4):27–48 (Fall)

Bernanke B, Gertler M, Gilchrist S (1996) The financial accelerator and flight to quality. Rev Econ Stat 78(1):1–15 (February)

Board of Governors of the Federal Reserve System (2008) Minutes of the Federal Open Market Committee. October 28–29, 2008. http://www.federalreseserve.gov/newsevents/press/monetary/fomcminutes20081029.pdf

Krugman PR (1998) It's baaack: Japan's slump and the return of the liquidity trap. Brookings Pap Eco Ac 29(2):137–205

Krugman PR (2008) Depression economics returns. New York Times, November 14

Mishkin SF (2009) Is monetary policy effective during financial crises? Am Econ Rev 99(2):573–577 (May)

Romer D (2000) Keynesian macroeconomics without the LM curve. J Econ Perspect 14(2):1449–1169 (Spring)

Sawyer M (2009) Fiscal and interest rate policies in the "new consensus" framework: a different perspective. J Post Keynesian Ec 31(4):549–565 (Summer)

Chapter 22
An Effective Demand Model of Corporate Leverage and Recession

Rilina Basu and Ranjanendra Narayan Nag

1 Introduction

A financial crisis refers to a loss of confidence in a country's financial assets. A crisis can arise in a variety of situations in which assets suddenly lose a large part of their value. Since the last two quarters of 2008, the world economy has witnessed a financial meltdown coupled with recessionary tendencies. Unemployment rate has been climbing up and any sign of immediate business prospect is not in sight. Some experts, notably Krugman (1999, 2003, 2008), attributed this crisis to high degree of leverage.[1] The basic idea is rather simple. Debt is used for investment in risky assets rather than investing in physical capital. Many firms, in recent times, are guided by speculative motives to earn quick returns on risky assets and, thus, they have borrowed very aggressively. In a deregulated financial system without adequate supervision, the whole economy becomes critically sensitive to fluctuations in asset market. If market sentiment changes suddenly and the credibility come under cloud, crisis can occur and spread very quickly with devastating effects on the real sector. Figure 22.1 shows declining growth rates of output across the world.

Let us consider some data to reflect on the nature and implications of crisis in the Indian context. Sensex, which had touched the 21,000 mark on 8 January 2008, dropped down to as low as 8,000 by the last quarter of the same year as is evident from Figs. 22.2 and 22.3.[2] Total net capital flows fell from US$ 17.3 billion in

[1] The leverage factor played a major role in the recent crisis related to the IT sector in India.
[2] The shock was initially felt in the real estate business and quickly spread to IT sector and finally across the rest of the economy.

R. Basu (✉)
Department of Economics, Rabindra Bharati University,
Kolkata 700050, West Bengal, India
e-mail: banerjee.rilina@gmail.com

R. N. Nag
St. Xavier's College (Autonomous), Kolkata 700016, West Bengal, India
e-mail: rnnag12@rediffmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_22, © Springer India 2014

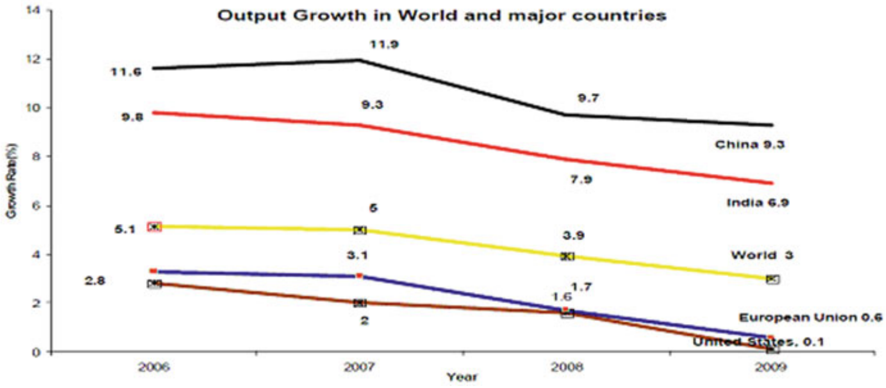


Fig. 22.1 Output growth in world and major countries. (Source: World Economic Outlook 2008)

April–June 2007 to US$ 13.2 billion in April–June 2008 (Mohan 2008). The 9.3 % growth rate of industries in the first half of 2007–2008 dropped down to 7.9 % in the second half. It further declined to 6.5 % in the first half of 2008–2009.[3] Although most of the Indian banks have limited exposure to global financial markets, the incidence of non-performing asset (NPA) looms large which is compounded by the withdrawal of foreign funds. External commercial borrowings of the corporate sector declined from US$ 7 billion in April–June 2007 to US$ 1.6 billion in April–June 2008, not only partially in response to policy measures in the face of excess flows in 2007–2008 but also due to the current turmoil in advanced economies (Mohan 2008). In this context, we note that there has been a 38.3 % growth in the volume of debt of major companies with sales more than Rs. 100 crore between 2005–2006 and 2006–2007 (Dey 2007). This is a clear indication of a very high degree of leverage of Indian companies.

There is an extensive and rapidly expanding literature on financial crisis.[4] This chapter is motivated by the role of balance sheet effect in the third-generation models of currency crisis.[5] One central feature of the balance sheet is the borrower's net worth, defined as the borrower's own funds plus the collateral value of his illiquid assets. If this net worth falls below a threshold value, the borrower may not be able to generate funds any more (Bernanke 1995). It, thus, becomes a compulsion of these companies to display an inflated balance sheet with a fundamental mismatch of assets and debts in term of maturity, called maturity mismatch (Roubini and Setser 2006).

[3] Source: Central Statistical Organisation. Available at: http://www.epwrf.res.in.

[4] In first-generation models, the collapse of financial market has been traced as the unsustainable fiscal policies. The classic first-generation models are those of Krugman (1979) and Flood and Garber (1984). The second-generation models by Obstfeld (1994, 1996) stated the conflict between a fixed exchange rate and the desire to pursue a more expansionary monetary policy as the cause of financial crisis. This genre of models exhibited multiple expectations, hence demonstrating speculative attacks being caused by self-fulfilling expectations.

[5] Krugman (1999, 2003, 2008) emphasised the role of companies' balance sheets in determining their ability to invest. In Calvo's (1998) model, the source of crisis is rooted in balance sheets of financial intermediaries.

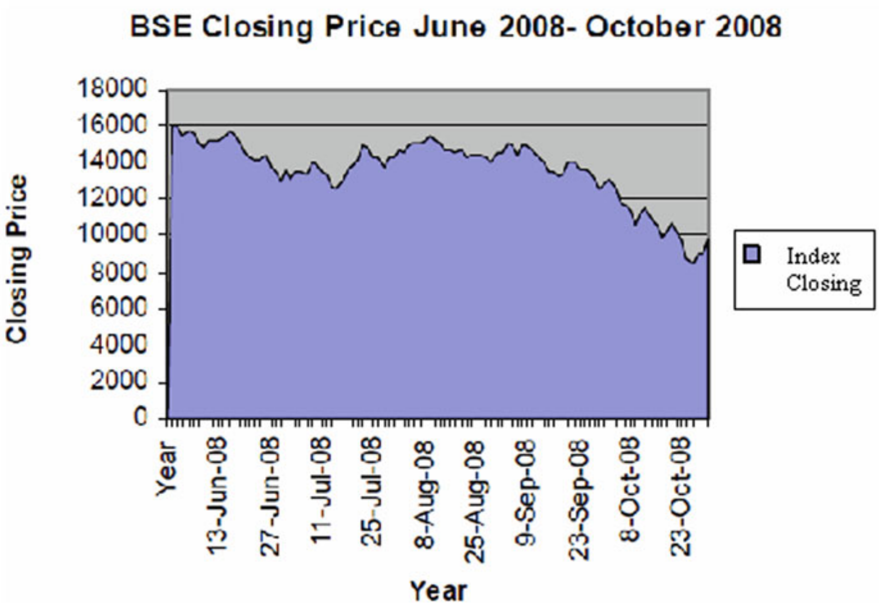


Fig. 22.2 Bombay Stock Exchange (BSE) closing price from June 2008 to October 2008. (Source: Malhotra 2008)

In the present scenario, the problem can be traced to be one of aggressive investment of highly leveraged firms and financial intermediaries in assets to earn significant return. Balance sheet channels may arise from the influence of monetary policy on the costs of intermediation, on the value of assets that are used as collaterals or through the risk externalities associated with financial intermediation (Knoop 2008). With a fall in the price of the risky asset and very bleak prospects of recovery, the balance sheets of these companies have been worst hit, and crisis becomes an inevitable outcome of self-fulfilling expectations.

Financial deregulation can generate myriad set of forces, which, without proper regulation, can lead to crisis in the form of sharp decline in the asset price. This produces an adverse effect on effective demand, and recession occurs through specific transmission mechanisms, which this chapter deals with. This chapter extends Krugman's model[6] not only to identify the trigger of the crisis but also to account for the recession that follows in the trail of the crisis. This chapter clearly demonstrates that the solution to the problem needs fiscal expansion, since the roots of the problem do not lie in liquidity.

[6] Specifically, we embed Krugman's partial equilibrium model in an overall macroeconomic structure.

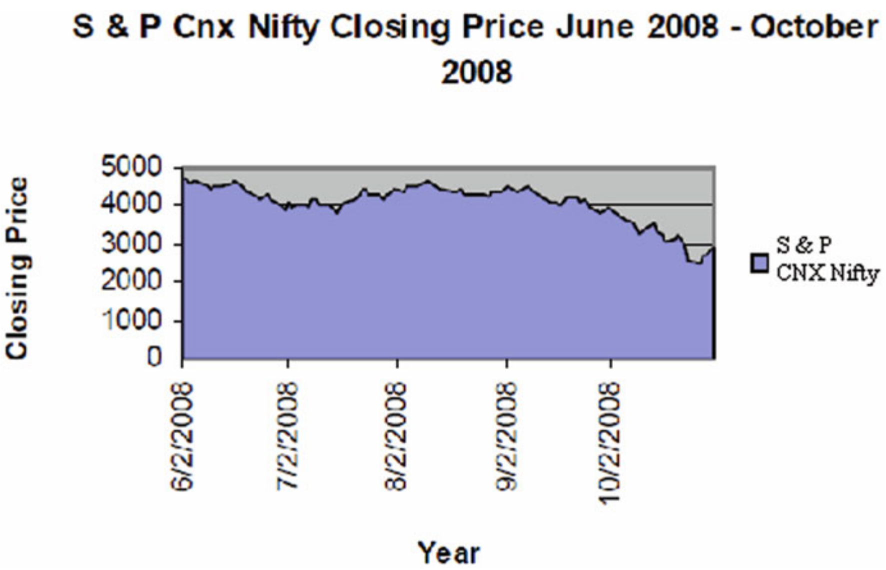


Fig. 22.3 S&P CNX Nifty closing price from June 2008 to October 2008. (Source: Malhotra 2008)

An important aspect of this model is the existence of endogeneity of money.[7] This chapter explains how money supply adjusts to clear the money market in response to macroeconomic developments. Once the real balance effect on consumption is incorporated, additional channels of transmission from financial to real sector open up.

The model is based on effective demand principle. In this context, we note that asset price volatility has not been accompanied by very major inflationary (deflationary) tendencies, not only in the Western world but also in some emerging market economies. However, the financial meltdown has led to loss of employment and output. Thus, the recent trends justify our choice of an effective demand model which incorporates price stickiness.

This chapter is organised as follows. Section 2 develops the basic model. Section 3 offers a comparative static analysis followed by concluding remarks in Sect. 4.

2 Model

There are three economic agents in the model, namely firms, banks and households. First, we present consolidated balance sheet of the economy:

[7] There is an extensive literature on endogeneity of money. In real business cycle models, money is primarily endogenous. The Neo-Keynesian theories also strongly argue in favour of endogeneity of money. Ireland (2005) suggests that the Central Bank and the banking system would have to supply money to clear money market.

Assets	Liabilities
FIRMS	
Risky Asset	Debt
Physical Asset	Equity
HOUSEHOLDS	
Risky Asset	Wealth
Equity	
Money	
COMMERCIAL BANKS	
Debt	Deposits
Reserves	
CENTRAL BANK	
Domestic credit	High Powered Money

In this context, we note that the value of the risky asset is measured at demand price of capital. However, physical capital is measured at replacement cost and, for all practical purposes, we take commodity price to represent replacement cost. It follows from the balance sheet that firms are highly leveraged and they use leverage to invest in risky asset.[8]

The following symbols will be used in this chapter:

Y	Output
C	Private consumption
T	Lump sum tax
I	Investment expenditure
q	Tobin's q
G	Government expenditure
λ	Maximum leverage ratio of the firms
N	Initial holding of the risky asset by the firms
D	Current debt
A	Fixed supply of the asset
H	Demand function of the households
q^e	Expected asset price
$\frac{q^e}{q} - 1$	Expected change in the price of the risky asset
β	Share of profit which is distributed as dividend
Π	Total profit
E	Number of equities outstanding
M	Nominal money stock
P	Fixed supply price of capital
L	Money demand function

[8] One can perceive of such firms as producer of financial products with limited investment in physical capital.

The model is represented by the following equations:

$$Y = C\left(\mathrm{Y} - \mathrm{T}\frac{M}{P}\right) + \mathrm{I(q)} + \mathrm{G}, I' > 0,\ I(1) = 0,$$
$$< C_1 = \frac{\partial C}{\partial (Y - T)} < 1, \quad C_2 = \frac{\partial C}{\partial\left(\frac{M}{P}\right)} > 0 \tag{22.1}$$

$$\lambda\left[\mathrm{N} - \frac{D}{q}\right] + \mathrm{H}\left(\frac{q^e}{q} - 1, \frac{\beta\pi(\mathrm{Y})}{E}\right) = \mathrm{A},$$
$$\mathrm{H}_1 = \frac{\partial H}{\partial\left(\frac{q^e}{q} - 1\right)} > 0,\ H_2 = \frac{\partial H}{\partial\left(\frac{\beta\pi(Y)}{E}\right)} < 0 \tag{22.2}$$

$$\mathrm{M} = \mathrm{L}\left(\mathrm{Y}\frac{q^e}{q} - 1, \frac{\beta\pi(Y)}{E}\right), \quad L_1 = \frac{\partial L}{\partial Y} > 0,$$
$$\mathrm{L}_2 = \frac{\partial L}{\partial\left(\frac{q^e}{q} - 1\right)} < 0, \quad \mathrm{L}_3 = \frac{\partial L}{\partial\left(\frac{\beta\pi(Y)}{E}\right)} < 0. \tag{22.3}$$

Equation (22.1) represents the commodity market. Industrial output (Y) is demand determined and the different components of aggregate demand are consumption, investment and government expenditure. We incorporate the real balance effect in the consumption function. Investment depends positively on q.[9] Government expenditure (G) is exogenously given.

Equation (22.2) represents market clearing for a risky asset.[10] There are two sources of demand for the risky asset[11]—household's demand and the firms' demand. The household's demand for risky asset depends on its own returns and return on equity.[12] The balance sheet[13] effect is incorporated in the model through financial leverage[14] and, accordingly, we get firm's demand for the risky asset. Given λ, the

---

[9] q is the ratio of demand price of capital to the supply price of capital. Since we consider a composite good, supply price of capital is the commodity price itself and, thus, q equals the demand price of capital which, in Krugman's model, is interpreted as asset price.

[10] We simply extend representation of asset market equilibrium in Krugman's (2008) model.

[11] This type of risky assets can be thought to be mortgaged-based securities like real estate as cited by Krugman (2008).

[12] Return on equity is measured in terms of dividend per equity which is a function of the profitability of the firm. As profit increases, the firms give away β proportion of its profit as dividend to the equity holders. In this case, real profit is identical to nominal profit as P = 1.

[13] The asset side of the firm includes the risky asset and the physical capital where as the liability includes both debt and equity. So from the balance sheet of the firm E + D = $P_k{}^d$N + pK.

[14] The leverage factor is given by $\lambda = \frac{1}{1 - \frac{D}{p^d{}_k N + pk}} = \frac{P_k{}^d N + PK}{E} = \frac{qN + K}{E}$, since p = 1.

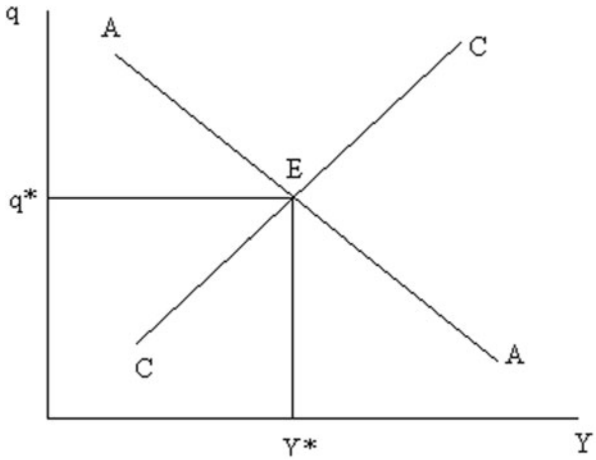


Fig. 22.4 Equilibrium in the model

firms' demand for the risky asset is given by $N^d = \lambda\left[N - \frac{D}{q}\right]$.[15] The asset is fixed in supply. Thus, Eq. (22.2) simply represents equality between given supply of risky asset and desired holdings of the asset from leveraged firms and the general public.[16]

Equation (22.3) represents money market equilibrium which endogenously determines money supply. The demand for money depends on output and different asset returns, namely expected change in price of the risky asset and the dividend per unit of equity. Demand for money falls as return on other assets rises and, thus, $L_2 < 0$, $L_3 < 0$. An increase in Y causes transaction demand for money to rise. On the other hand, output expansion causes dividend to increase and, hence, people shift towards equity. We assume that demand for money rises with an increase in output.[17]

Equation (22.2) can be rewritten as

$$\lambda\left[N - \frac{D}{q}\right] = A - H\left(\frac{q^e}{q} - 1, \frac{\beta\pi(Y)}{E}\right). \tag{22.4}$$

From Eqs. (22.1) and (22.2), we get output and price of the risky asset. Equation (22.3) determines the amount of money required to clear the money market.

Let us now undertake a diagramatic representation of the model. From Eq. (22.1), we get the CC curve which is positively sloped: As q increases, investment increases with a rise in Y. From Eq. (22.4), we get AA curve. As Y increases, dividend per unit of equity goes up and, hence, general public shift out of risky asset. The fall in demand causes q to fall. This is illustrated in Fig. 22.4.

---

[15] The value of the asset is given by: $qN^d = \lambda E - K = \lambda\,[qN + K - D - K]$, or $N^d = \lambda\left[N - \frac{D}{q}\right]$.

[16] For stability condition, we require $\frac{\partial ED}{\partial q} < 0$, i.e. $\lambda D < H_1 q^e$.

[17] $\frac{\delta L}{\delta Y} = L_1 + L_3\frac{\beta\pi'}{E} > 0$, where $L_1 > 0$, $L_3 < 0$.

3 Comparative Static Analysis

First, this chapter explains how a self-fulfilling crisis can occur. This is modelled by a fall in expected q. We will discuss how dividend policy of corporations work and how interventionist policies can be chosen to avert off crisis of this type.

3.1 Self-Fulfilling Crisis

In this chapter, self-fulfilling crisis is interpreted in Krugman's (2003, 2008) sense. The model shows how actual asset price declines and recession occurs. A fall in q^e reduces household sector's demand for the risky asset. This is represented by a leftward shift of AA curve. The resulting increase in the supply of risky asset to the highly leveraged institutions reduces q. Investment declines and output falls. The recession may be further aggravated by fall in money and the associated fall in consumption through the real balance effect. However, CC curve will not stay put. Depending on the direction in which money will change, CC curve will shift. In case of a fall in money, the recessionary tendencies will be exacerbated through the real balance effect on consumption.

The results of the earlier mentioned comparative static exercise are as follows:

$$\frac{\partial Y}{\partial q^e} = \frac{I_q H_1 q^2 - C_2 L_2 \lambda D}{q^3 \Delta} > 0$$

$$\frac{\partial q}{\partial q^e} = \frac{(1 - C_1)H_1 + C_2 L_2 a - C_2 H_1 c}{q\Delta} > 0^{18}$$

$$\frac{\partial M}{\partial q^e} = \frac{(1 - C_1)(H_1 d - bL_2) + I_q(aL_2 - cH_1)}{\Delta} > 0^{19}$$

where $$\Delta = \begin{pmatrix} 1 - C_1 & -I_q & -C_2 \\ a & b & 0 \\ c & d & -1 \end{pmatrix} > 0.^{20}$$

The preceding discussion leads to the following proposition.

Proposition 1 Self-fulfilling financial crisis not only represents fall in asset price, but also generates demand deficiency leading to output and employment loss.

---

[18] $(1 - C_1)H_1 + aC_2L_2 > C_2H_1c$

[19] If $H_1(L_1 + L_3) < L_2H_2 \dfrac{\beta\pi_y}{E}$

[20] See Mathematical Appendix for detailed calculations. The notations in the matrix are $a = \dfrac{\beta}{E}H_2\pi_y < 0, b = \dfrac{\lambda D - q^e H_1}{q^2} < 0, \quad c = L_1\dfrac{\beta}{E}L_3\pi_y > 0, \ d = -L_2\dfrac{q^e}{q^2} > 0.$

3.2 Dividend Policy of the Firm

Rise in dividend may not work well in the presence of high degree leverage.[21] In the present model, we represent a dividend policy by a parametric increase in β. This causes a portfolio diversification in favour of equity away from the risky asset. Hence, q falls, investment declines and recession is round the corner. This can be represented by a leftward shift of the AA curve.

The results of the earlier mentioned comparative static analysis are as follows:

$$\frac{\partial Y}{\partial \beta} = \frac{\pi(Y)(-I_q H_2 + C_2 H_2 d - C_2 L_3 b)}{E\Delta} > 0^{22}$$

$$\frac{\partial q}{\partial \beta} = \frac{\pi(Y)\big((1 - C_1)H_2 + aC_2 L_3 - H_2 c\big)}{E\Delta} > 0^{23}$$

$$\frac{\partial M}{\partial \beta} = \frac{\pi(Y)((dH_2 - bL_3)(1 - C_1) + I_q(-cH_2 + aL_3)}{E\Delta} > 0.^{24}$$

The effects of dividend policy are summed up in proposition 2.

Proposition 2 An increase in dividends could intensify the recession in the presence of high degree of financial leverage used to finance investment in risky asset.

3.3 Expansionary Fiscal Policy

Once recession is entrenched, Keynesianism is back in the sense that the Keynesian recommendations work perfectly well. As suggested in the introduction, the real problem is not liquidity crunch, but demand deficiency. The model shows how an expansionary fiscal policy can mitigate the contractionary effects of a financial crisis. Fiscal expansion generates demand: It causes a rightward shift of the CC curve. Output expansion increases dividend, which induces individuals to move out of risky asset and, hence, q falls which causes investment to decline. Thus, fiscal expansion may not completely reverse the original shock but can undo the adverse balance sheet effect of that shock.

[21] The newly elected US President Barrack Obama raised concern against increasing dividends by highly leveraged firms given the current financial scenario.

[22] Given the condition $H_2(C_2 d - I_q) > L_3 b C_2$

[23] Given the condition $aC_2 L_3 > (1 - C_1 + c)H_2$

[24] Given the condition $I_q(aL_3 - cH_2) > (1 - C_1)(dH_2 - bL_3)$

The result of the earlier mentioned comparative static analysis is as follows:

$$\frac{dY}{dG} = \frac{H_1 q^e - D}{q^2 \Delta} > 0$$

$$\frac{dq}{dG} = \frac{H_2 \beta \pi_y}{E \Delta} < 0$$

$$\frac{\partial M}{\partial G} = \frac{ad - bc}{\Delta} > 0.$$ [25]

The comparative static analysis rise in government expenditure leads to proposition 3.

Proposition 3 Fiscal expansion could mitigate the recession that financial crisis generates.

3.4 Policy of Bailout

Though bail out itself creates moral hazard, which lies at the root of currency crisis, it is an immediate measure to avert off financial collapse and loss of confidence. The working of bail out in this model is as follows. Suppose that leveraged corporations receive some financial support from treasury. Let X be the value of financial assistance which a bailout programme involves. Accordingly, Eq. (22.4) can be written as

$$\lambda\left[\mathrm{N} - \frac{D}{q} + X\right] = \mathrm{A} - \mathrm{H}\left(\frac{q^e}{q} - 1, \frac{\beta\pi(Y)}{E}\right).$$

It causes an upward shift of the AA curve. Thus, a bailout programme arrests the trend of declining asset price by generating demand for assets.[26]

Thus, we get the following proposition.

Proposition 4 A bailout programme is inevitable in view of financial crisis of an alarming proportion.

4 Conclusion

The present crisis has two dimensions, namely plummeting asset prices and economy-wide recession. This chapter shows how self-fulfilling financial crisis occurs due to high degree of leverage and unsustainable debt to equity ratio of corporate

[25] Given the condition ad > bc

[26] However, bailout should not be a permanent feature of a system in the presence of high degree of financial mobility. What is required is appropriate regulatory framework to govern and regulate the direction of flow of funds.

sector. We also examine how financial crisis precipitates recession and unemployment, and, accordingly, we construct an effective demand model.

The message of this chapter is very clear. Excessive risk-taking by the corporate sector lies at the roots of the crisis and state intervention at specific nodal points is required for the economy to tide over this crisis. A higher dividend may not work and capital injection by the government should be an integral component of a rescue package. More specifically, IMF-style stabilization programme would turn out to be a perilous undertaking.

This chapter can be extended in different directions. The model in this chapter is essentially static and is a good starting point. One can introduce perfect foresight and examine behaviour of asset price over time. The model can be further extended to capture global contagion.

Acknowledgement We are grateful to Bhaskar Goswami for his comments on a preliminary draft of the paper. However, the usual disclaimer applies.

5 Appendix

The model is represented by the following three equations:

$$Y = C\left(Y - T, \frac{M}{P}\right) + I(q) + G \tag{22.5}$$

$$\lambda\left[N + \frac{K}{q} - \frac{D}{q}\right] = A - H\left(\frac{q^e}{q} - 1, \frac{\beta\pi(Y)}{E}\right) \tag{22.6}$$

$$M = L\left(Y, \frac{q^e}{q} - 1, \frac{\beta\pi(Y)}{E}\right). \tag{22.7}$$

a. Self-fulfilling expectations

To get the effect of a fall in q^e, using chain rule, we differentiate Eqs. (22.5)–(22.7) with q^e and form the following matrix:

$$\begin{pmatrix} 1 - C_1 & -I_q & -C_2 \\ a & b & 0 \\ c & d & -1 \end{pmatrix} \begin{pmatrix} \dfrac{\partial Y}{\partial q^e} \\ \dfrac{\partial q}{\partial q^e} \\ \dfrac{\partial M}{\partial q^e} \end{pmatrix} = \begin{pmatrix} 0 \\ -\dfrac{H_1}{q} \\ -\dfrac{L_2}{q} \end{pmatrix},$$

where $a = \frac{\beta}{E} H_2 \pi_y < 0,\ b = \frac{\lambda D - q^e H_1}{q^2} < 0,\ c = L_1 + \frac{\beta}{E} L_3 \pi_y > 0,\ d = -L_2 \frac{q^e}{q^2} > 0.$

Solving the above equations using Cramer's rule, we get

$$\frac{\partial Y}{\partial q^e} = \frac{\begin{pmatrix} 0 & -I_q & -C_2 \\ -\dfrac{H_1}{q} & b & 0 \\ -\dfrac{L_2}{q} & d & -1 \end{pmatrix}}{\Delta} = \frac{I_q H_1 q^2 - C_2 L_2 \lambda D}{q^3 \Delta} > 0, \text{ where}$$

$$\Delta = \begin{pmatrix} 1 - C_1 & -I_q & -C_2 \\ a & b & 0 \\ c & d & -1 \end{pmatrix} > 0,$$

$$\frac{\partial q}{\partial q^e} = \frac{\begin{pmatrix} 1 - C_1 & 0 & -C_2 \\ a & -\dfrac{H_1}{q} & 0 \\ c & -\dfrac{L_2}{q} & -1 \end{pmatrix}}{\Delta}$$
$$= \frac{(1 - C_1)H_1 + C_2 L_2 a - C_2 H_1 c}{q\Delta} > 0,$$

given the condition $(1 - C_1)H_1 + aC_2L_2 > C_2H_1c$,

$$\frac{\partial M}{\partial q^e} = \frac{\begin{pmatrix} 1 - C_1 & -I_q & 0 \\ a & b & -\dfrac{H_1}{q} \\ c & d & -\dfrac{L_2}{q} \end{pmatrix}}{\Delta} = \frac{(1 - C_1)(H_1 d - bL_2) + I_q(aL_2 - cH_1)}{\Delta} > 0,$$

if $H_1(L_1 + L_3) < L_2 H_2 \dfrac{\beta \pi_y}{E}$.

b. Dividend policy of the firms

To obtain the effect of an increase in β, we differentiate Eqs. (22.5)–(22.7) with β and get the following matrix:

$$\begin{pmatrix} 1 - C_1 & -I_q & -C_2 \\ a & b & 0 \\ c & d & -1 \end{pmatrix} \begin{pmatrix} \dfrac{\partial Y}{\partial \beta} \\ \dfrac{\partial q}{\partial \beta} \\ \dfrac{\partial M}{\partial \beta} \end{pmatrix} = \begin{pmatrix} 0 \\ -H_2 \dfrac{\pi(Y)}{E} \\ -L_3 \dfrac{\pi(Y)}{E} \end{pmatrix}.$$

Solving by Cramer's rule, we get

$$\frac{\partial Y}{\partial \beta}=\frac{\begin{pmatrix}0 & -I_q & -C_2\\ -H_2\dfrac{\pi(Y)}{E} & b & 0\\ -L_3\dfrac{\pi(Y)}{E} & d & -1\end{pmatrix}}{\Delta}=\frac{\pi(Y)\left(-I_qH_2+C_2H_2d-C_2L_3b\right)}{E\Delta}>0,$$

given the condition $I_qH_2 > C_2dH_2 + C_2bL_3$

$$\frac{\partial M}{\partial \beta}=\frac{\begin{pmatrix}1-C_1 & 0 & -C_2\\ a & -H_2\dfrac{\pi(Y)}{E} & 0\\ c & -L_3\dfrac{\pi(Y)}{E} & -1\end{pmatrix}}{\Delta}=\frac{\pi(Y)\big((1-C_1)H_2+aC_2L_3-H_2c\big)}{E\Delta}>0,$$

given the condition $(C_2cH_2 + C_2aL_3) > (1 - C_1)H_2$

$$\frac{\partial M}{\partial \beta}=\frac{\begin{pmatrix}1-C_1 & -I_q & 0\\ a & b & -H_2\dfrac{\pi(Y)}{E}\\ c & d & -L_3\dfrac{\pi(Y)}{E}\end{pmatrix}}{\Delta}$$
$$=\frac{\pi(Y)((dH_2-bL_3)(1-C_1)+I_q(-cH_2+aL_3)}{E\Delta}>0,$$

given the condition $I_q(aL_3 - cH_2) > (1 - C_1)\,(dH_2 - bL_3)$.

c. Expansionary fiscal policy

To study the effect of an expansionary fiscal policy, we differentiate Eqs. (22.5)–(22.7) by G and get the following matrix:

$$\begin{pmatrix}1-C_1 & -I_q & -C_2\\ a & b & 0\\ c & d & -1\end{pmatrix}\begin{pmatrix}\dfrac{\partial Y}{\partial G}\\ \dfrac{\partial q}{\partial G}\\ \dfrac{\partial M}{\partial G}\end{pmatrix}=\begin{pmatrix}1\\0\\0\end{pmatrix}.$$

Solving using Cramer's rule, we get

$$\frac{dY}{dG}=\frac{\begin{pmatrix}1 & -I_q & -C_2\\ 0 & b & 0\\ 0 & d & -1\end{pmatrix}}{\Delta}=\frac{H_1q^e-D}{q^2\Delta}>0$$

$$\frac{dq}{dG} = \frac{\begin{pmatrix} 1 - C_1 & 1 & -C_2 \\ a & 0 & 0 \\ c & 0 & -1 \end{pmatrix}}{\Delta} = \frac{H_2 \beta \pi_y}{E\Delta} < 0$$

$$\frac{dM}{dG} = \frac{\begin{pmatrix} 1 - C_1 & -I_q & 1 \\ a & b & 0 \\ c & d & 0 \end{pmatrix}}{\Delta} = \frac{ad - bc}{\Delta} > 0, \text{ if ad} > \text{bc},$$

where $$\Delta = \begin{pmatrix} 1 - C_1 & -I_q & -C_2 \\ a & b & 0 \\ c & d & -1 \end{pmatrix} > 0.$$

References

Bernanke B (1995) Money, credit, and banking literature. J Money Credit Bank 27(1):1–28

Calvo GA (1998) Capital flows and capital-market crises: the simple economics of sudden stops. J Appl Econ 1(1):35–54. Reprinted in Calvo GA (2005) Emerging capital markets in turmoil: bad luck or bad policy? MIT Press, Cambridge

Dey P (2007) Companies show high debt to equity ratio. The Financial Express. http://www.financialexpress.com/news/companies-show-high-debttoequity-ratio/211206/ (Accessed 10 Nov 2008)

Financial Stress, Downturns and Recoveries, World Economic and Financial Surveys, World Economic Outlook, October 2008, International Monetary Fund.

Flood R, Garber P (1984) Collapsing exchange rate regimes: some linear examples. J Int Econ 17:1–13

Ireland P (2005) The monetary transmission mechanism. FRB Working Paper No.—06–1. http://www.bos.frb.org/economic/wp/wp2006/wp0601.pdf (Accessed 24 Oct 2008)

Knoop AT (2008) Modern financial macroeconomics. Blackwell, Oxford

Krugman P (1979) Model of balance of payments crises. J Money Credit Bank 11:311–325

Krugman P (1999) Balance sheets, the transfer problem, and financial crises. Int Tax Public Finan 6:459–472

Krugman P (2003) Crises: the next generation. In: Helpman E, Sadka E (eds) Economic policy in the international economy. Cambridge University Press, Cambridge

Krugman P (2008) The international finance multiplier. http://www.princeton.edu/∼pkrugman/finmult.pdf (Accessed 28 Jan 2009)

Malhotra M (2008) Financial crises: a cascading effect on India. http://advances.mse.ac.in/making/Financial%20Crises%20madhuri.pdf (Accessed 28 Jan 2009)

Mohan R (2008) Global financial crisis and key risks: impact on India and Asia. http://rbidocs.rbi.org.in/rdocs/Speeches/PDFs/87784.pdf (Accessed 28 Jan 2009)

Obstfeld M (1994) The logic of currency crises. Cah Econ Monet 43:189–212

Obstfeld M (1996) Models of currency crises with self-fulfilling features. Eur Econ Rev 40: 1037–1047

Roubini N, Setser B (2006) Bailouts or Bail-ins? Viva Books, New Delhi

Chapter 23
Empirical Evidence on the Relationship Between Stock Market Development and Economic Growth: A Cross-Country Exploration in Asia

J. Felix Raj and Samrat Roy

1 Introduction

Considerable attention has been paid to the empirical and theoretical discussions between financial market development and economic growth in macro-literature. The focus is gradually shifted to the issue of stock market–economic growth nexus. Levine and Zarkos (1998) argued that various measures of stock market development have explained part of the variation of economic growth. The idea that financial markets may be related with economic growth is a dynamic concept. Gurley and Shaw (1955) were the first to study the relationship between financial markets and economic growth. They argued that the basic distinction between a developed and a developing country arises due to the differences in the strengths of the financial systems. The argument was that financial markets could extend a borrower's financial capacity and improve the efficiency of trade. With well-developed financial markets, investors can be provided with the necessary funds for their projects. They concluded that financial markets contribute to economic development through enhancing physical capital accumulation.

An extensive volume of literature has attempted to address the above nexus at both the theoretical and empirical level. The findings and views expressed in these works have been generally conflicting in nature. Some studies like King and Levin (1993a, 1993b), Levine and Zervos (1998), and Demirguc and Maksimovic (1996) have found positive causal effects of financial development on economic growth in line with the 'supply leading' hypothesis. These studies claim that countries with better-developed financial systems, particularly those with large efficient banks and large, well-organized and smoothly functioning stock markets, tend to grow much faster by providing access to much needed funds for financially constrained

J. F. Raj (✉) · S. Roy
St. Xavier's College (Autonomous), Kolkata 700016, West Bengal, India
e-mail: felixrajsj@gmail.com

S. Roy
e-mail: samratsxc@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_23,

economic enterprises. Kletzer and Pardhan (1987) and Beck (2002) also argue along similar lines but they also tried to establish that financial development is much more effective in promoting economic growth in more industrialized economies than in agricultural economies. Their view has been contradicted in some other studies, which argue that countries at their early stage of development benefit more from financial sector development than their older and mature counterparts (Fry 1995). However, most of these studies being cross-country regression-based studies, there were some inherent weaknesses in such analyses that drew considerable criticism from contemporary researchers. Levine and Renelt (1992) talk about omitted variable bias or misspecification; Evans (1995) and Pesaran and Smith (1995) highlight the effect of heterogeneity of slope coefficients across countries; Demetriades and Hussain (1996) and Harris (1997) explore problems of causality and endogeneity. Motivated by such criticism, Levin et al. (2000) examined empirically the same issue by incorporating adequate corrections for the effects of simultaneity bias and country-specific effects, effects of other determinants of growth and biases arising from model-specific errors like omitted variables. Their conclusions identified a causal relationship running from financial development indicators to economic growth even after controlling for such factors. The 'demand following' argument is also there in the research works over the past four or five decades. Robinson (1952) argued that financial development primarily follows growth in the real economy, as a result of increased demand for financial services. Lucas (1988) stated that the role of financial sector development in causing economic growth of a country has been 'badly overstressed'.

Against this aforesaid background, this chapter focuses on Asian economies. Changes in stock market performance play an important role in attaining strong economic fundamentals in Asian economies. The magnitude dependency of the performance of stock market on economic growth is valuable information to policymakers. The linkage between each country's measures of stock prices and economic activity is an important area to study in order to obtain meaningful findings. This motivates the study to explore the relationship between stock market development and economic growth.

The objective of this study is to examine the dynamic relationship between stock markets and economic growth in eight Asian countries from 1980 to 2010. The eight Asian countries selected in this study are China, India, Indonesia, Malaysia, Pakistan, the Philippines, Sri Lanka and Singapore. It is interesting to investigate how the stock market performance spurs economic growth in these countries in order to understand the interrelationship between stock market and real output in different Asian economies context. More specifically, we intend to examine the dynamic causal linkage between stock markets and economic growth in these Asian countries using time series econometric methods.

2 Review of Literature

From an economic point of view, issue on whether or not stock markets promote economic growth has been an interesting topic that has prompted tremendous empirical studies to be carried out on this subject.

Atje and Jovanovic (1993) examine a cross-country study of stock markets and economic growth for 39 countries. The most notable detection, they claim, is the existence of a noteworthy relationship between level and growth rate of stock market towards the economic growth over the period 1980–1988. However, the equal effect anticipated for the banking lending is absent. Levine and Zervos (1996) employ pooled cross-country, time series regression analysis to examine the data compiled from 41 countries over the period of 1976–1993. The outcome reflects the existence of a strong bond between stock market development and economic growth. Furthermore, the instrumental variables procedures reveal that economic growth in the long run is heavily inclined towards the predetermined component of stock market development.

In contrast, the arguments of Singh (1997) and Harris (1997) suggest otherwise. Singh (1997) focuses his research on developing countries and seeks to understand the role of stock markets towards long-run economic growth in the 1980s and 1990s. He concludes that in developing countries, long-run economic growth does not show dependency towards the stock market. One of the main reasons for this scenario is that the volatility and arbitraging process will deficient investment allocation. Also, the reaction between stock market and currency markets in the wake of negative economic shocks will alter the constancy macroeconomic variables. Consequently, long-run economic growth would be hindered.

A study by Harris (1997) on 49 countries from 1980 to 1991 found out that there is no significant relationship between stock market and economic growth. He utilizes a two-stage least-squares technique whereby the sample size is divided into two sectors: developed and less-developed countries. Empirical results suggest that stock market has some explanatory power in developed countries, whereas in developing countries, stock market and economic growth do not appear to be robustly correlated.

Singh and Weisse (1998) dispute that stock markets are, in fact, likely to destruct economic expansion due to their vulnerability to market breakdown. They assert that stock market developments are not likely to assist developing countries in attaining faster industrialization and long-term economic growth in the 1980s and 1990s due to several reasons. First, the high unpredictability of share prices in rising markets renders the prices inefficient as signals for resource allocation. Second, rather than allowing corporate managers to take a long-term view of investment, stock markets, in fact, habitually promote short-term profits. Third, the supremacy of stock markets may weaken the role of the banking system to facilitate the economy in developing countries, particularly those in the East and Southeast Asian countries.

Levine and Zervos (1998) seek to measure and evaluate how the stock market, banks and economic growth interact with one another using data of 47 countries from 1976 to 1993. They provide empirical evidence that stock market liquidity and banking development are both positively and robustly correlated with future economic growth. By employing Sims' causality test, Tuncer and Alovsat (2000) investigate the causal relationship between stock markets and economic growth in 20 countries from 1981 to 1994. The panel data analysis indicates that there is a bidirectional causation between stock market development and economic growth. Although a concrete conclusion cannot be reached on the country-based analysis, they do claim that there is a rather strong linkage among the variables under study in developing countries.

Choong et al. (2003) utilize the autoregressive distributed lag (ARDL) bounds test and the Granger causality test based on a vector error correction model (VECM) to study the finance-led growth hypothesis in Malaysia for the period of 1978–2000. They suggest that stock market development poses a noteworthy constructive long-run impact on economic growth. In the short run, on the other hand, the stock market is identified as a foremost segment in stipulating domestic growth. They also state that in circumstances where the market is more stable and liberalize as well as there is an improvement in the size and the regulations of the stock market, the evolution of financial sector in particular the stock market has the inclination towards stimulating and promoting economic growth.

Caporale et al. (2005) re-examine the dynamic interactions between investments, stock market development and economic growth in Chile, Korea, Malaysia and the Philippines from 1977:1 to 1998:4. They employ the Granger noncausality techniques recently developed by Toda and Yamamoto (1995) to test the hypothesis that stock markets can enhance economic growth through investment productivity. Their findings are supporting to Leigh's (1997) argument in which a well-developed stock market can promote economic growth in the long run.

The contribution of our chapter to the above existing literature lies in focusing on the link between stock market developments and the growth for Asian economies, rather than constructing a composite index that simultaneously reflect some of these indicators, as in the study by Agrawalla and Tuteja (2007). It was felt that such a composite index may not adequately capture the influence of each indicator of stock market development separately on economic growth. Secondly, this chapter focuses only on stock market development and its causal linkage with economic growth, rather than interactions between the growth of real GDP and broad indicators of overall financial development, viz., financial depth, bank credit, etc., as is evident in the study by Chakraborty (2008).

3 Data and Methodology

This study employed annual data spanning from 1980 to 2010. The following stock exchanges are the sources of the data on stock prices: Shanghai Stock Exchange (SSE), Bombay Stock Exchange (BSE), Kuala Lumpur Stock Exchange (KLSE), Stock Exchange of Singapore (SGX), Philippines Stock Exchange (PSE), Colombo Stock Exchange (CSE), Pakistan Stock Exchange (PSE) and Indonesia Stock Exchange (ISE).

The data set consists of two variables for each of the Asian countries, which are real gross domestic product (RGDP) and stock indices. All the variables in the data set are transformed into natural logarithms for the usual statistical reasons. For the stock indices, all the values are extracted from various issues of each of the eight Asian countries' stock exchange statistical report. Meanwhile, RGDP data are compiled from world development indicators published by World Bank by deflating the nominal GDP with consumer price index (CPI).

This chapter uses time series methodology to address this issue of causality. As a matter of fact, the unit root test followed by Granger causality test is conducted.

Unit root test There are several reasons why the concept of nonstationarity is important. A stationary series have a constant mean, constant variance and constant auto-covariance for each given lag. Many factors can make the series nonstationary. In particular, seasonal effects, trend, shocks and so on can cause a non-stationary series. Time series should be separated from all these effects to make a correct evaluation with correct models. Stationarity could be achieved by appropriate differencing, and this appropriate number of differencing is called order of integration. If some series are nonstationary, differences should be taken until series are stationary at some level. If it is a nonstationary series, $t\,Y$ must be differenced one time before it becomes stationary, then it is said to be integrated of order (1). This would be written as $t\,Y \sim I(1)$. So if $t\,Y \sim I(1)$, then $d\,D\,t\,Y \sim I(0)$. This latter piece of terminology states that applying the difference operator, one time, leads to $I(0)$ process, a process with no unit roots. One of the methods to test whether series is stationary or not is Dickey–Fuller (DF) (1979). DF test is very important in terms of measuring which degree stationary series have, but it does not consider an autocorrelation in disturbance term. If disturbance term contains autocorrelation, DF test is invalid. In this situation, by adding lagged terms of dependent variable to explanatory variable, generalized Dickey–Fuller (augmented Dickey–Fuller, ADF) is used. The ADF unit root test is used for this purpose.

Granger causality test Granger causality is used for testing the long-run relationship between stock market development and economic growth. The Granger procedure is selected because it consists of more powerful and simpler way of testing causal relationship (Granger 1986). Granger's operational causality definition depends on two hypotheses. First, next cannot be reason of past. Certain causality is possible only with past causes present time or future time. Cause is always to be come true before the result. In addition, this makes the time lagged between causes and results. Second, causality can be determined only in stochastic process. It is not possible to determine the causality between two deterministic processes. Granger test states that if past values of a variable Y significantly contribute to forecast the value of another variable X, then Y is said to Granger cause X and vice versa.

Table 23.1 Stationary tests and Jarque–Bera test results

Augmented Dicky–Fuller test				Philips–Perron test		Jarque–Bera test
Country	Variable	H_0:$I(0)$	H_0:$I(1)$[a]	H_0:$I(0)$	H_0:$I(1)$[a]	H_0:$\mu_3 = 0$ and $\mu_4 = 3$[b]
China	LRGDP	– 1.53	– 4.92	– 6.45	– 7.65	0.673(0.53)
	LSTOCK	– 2.30	– 5.61	– 0.447	– 8.18	6.456(0.456)
Pakistan	LRGDP	– 6.228	– 4.47	– 6.05	– 4.06	4.654(0.543)
	LSTOCK	– 2.79	– 4.031	– 2.006	– 2.63	3.543(0.321)
Sri Lanka	LRGDP	– 1.579	– 3.85	– 2.62	– 3.94	5.432(0.453)
	LSTOCK	– 0.491	– 4.99	– 1.35	– 4.98	5.543(0.321)
Malaysia	LRGDP	– 2.25	– 6.57	– 1.29	– 6.55	9.754(0.221)
	LSTOCK	– 1.29	– 4.66	– 2.27	– 4.56	11.542(0.322)
Indonesia	LRGDP	– 2.31	– 4.56	– 2.45	– 4.62	7.543(0.231)
	LSTOCK	– 1.65	– 4.69	– 1.03	– 4.48	9.549(0.438)
Philippines	LRGDP	– 2.55	– 4.56	– 2.63	– 4.50	11.543(0.327)
	LSTOCK	– 2.67	– 6.46	– 2.64	– 6.75	10.573(0.111)
Singapore	LRGDP	– 0.92	– 4.55	– 1.76	– 4.45	8.563(0.558)
	LSTOCK	– 1.309	– 5.44	– 1.02	– 6.46	6.457(0.245)
India	LRGDP	– 2.48	– 4.17	– 2.17	– 4.28	5.432(0.378)
	LSTOCK	– 2.65	– 4.65	– 2.40	– 4.70	8.563(0.345)

[a] Values are significant at 5 % level
[b] Figures in parentheses are p values

4 Empirical Results

A prerequisite for testing for cointegration in a set of variables is to test for stochastic trends (unit roots) in the autoregressive representation of each individual time series. They should be integrated of the same order to be cointegrated. In other words, the variables should be stationary after differencing each time series the same number of times. To do the task, we rely on ADF unit root test, which is one of the most commonly used univariate unit root tests proposed by Dickey and Fuller (1981). This test runs an ordinary least-square (OLS) regression such as the following:

$$\Delta Yt = \alpha + \alpha Yt + \Sigma \beta i \Delta Yt + eit,$$

where ΔYt is the first difference of the Yt, $\alpha 0$ is the intercept, α is the coefficient, t is the time or trend variable and p is the number of lagged terms chosen to ensure that εt is white noise. The optimal lag length of p is selected by using Akaike's information criteria (AIC) suggested by Akaike (1977). If the observed t-statistic is found to be negative and statistically significant, we can reject the null hypothesis of a unit root.

Table 23.1 presents the results of the ADF test. The report indicates that we cannot reject the null hypothesis of a unit root for the variables under study in level for all the eight Asian countries. However, the ADF tests show that all the variables are stationary at first difference, implying the generalization of $I(1)$ variables.

We next conducted a Granger causality test to determine the direction of causation between these two variables. Table 23.2 presents the results of the Granger causality test. The null hypothesis is always that of noncausation.

Table 23.2 Granger causality test results

Country	Null hypothesis	Observations	Number of lags	F-statistic[a]	Prob value	Concluding remarks
China	LRGDP dngc LSTOCK	26	2	6.6860	0.0040	Bidirectional
	LSTOCK dngc LRGDP	26	2	0.1646	0.0046	
Pakistan	LRGDP dngc LSTOCK	26	2	8.79002	0.10168	No causality
	LSTOCK dngc LRGDP	26	2	0.67130	0.52167	
Sri Lanka	LRGDP dngc LSTOCK	27	1	0.9790	0.10029	No causality
	LSTOCK dngc LRGDP	27	1	0.92801	0.34489	
Malaysia	LRGDP dngc LSTOCK	26	2	4.30636	0.00192	Bidirectional
	LSTOCK dngc LRGDP	26	2	5.21743	0.00637	
Indonesia	LRGDP dngc LSTOCK	26	2	3.1790	0.04645	Bidirectional
	LSTOCK dngc LRGDP	26	2	5.1711	0.0456	
Philippines	LRGDP dngc LSTOCK	27	1	7.75121	0.0103	Bidirectional
	LSTOCK dngc LRGDP	27	1	5.82691	0.0026	
Singapore	LRGDP dngc LSTOCK	26	2	5.97034	0.0088	Bidirectional
	LSTOCK dngc LRGDP	26	2	5.2824	0.0012	
India	LRGDP dngc LSTOCK	27	1	5.506	0.02734	Bidirectional
	LSTOCK dngc LRGDP	27	1	4.8109	0.0517	

dngc do not Granger cause

[a] Values are statistically significant at 5 % level

In the case of China and India, there is evidence of bidirectional (feedback) causality between stock markets and economic growth in the short run. In the economy of the Philippines, we find unidirectional Granger causality. In the case of Indonesia, Malaysia and Singapore, the pairwise Granger causality tests indicate that there is a feedback effect in the system, or a short run bidirectional causality between stock market and economic growth. No causality exists for Sri Lanka and Pakistan.

5 Conclusion

Our finding suggests that stock markets in most of the countries under study play a fundamental role in promoting economic growth. Therefore, the corresponding countries' authorities should take capital market measures to improve relation and simplicity with the aim to facilitate a sustainable stock market. More stringent regulations should also be implemented to shield investors, to enhance the corporate ascendancy practices and also to ensure a systematic and fair market in the stock market trading of securities. In addition, the empirical results also suggest that stock markets can stimulus economic growth and vice versa in China, Indonesia, Malaysia and Thailand. Since the pragmatic results show a mixture of findings of causal interactions between stock markets and economic growth, the policy implication, as a result, should not be generalized but it should be designed differently to best fit the economic environment in different countries.

It is to our utmost optimism that these findings might cast some imminent to investors in helping them to predict upcoming market movement in accordance to stock market activities. It also has the purpose of serving as a constructive assistance and reminder for government and private sectors to always scrutinize the effectiveness of each policy they implement. Causality test results suggest that stock market development leads to economic growth at least for the period under study for the consideration, which is in line with the 'supply leading' hypotheses. The funds raised by the corporate from the financial markets during the study period, thus, played an important role for the appreciable growth of the Indian economy. With the Indian stock market assuming more and more importance, this finding could have significant policy implications for the market regulators and economic planners in future.

References

Akaike H (1977) On entropy maximization principle. In: Krishniah PR (ed) Applications of statistics. North-Holland, Amsterdam, pp 27–41

Atje R, Jovanovic B (1993) Stock markets and development. Eur Econ Rev 37:632–640

Bencivenga VR, Bruce DS, Ross MS (1996) Equity markets, transactions costs, and capital accumulation: an illustration. World Bank Econ Rev 10(2):241–265

Bergstrom AR (1990) Continuous time econometric modeling. Oxford University Press, Oxford

Caporale GM, Howells P, Soliman AM (2005) Endogenous growth models and stock market development: evidence from four countries. Rev Dev Econ 9(2):166–176

Choong CK, Yusop Z, Law SH, Liew KS (2003) Financial development and economic growth in Malaysia: the stock market perspective. Econ, WPA, Macroeconomics Series, No. 0307010

Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431

Harris R (1997) Stock market and development: a re-assessment. Eur Econ Rev 41:139–146

International Monetary Fund. International Financial Statistics, various issues. IMF, Washington D.C.

Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254

Johansen S, Juselius K (1990) Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. Oxford B Econ Stat 52(2):169–210

Leigh L (1997) Stock market equilibrium and macroeconomic fundamentals. (IMF Working Paper WP/79/15.) IMF, Washington, D.C.

Levine R (1991) Stock market, growth and tax policy. J Finan 46:1445–1465

Levine R, Zervos S (1996) Stock market development and long run growth. World Bank Econ Rev 10(2):323–339

Levine R, Zervos S (1998) Stock market, banks and economic growth. Am Econ Rev 88:537–558

Mayer C (1988) New issues in corporate finance. Eur Econ Rev 32:1167–1188

Singh A (1997) Financial liberalisation, stock markets and economic development. Econ J 107: 771–782

Singh A, Weisse BA (1998) Emerging stock markets, portfolio capital flows and long-term economic growth: micro and macroeconomic perspectives. World Dev 26(4):607–622

Toda HY, Yamamoto T (1995) Statistical inference in vector autoregressive with possibly integrated processes. J Econometrics 66:225–250

Tuncer GC, Alovsat M (2000) Stock markets and economic growth: a causality test. Doğuş Univ J 1(2):124–131

Chapter 24
Financial Development in India: An Empirical Test of the McKinnon–Shaw Model

Mahendra Pal

1 Introduction

Over the past four decades, a body of literature has emphasized the role of financial intermediation in the process of economic growth and it has also been argued that the increased financialization of assets is instrumental in promoting economic development (Gurley and Shaw 1955; Goldsmith 1969; Patrick 1966; McKinnon 1973; Shaw 1973; Gupta 1984; Jung 1986).

Schumpeter (1934) regarded the banking system as being capable of promoting economic development. In recent years, inspired by the seminal work of McKinnon (1973) and Shaw (1973) and in the light of its policy implications for less-developed countries (LDCs), the literature on development has focused considerable attention on the relationship between financial intermediation and economic growth (Fry 1988).

The McKinnon–Shaw (1973) hypothesis received widespread acceptance throughout the world and several developing countries adopted interest rate liberalization and the McKinnon–Shaw thesis has become the new conventional guiding light. Until the early 1980s, commercial banks and other financial institutions in almost all these countries were nationalized, interest rates were fixed below market clearing level, directed credit programs with minimum targets for priority sectors like

This chapter is based on the research conducted under the UGC MRP-Financial Deepening and Economic Growth in India- a Co-Integration Analysis. Financial Support from UGC is highly acknowledged. The author is highly thankful for the helpful comments from discussants Prof. R. P Sinha, Micheal J. Hannan, and other participants in Pennsylvania Economic Conference, USA and also to Prof. V. N. Pandit, Prof. K. L. Krishna, Prof. J. P Sharma, Prof. K. V. Bhanumurthi (DSE) for a close discussion during the research period and Dr. Rajeev Singh and Mr. Sachin Kumar for technical assistance.

M. Pal (✉)
Department of Commerce, Delhi School of Economics, Delhi, India
e-mail: mahendra_dsc@yahoo.com

Department of African Studies, University of Delhi, Delhi, India

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_24, © Springer India 2014

agriculture and small industries were exclusively in vogue, and governments apportioned a substantial share of total credit at lower rates of interest (Hansan and Craig 1986). The resulting "financial repression" was regarded as one of the most important sources of inefficiency, low saving, and poor growth in developing countries (McKinnon–Shaw 1973). However, in a large number of Third World nations under the structural adjustment lending (SAL), financial deregulation was opposed (Corbo and de Melo 1985). McKinnon (1988) also pointed out that "all is not well in the liberal camp," and acknowledged that how best to achieve financial liberalization, remains seriously incomplete. Financial repression in South Asia is the result of a set of policies that often takes the form of various administrative controls, distorting the domestic financial market. Some of the widely used instruments of financial repression are; institutionally controlled nominal interest rates, combined with a high inflation rate, and a negative real interest rate, which in turn, results in an extremely low savings rate. Accelerating inflation encourages people to invest their savings in tangible assets such as land or gold, fuelling further inflation (Fry 1988).

In view of the continuing controversy about financial liberalization, this chapter studies the financial sector reforms in India, especially interest rate liberalization through the policy of high real rate of interest. We test whether higher real interest rates, a proxy for financial liberalization lead to increased financial intermediation, and whether the financial intermediation increases the level of GDP in the economy. We also try to test additional hypothesis regarding the neostructuralist crowding out concept. The chapter is structured as follows: Section 24.2 discusses financial liberalization in India, Sect. 24.3 discusses the financial repression model of McKinnon–Shaw with its theoretic and empirical application, Sect. 24.4 gives a snapshot review of empirical literature, Sect. 24.5 deals with the data and methodology, empirical results, Sect. 24.6 discusses the model specification and empirical estimation and Sect. 24.7 summarizes the chapter and policy implications.

2 Financial Liberalization in India

India has been passing through the process of financial liberalization, where the objective is to accelerate the pace of economic development in an environment of reasonable price stability. Before the institution of the new economic policy (1991), the Indian economy was a financially repressed economy. The first phase of financial sector reforms were guided by the recommendations of the Committee on the Financial System: "Issues of competitive efficiency and profitability are ownership neutral. It is how the institution functions or allowed to function that is more important" (GOI 1991).[1] In India, the policy environment adversely affected the operational and allocative efficiency of the financial system (Rangarajan 1998). This was reflected in high transaction costs and poor quality and spread of financial services, which in turn

[1] Report of the Committee on the Financial System (1991), (Narasimham Committee Report), Ministry of Finance, Government of India, New Delhi, 1991.

resulted in low profitability, deteriorating quality of assets, and erosion in the capital base. Consequently, removal of policy constraints on the functioning of the financial system through suitable modifications in the policy framework within which the financial system operates was assigned priority in the sequencing of financial sector reforms.

India has now undertaken an ambitious financial sector reforms program with a view to facilitating an efficient and effective monetary management through, first the introduction of indirect monetary controls to remove and replace the use of credit ceiling policy, second removing the distortions and segmentation of financial market for government debt instruments, and third, from administered interest rate to market interest rate determination by initiating a regular auction programme of the government debt.

The banking sector in India remained heavily regulated for a long time. Public sector banks accounted for about 93 % of total deposits. After reforms, the government removed the restrictions on entry and expansion of private banks, and reduced those on foreign banks. Now, the monopoly of public sector banks has been dismantled because of the entry of foreign and private banks and their share increased in the deposit and credit markets. Even now, the public sector banks continue to dominate the banking industry.

The reserve requirements, which ensure the liquidity of banks, also impose a significant cost on the banks since the interest received on the reserves is low, i.e., around 5 % per annum. This requirement ranges from 5 to 15 % in India. Since the reforms of 1991, this ratio has progressively come down to less than 5 %, however, with certain fluctuations recently. The statutory liquidity ratio (SLR) which reached to 38.5 % in 1991 was reduced to its lowest level, i.e., around 25 %.[2] The allocation of credit in India is heavily controlled by the state through the government, directed bank lending, i.e., priority sector lending. In India, it accounts for about 40 % of total bank credit since the early 1970s; however, it has come down to around 30 %.[3]

India has partially liberalized the interest rates structure and continues to regulate interest rates on priority sector loans. The interest rates in India were mostly controlled by the government prior to the reforms. India retained a ceiling on the maximum deposit rate that can be paid on fixed deposits with banks. Thus, it is clear that in the prereform period, interest rates were significantly below market clearing level in India. However, following the reforms nominal interest rates were raised significantly. The reforms were followed by significant foreign capital inflows, which led to excess liquidity and thus, slightly lowered interest rates. The administered interest rate structure was dismantled and thus financial system was given freedom

---

[2] The liquidity requirement stipulates that banks must invest a certain fraction of their total deposits in government or government approved securities, which typically carry below market interest rates. It is often used to divert the bank deposits of the households to finance the government budget deficits and other credit needs of the public sector at subsidized interest rates.

[3] In India, the extent of directed credit has not been reduced but interest rates subsidies on priority sector lending have been largely phased out following the financial reforms in 1992 and there has been renewed emphasis on ensuring the recovery of loans. Thus, the state continues to play a significant, though declining role in the allocation of bank credit.

in the determination of the interest rate level. Money market rate, TR Bill Rate, real rate of interest, nominal deposit rates, State Bank of India (SBI) advance rate, commercial paper (C.P), and Certificate of Deposit (C.D) rates have been liberalized completely.

Money market rates before 1970 were very low, around 4.5 %. During the 1970s, there was an upward trend with high fluctuations; however, during the 1980s, the moved upwards. During the 1960s, the average real interest rates remained negative (i.e., − 1.20). Even during the early 1970s, they were negative but have been moderately positive since then. During the 1980s, especially 1984 onwards, real rates remained positive. Further in 1991, interest rate channel has assumed a key role as an instrument of macroeconomic policy. Since the 1970s, driven by McKinnon and Shaw hypothesis, policy authorities have rolled back repression regimes and have undertaken concerted reform measures in their endeavor to strengthen competition and improve the functioning of financial markets. Deregulation of interest rates is the most common element of financial reforms occurring around the world. Shifts in operating procedures of monetary control have accompanied these changes in the policy environment, paving the way for a shift from direct instruments to indirect instruments of monetary control. Short-term interest rates have become the key instruments through which central banks transmit policy impulses to the financial market.[4] In India, financial deepening increased from 26 % in 1960 to 44 % in 1985 and a spectacular increase was observed during the liberalization period (i.e., from 45 % in 1990 to around 80 % in 2010).

3 Financial Repression Model of McKinnon and Shaw

McKinnon and Shaw (1973) provide a forceful critique of the conventional wisdom of classical theory of Wicksell and Ricardo and later with Keynes, which states that below-market interest rates stimulate private investment. According to the McKinnon–Shaw thesis, below-market interest rates discourage financial savings and at the same time permit investments to take place where marginal returns to invested capital are poor, leading to an overall low efficiency of investment. Consequently, against the traditional belief, below-market interest rates are responsible for economic stagnation, rather than growth.[5]

[4] Loans above Rs. 2 lac have only a minimum lending rate prescribed. Whereas loans below that size have been broadly divided into two categories, each with a rate assigned to it. Money market rates such as call money rate, certificate of deposits (CDs), bill discounting rates, rates on commercial papers (CPs) have been completely freed. Debenture rate has also been freed.

[5] Financial Repression in McKinnon's phrase consists in setting a price for external finance (the interest rate) below the equilibrium level and discriminating in favor of the Government, firms engaged in foreign trade, and perhaps a few others, at the expense of other would be borrowers who are either excluded from the market altogether and made to depend on internal finance, limited to high priced credit from money lenders or unable to find any capital with which to grow.

3.1 Theoretical Foundation: McKinnon

The complementarity hypothesis between money and physical capital in the McKinnon (1973) framework can be seen from his construction of the demand for money and investment function:

$$(M/P) = f(Y/P, I/Y, d - i) \tag{24.1}$$

where M is the money stock, P is the price level, M/P is the real money stock, Y is the gross national product (GNP), Y/P is the real GNP, I/Y is the ratio of gross investment to GNP, d is the deposit rate, I is the inflation rate, and $d - i$ is the real deposit rate of interest (RR). The investment/income ratio is included as one of the determinants of the real stock of money. McKinnon (1973) argues that such a construction incorporates the demand for money that arises directly from the process for capital accumulation itself. The first independent variable still captures the conventional transaction motive for holding money, while the last variable measures the real return on holding money. If the deposit rate is kept constant for a long time while inflation fluctuates, the last variables $(d - i)$ effectively show the impact of inflation on money demand. The necessary condition for money to complement capital is that the partial derivative of the second explanatory variable is positive, such as:

$$\delta(M/P)/\delta(I/Y) > 0. \tag{24.2}$$

McKinnon (1973) draws attention to the indirect link of $(d - i) \rightarrow M/P \rightarrow I/Y$ if money is treated as a CONDUIT through which accumulation takes place. The demand for money increases with the productivity of capital. The desire to hold money depends on the real return; a positive real return raises investment and savings propensities because of the importance of money as a store of value. If the attractiveness of holding money is high, the partial derivative of the real return variable $(d-i)$ is large and positive, such as:

$$\delta(M/P)/\delta(d - i) > 0. \tag{24.3}$$

The point to be noted is that a small derivative indicates the presence of financial repression. If domestic saving is equal to domestic investment, a saving function with real income and an additional independent variable of the real rate of interest $(d-i)$ can be constructed as follows:

$$(S/Y) = f(Y/P, d - i) \tag{24.4}$$

where S/Y shows gross national saving and Y/P is the real income. The coefficient estimates can be used as an indication of complementarity. In the McKinnon (1973) construction, complementarity can also be seen in investment function in the following form:

$$(I/Y) = f(R, L - i) \tag{24.5}$$

where R is the average return on physical capital and L is the loan rate. If we start from a situation of financial repression with low real cash balances, the complementarity effect as indicated by the partial derivative is as follows:

$$\delta(I/Y)/\delta(L-i) > 0. \tag{24.6}$$

If cash balances are already attractive to hold and the economy is already liquid with cash, further increases in the real deposit rate reduce the propensity to save so that the direction of Eq. (24.6) is reversed and money and capital revert to the neoclassical case of a "competing asset" affect.

McKinnon's indirect link can be reformulated by substituting (M/P) for $(L-i)$ as the explanatory variable in Eq. (24.5). The monetary variable will then have a direct impact on the investment/income ratio. Such a direct formulation becomes.

$$(I/Y) = f(R, M/P). \tag{24.7}$$

By including (M/P) in the investment, the two Eqs. (24.4) and (24.7), become the two equations system employed in the Thornton and Poudyal's (1990) study on the Nepal's Economy.

3.2 Theoretical Foundation: Shaw

Shaw's formal theoretical contribution is not based on the stringent assumption of a fiat money world, but on the role of financial intermediaries in the process of economic development. If domestic saving (S) is assumed to equal domestically financed investment, McKinnon's demand for money function can be modified by replacing (I/Y) with (S/Y). Shaw (1973) proposes the debt intermediation view (DIV) which looks at the income effect of money and real capital generated from a reduction in transaction and information costs. His demand for money functions is as follows:

$$(M/P) = f(Y/P, \mathrm{V}, d-i) \tag{24.8}$$

where V is the vector of the opportunity cost of holding money in real terms. Shaw (1973) expects real yield on all forms of wealth including money to have a positive effect on the domestic savings ratio. On the contrary, financial intermediation is repressed when interest rates are fixed administratively below equilibrium levels.[6]

4 Review of Empirical Literature

Fry (1978) estimated the money demand function utilizing annual data obtained from a sample of 10 Asian LDCs for the time period 1962–1972, considering simultaniety between money demand and saving. He formulated the simultaneous equation system

[6] For detailed discussion see (Fry 1988; Gelb 1989; Li 1991; Ahmad and Ansari 1995).

to estimate the demand for money. Fry tried to find out a negative relationship between the domestic the savings ratio and real money balances in the demand for money function, which is inconsistent with McKinnon's complementarity hypothesis. He however found a positive effect of the real interest rate on domestic saving and economic growth. Fry explained this result by concluding that countries included in his sample have achieved a level of financial development beyond that of which the conditions for complementarity might hold. He stated that if one desires to find evidence supporting McKinnon's complementarity hypothesis, then one would need to look to the world's least developed economies, those which have achieved a level of economic and financial development well below those included in his study.

Fry and Mason (1982), using a life cycle saving model, estimated a saving function using pooled time series data for seven Asian economies. They found that interest rate coefficient is positive and statistically significant in the saving function. More recently, Fry (1988) estimated the International Monetary Fund (IMF)'s cross-section study for 1971–1980 with 22 countries and found the positive relationship between real output and real deposit rate.

In a more comprehensive background study for the World Bank, Alan Gelb (1989) analyzed the relationship between real deposit rate of interest (RR) and output growth (Y) and found strong correlation over the period of 1965–1985. With the breakdown of the Bretton Woods system of fixed exchange rates after 1973, measured average growth in real GDP fell from 6 % per year to 4 % per year in the sample of 33 countries. Hence, Gelb introduced a dummy variable SHIFT, which has the value 0 for 1965–1973 and 1 for 1974–1985 and then calculated countrywise averages for RR and Y for each of the two subperiods. He then reran the regression pooled over the two subperiods. The SHIFT variable indicates a marked decline in output growth after 1973. He found that for every 1 % increase in the real deposit rate, output growth increases by 0.2–0.25 % respectively.[7]

Following Fry's methodology, Thornton and Poudyal (1990) tested the complementarity utilizing data from Nepal over the period 1974–1987. According to the World Development Report, which classifies countries according to GNP per capita, Nepal was the eleventh poorest country in the world as of 1989. Thornton and Poudyal contend that we should expect complementarity to be a feature of the demand for money in an economy which occupies such a low position on the world's development ladder. The authors employed the same estimation procedures as Fry, but slightly modified Fry's demand for money and saving function specifications. Their findings provide support of McKinnon's assertion that the conditions necessary for complementarity are characteristics of the world's least developed economies.

Seck and Nil (1993), in their study of nine developing countries of Africa, found that the real deposit rate had a positive effect on financial savings (measured by the growth of M_2/GDP), on ratio of gross investment to GDP and on the growth rate of output. Haque et al. (1993) empirically examined the determinants of income velocity of money in Bangladesh. Their empirical results indicate that inflation and income variables affect velocity positively. The proxy for financial development

[7] World Development Report 1989; Gelb (1989).

(demand deposit over time deposits) affects velocity negatively, implying that the lower the proxy, the greater the level of financial development and the higher the velocity of money. Hassan (1995) tested for complementarity between money and capital in Bangladesh by using a two-stage least square technique over the time period 1970–1995. The estimated results provide support for McKinnon's hypothesis, suggesting that Bangladesh's financial system has not yet achieved a level of development whereby alternative nonmonetary assets have replaced money as the primary repository for domestic savings.

In a more recent study, Watson (1992) obtained results confirming the McKinnon–Shaw financial liberalization hypothesis but could not fully substantiate McKinnon's hypothesis for Trinidad and Tobago. Recently, Khan and Hassan (1998) using unit roots, co-integration, and error correction methodology found strong support for McKinnon's hypothesis in Pakistan. The coefficients of saving ratio in the money demand function and of real money balances in the saving function are positive and statistically significant.

4.1 Indian Case Studies

Some studies have been published that deal exclusively with Indian economy. Thornton (1989) found strong support for the complementarity hypothesis in the case of India. Demetriades and Luintel (1994) calculated the direct cost of financial repression policies in India using an index of repression policies. Their findings show that a removal of ceilings on deposit rates would result in an increase in financial depth by around 11 %. Furthermore, if required reserve requirements are reduced from their 1991 level of 15 % to, say, 5 %, it would result in a long run increase in financial depth of 5.4 %. Recently, Bill and Rousseau (2000) evaluated the strength and direction of the links between means of formal intermediation and various economic aggregates in India. Pradeep Aggarwal also tested the relationship between financial deepening (M_2/GDP) and deposit rate (DR) and inflation rate (P) in five Asian sample countries (i.e., India, Pakistan, Sri Lanka, Nepal, and Bangladesh) for the period of 1970–1993. His empirical findings show the positive relationship between financial deepening and DRs. His findings reveal that a 1 % rise in DR causes a 2.25 % rise in financial deepening. He also finds negative relationship between inflation rate and financial deepening. His results show that 1 % rise in inflation rate reduces financial deepening by 0.74 %. His results are significant at the 10 % level. Results are satisfactory, though, not with much explanatory power; but, they confirm the nature of McKinnon–Shaw hypothesis. Increase in a bank's nominal rate may result in an improvement in the financial deepening, which leads to economic growth in these economies. In these countries, interest rates have been partially liberalized. Given the controversy between repressionist school (McKinnon–Shaw) and neostructuralists and the oligopolistic nature of banking sector in these countries, especially in India, some regulation on interest rates will have to be continued.

5 Data and Methodology

To achieve this objective, we shall use McKinnon's complementarity hypothesis, because of its succinctness and easy amenability to empirical verification. McKinnon's hypothesis works as follows: An increase in real-interest rates will increase the demand for real money balances and since investment finance is an important motive in the demand for money, the increase in the demand for real balances will increase investment. The formulation will be made operational by estimating the real money demand and saving functions simultaneously. McKinnon's hypothesis will be tested using the time-series data for the period of 1971–2010. The main reason of selection of this period is because many countries like India started her liberalization process of their economies and financial system during this period. Second, India started her financial sector reforms; however, it was started in 1985, but got momentum after new economic policy of 1991. During this period i.e., 1971–2010, financial deepening i.e. M_3/GDP, private corporate saving, GDP growth rate, Pcy, quasi money/GDP ratio, real money balances (M/P), openness (X + M/GDP), have shown a rapid rising trend. For estimation purpose, we use the two-stage least square (2SLS) and ordinary least square (OLS) regression techniques.

Data used in this study are published, unpublished and self- generated data. Published data are available from various publications of the Reserve Bank of India (RBI), World Development Indicators (World Bank), International Financial Statistic (IFS) CD ROM (IMF); Handbook of Statistics on the Indian Economy, RBI bulletins and Report on Currency and Finance, Economic Survey, and various issues the Government of India.

6 Empirical Analysis

6.1 McKinnon's Complementarity Hypothesis

Following the methodology of Fry (1978) and Thornton (1989), we shall use McKinnon's complementarity hypothesis, because of its succinctness and easy amenability to empirical verification. McKinnon's hypothesis works as follows: An increase in real interest rates will increase the demand for real money balances and since investment finance is an important motive in the demand for money, the increase in the demand for real balances will increase investment. The formulation will be made operational by estimating the real money demand and saving functions, simultaneously.

$$\mathrm{In}\,(M/P) = f\,(\overset{+}{Sd/Y}), \overset{+}{\mathrm{In}\,(Y/P)}, \overset{-}{P}, \overset{+}{\mathrm{In}}\ (M/P) - 1 \tag{24.9}$$

Table 24.1 Money demand function: (In (M/P)) as dependent variable

C	Sd/Y	In (Y/P)	P	In (M/P)_1	R^{-2}	D.W.
4.212 (5.80)	0.865** (2.13)	0.841* (5.12)	– 0.049** (2.68)	0.904* (12.46)	0.99	2.05

t values are in parenthesis below the coefficient
*Indicates significant at 1 % level
**Indicates significant at 5 % level

$$S/Y = f(\overset{+}{\text{In}\,(M/P)}, \overset{+}{\text{In}\,(Y/P)}, \overset{-}{Sf/Y}, \overset{+}{(Sd/Y) - 1} \tag{24.10}$$

[+ sign stands for positive relationship, i.e., > 0, and − sign stands for negative relationship, i.e., < 0]

where In (M/P) is the per capita real money balances expressed in natural log (M/P) is real money stock which, in monetary equilibrium, must equal real money demand (M/P), i.e. $(M/P) = (M^d/P)$. (M/P) is broadly defined, which includes M^1 + quasi money (time deposit and saving deposit).[8] Y/P represents real per capita income expressed in natural logarithm, Sf/Y is foreign saving. The Harrod–Domar model states that the rate of growth of output is equal to the savings rate divided by the incremental capital output ratio, i.e. $g = s/v$ where g is the growths rate, s is saving rate while v stands for capital output ratio. Chenery and Strout (1966) argued that foreign saving acted as supplement to domestic savings and hence raised the growth rate to $(s + a/v)$, where a is foreign aid. Inflow of foreign aid would have the effect of raising the saving rate in subsequent period. In the money demand function, domestic savings ratio (Sd/Y), real per capita income (Y/P) and lagged money balances are expected to be positively related to money demand, while inflation (P) is expected to be negatively related to money demand.

In Table 24.1 results show strong support for McKinnon's complementarity hypothesis in both the demand for money and saving function. The coefficient Sd/Y in the money demand function and the coefficient of (M/P) in the saving function are both positive and significant. This is consistent with higher average money balances being held for domestically financed investment. In the demand for money function, the expected rate of inflation has a negative and significant impact on the demand for real money balances, by implications, therefore, expected real rates of return have a positive impact. In Table 24.2 if we see the saving function, the coefficient of the lagged dependent variable, the short run demand for real money balances appears to adjust only slowly to desired real money balances.

---

[8] M^2 is consisted of the sum of line 34 and line 35. Line 34 stands for money and line 35 stands for quasi money. These two lines are obtained from the International Financial Statistics (IFS) a publication of the International Monetary Fund (IMF).

Table 24.2 Saving function: (Sd/Y) as dependent variable

C	In(M/P)	In (Y/P)	Sf/Y	Sd/Y_1	R^{-2}	D.W.
0.130	0.0639**	– 0.0308*	– 0.450**	0.071*	0.89	1.80
(1.96)	(2.173)	(0.983)	(1.93)	(0.663)		

t values are in parenthesis below the coefficient
*Indicates significant at 1 % level
**Indicates significant at 5 % level

Table 24.3 Financial depth, saving and growth in developing countries (1965–1987). (Source: World Bank (World Development Report 1989))

Country group by GDP growth rate	M^2/GDP	S/GDP	ICOR
High growth (over 7 %) 7 countries	43.0[a]	28.0	3.80
Medium growth (3–7 %) 51 countries	31.2	18.5	4.24
Low growth (less than 3 %) 22 countries	23.8	19.0	9.9

Data are weighted average times 100 and are based on a sample of 80 developing countries. M^2 is currency in circulation plus demand, time and savings deposits at banks. Because of lack of data, average is for 1977–1987 only

6.2 Shaw Intermediation Hypothesis

6.2.1 Financial Deepening and Economic Development

We also test the financial deepening hypothesis postulated by Shaw (1966) which traces the relationship between financial deepening and economic growth.[9] Table 24.3 shows that high economic growth in developing countries is associated with greater financial depth, high savings rates and lower incremental capital-output ratio (ICOR). Moreover, the indicators of financial deepening are reflected through different ratios such as M_1/GDP, M_3/GDP, and TD/M_3 in India. We do not report all the trends here. Data is available with the author and can be made available on request.

Thus, the equation estimated for the period of 1971–2010 is as follows:

$$\mathrm{In(GDP)} = \mathrm{C} + 1.99\ \ln(\mathrm{M}_3/\mathrm{GDP})\mathrm{R}^{-2} = 0.84 \qquad (24.11)$$

(3.60).

Our data fitted to this equation yields that the coefficient of M^3/GDP is found to have the positive sign predicted by McKinnon–Shaw hypothesis. Results show that a 1 % rise in financial deepening (M_3/GDP) causes GDP growth rate to rise by 1.99 %. The b coefficient is significant at the 1 % level.

[9] Financial depth, as measured by, for example, the ratio M3/GDP, is an indicator of how well developed a country's financial system is when the ratio is relatively low, the flow of loanable funds from lenders is restricted. Encouragement of greater financial depth itself depends, to some extent, on the willingness of wealth holders to place saving with financial intermediaries (e.g. deposit banks, mutual funds, pension funds, life insurance companies) or to hold bonds or equity. Crucial here is the real rate of interest on, say, deposits. When this is positive, financial deepening is encouraged, but if it is negative, wealth holders will seek other less liquid means in which to hold their wealth.

6.3 Financial Crowding Out Hypothesis

So far, we have examined in detail the implications of financial reforms only with respect to the formal money market (FMM). Here, we consider the neo-structuarlists crowding out argument against financial liberalization (Wijnbergen 1983; Bufee 1984; Taylor 1988). They argue that raising bank DR simply result in transfer of funds from informal money market (IMM) which may be more efficient in providing credit to investors (Edwards 1988; van Wijnbergen 1982). Some funds may also be transferred from physical capital and share markets (Morriest 1993). However, it is to be noted that under the impact of financial dualism, share markets in India and other South Asian countries still account for a very small fraction of total financial assets. Moreover, IMMs are not well integrated with the FMMs because IMMs mostly utilize unaccounted money. Moreover, the interest rate in IMMs are very much high, and thus, minor changes in DRs may not really lead to significant transfer from IMMs to FMMs. Because of lack of time-series data from Asian countries, the neo-structuralist argument can be tested indirectly only. If we estimate the GFCF/GDP as a function of bank (DR), the coefficient of the DR should be positive if the McKinnon–Shaw argument dominates. On the other hand, if the neo-structuralist argument is valid, the coefficient of bank DRs should be negative. We have tested the following relationship in India for the period of 1971–2010 and found the positive results:

$$\text{In (GFCF/GDP=C+1.99 In (DR), R}^{-2} = 0.836 \qquad (24.12)$$

(2.61).

The *b* coefficient is significant at 10 % level.

7 Policy Implications and Conclusions

First, care must be taken in deciding which variables are exogenous and which are endogenous. The positive correlation between growth in financial assets and growth in GDP does not show which way the causality operates. However, for the purpose of portfolio choice by an individual investor, a case can be made for treating the real rate of interest on depository claims on banks as exogenous. Governments usually intervene to set ceilings on nominal rates of interest on bank deposits, and at the same time they determine the aggregate rate of price inflation, the real DR of interest is therefore, more or less determined by public policy. Nevertheless, this assumption that the real DR is an exogenous indicator of financial policy remains to be considered in the context of a structural model of how process work them out.

When the interest rate is completely flexible, a macro economic management problem may arise, because bank and non-bank financial intermediaries are permitted to mushroom and quantitative controls over credit are ruled out. Macroeconomic management would generally be beyond the monetary authority of any developing country. It is essential that in case of India. The RBI needs to have some blunt instruments e.g. credit ceilings or quantitative restrictions on credit to traders (Rakshit 1994). Deregulation of interest rate should be based on gradual process. A completely deregulated interest rate is a phenomenon of the 1980s.

We have already argued that low price-distortion economies have performed better then high price-distortion economies but it does not follow that the immediate removal of all price distortions is desirable. For this there should be proper using of reform process. (McKinnon 1982; Edward 1984; Corbo and de Melo 1987; Fry 1988; World Bank 1989; Collier and Gunning 1992). The economic collapse of the Southern Cone Countries of Latin America in the early 1980s, following incorrect order and hasty financial liberalizations, demonstrates the necessity of sequencing the correct order of financial liberalization.

Second, importance of pace is to be recognized. In the light of the poor results of Latin American financial liberalization in the 1970s and early 1980s, many economists argue that financial liberalization in developing countries is most successful when it is gradual (McKinnon 1989; Villanueva and Mirakhor 1990). According to Viollaueva and Mirakhor (1990), there are two groups of countries: the "rapid liberalizers" (Argentina, Chile, Malaysia, Philippines, Turkey, and Uruguay) and the "gradual liberalizers" (Indonesia, South Korea, Sri Lanka, Taiwan, and Singapore). One reason for the relative success of the second group was found to be that they first provided a relatively stable macroeconomic environment—i.e., manageable budget deficit and slower rates of money growth—before reducing the degree of interest rate management and credit rationing. With the interest rate stabilized, widespread bankruptcies were avoided and threat of investment was maintained. The main problem for the rapid liberalizers was that they virtually abandoned management of interest rates and allocation of credit before the fiscal position was made sustainable. When they were free to do so, the bank sharply raised interest rates and this led to wide spread bankruptcies and the distortion of investment into more risky ventures. An order of liberalization is now accepted by some economists, i.e., reduced fiscal deficit is considered a prerequisite for interest rate liberalization in the literature. Without reducing fiscal deficit, financial liberalization would be seen by the private parties as unsustainable and reversible as we have already discussed the cases of Argentina and Uruguay. The intertemporal distortion would then be very likely to affect adversely the amount of intermediation that the domestic bank would perform. This is because the increase in the real DR of interest will be considered as temporary, hence, it might attract investors or wealth holders to diversify into bank deposits, but they might invest in inflation hedges or foreign currencies.

Third, in the context of interest rate structure Chakravorty committee's views are discussed here. The Committee discussed mainly three heads, i.e., a link between interest rate (i) and expected rate of inflation (P^e); interest rate on government securities and interest rates on deposits and loan of banks. The link between administered interest rates with appropriate P^e clearly draws upon Irving Fisher's well-known hypothesis about the relation between money (nominal) and real rates of interest which says that, in a risk free world, in equilibrium, we have $i = r + P^e$ where i stands for money rate of interest and $r =$ real rate of interest.

In this way i is put into actual practice. The Committee recommended also the upward revision of all the rates on government securities so that they become competitive in the open market. Regarding the deposit and lending rates, the Committee recommended that the maximum DR applicable to deposits with a maturity of 5 years and above should be fixed at long term $P^e +$ a positive real r of not less than 2 % per annum. The 1-year DR should be marginally positive in real terms.

The basic lending rate should be fixed at 3 percentage points higher than maximum rate on deposits. Since the maximum DR recommended long-term $P^e + 2\,\%$ p.a., the basic lending rate will come to long-term $P + 5\,\%$ p.a. The 3 % will constitute the minimum administered spread and one rate below it applicable to all concessional lending. In other words, the present proliferous interest rate structure could be replaced simply by a maximum DR, a minimum lending rate one below it, leaving the rest to be decided by the banks themselves.

Interest rate structure in India should be based on the Chakravorty Committee recommendations. With excellent performance of Indian financial sector reform, the reform process would be able to capture the emerging changes and will continue to play an important role in the economic growth as the Indian economy was being rapidly integrated into the global economy in the last 20 years.

7.1 Conclusions

In this study, we have tested empirically the McKinnon–Shaw complementarity hypothesis in India. India administratively regulated deposits and lending rates, high reserve requirements, heavy priority sector lending, high inflation rate—all these factors led financial repression in India till the introduction of new economic policy of 1991. The Govt. of India has recently introduced some limited financial liberalization measures such as liberalization of interest rates structure. The success of the interest rates deregulation policies, however, depends on the sensitivity of saving and the demand for real money balances to interest rate changes. With those objectives, we have tested a simultaneous money demand and saving function. In the money demand function, saving income ratio, and in saving function, real money balances have positive and statistically significant effect, leading support to the complementarity hypothesis. We have also tested Shaw's financial deepening hypothesis and Wijnbergen and Taylor's crowding out hypothesis. All results presented in this study are consistent with McKinnon and Shaw' view that financial liberalization increases savings, improves the efficiency with which resources are allocated among alternative investment projects and therefore raises the rate of economic growth. These results support the continued financial liberalization with effective macroeconomic management policies in India.

Acknowledgments This chapter is based on the research carried out under the UGC MRP-Financial Deepening and Economic Growth in India- a Co-Integration Analysis. Financial Support from UGC is highly acknowledged. The author is very thankful for the helpful comments from discussants

Prof. R. P Sinha, Micheal J. Hannan and other participants in Pennsylvania Economic Conference, USA and also to Prof. V. N.Pandit, Prof. K. L. Krishna, Prof. J. P Sharma, Prof. K. V. Bhanu Murthy (DSE) for a close discussion during the research period and Dr. Rajeev Singh and Mr. Sachin Kumar for technical assistance.

References

Akhtar MA (1974) The demand for money in Pakistan. Pakistan Dev Rev 13:40–54

Berthelemy JC, Varoudakis A (1996) Economic growth, convergence clubs and the role of financial development. Oxford Econ Pap 48:300–328

Chakravorty S (1985) Report of the committee to review the working of the monetary system. Bombay, Reserve Bank of India

Clarke R (1996) Equilibrium interest rates and financial liberalization in developing countries. J Dev Stud 32:391–413

Edward S (1984) The Order of Liberalization of the External Sector in Developing Countries, Essay in International Finance, No. 156, Princeton. See also Vijai J (1994) Macro Economic Policy Reform in India, Exim, Bombay

Fisher B (1981) Interest rate ceilings, inflation, economic growth in developing countries. Economics 75–93

Fry MJ (1978) Money and Capital or financial deepening in economic development. J Money Credit Bank 10:464–475

Fry MJ (1995) Money, interest and banking in economic development. John Hopkins University Press, Baltimore

Fry MJ, Mason A (1982) Variable rate of growth effect in the life cycle saving model: children, capital inflows, interest and growth in a new specification of the life cycle model applied to seven Asian developing countries. Econ Inq 20:426–442

Galbis V (1977) Financial intermediation and economic growth in less developed countries: a theoretical approach. J Dev Stud 13:2 (January)

Gelb A (1989) Financial Polices, Growth and Efficiency. Policy Planning and Research Working Paper, World Bank, 2002

Gelb A, Honohan P (1989) Financial sector reforms in adjustment programs. Policy, Planning and Research Working Paper 169, Washington, D.C.: World Bank, Country Economics Department, Processed

Ghatak S (1981) Monetary economics in developing countries. Macmillan, London

Giovannini A (1983) The interest elasticity of savings in developing countries: the existing experience. World Dev 11:601–607

Goldsmith R (1969) Financial structure and development. Yale University Press, New Haven, pp 35–70

Gupta KL (1987) Aggregate savings, financial intermediation and interest rate. Rev Econ Statist 69:303–311

Gurley JG, Shaw ES (1955) Financial aspects of economic development. Amer Econ Rev 45:515–533

Gurley JG, Shaw ES (1960) Money in a theory of finance. Brookings Institution, Washington, D.C.

Harris JN (1979) An empirical note on the investment content of real output and the demand for money in the developing economy. Malays Econ Rev 23:49–62

Hossain A (1993) Financial reforms, stability of the money demand function and monetary policy in Bangladesh: an econometric investigation. Indian Econ Rev 28:85–100

Jadhav N (1994) Monetary economics for India. Macmillan, New Delhi

Jung WS (1986) Financial development and economic growth: international evidence. Econ Dev Cult Chang 34:333–346

Li KW, Skully M (1991) Financial deepening and institutional development. 15th June, 1987:147–164
Levine R (1997) Financial development and economic growth: views and agenda. J Econ Lit 35:688–626
McKinnon RL (1973) Money and capital in economic development. The Brookings Institution, Washington, D.C.
McKinnon RL (1986) Financial liberalization in retrospect: interest rate policies in LDCs. Stanford: Centre for economic policy research publication, 74
Klein LR (1987) Financial innovation, urgent need for regulation. Fourth C.D. Memorial Lecture, Bombay, Reserve Bank of India
Pal M (1996) Financial globalization: some theoretical and empirical issues. Gwalior: Indian Economic Association, Annual Conference Volume, 79, pp 116–124
Pal M (1997) Financial deepening, real interest rate & growth in India (Theory & Evidence). In: Jani BM (ed) Impact of financial liberaliztion policy in India. Himalaya, Bombay
Pal M (2012) Financial sector reforms and economic growth in Zambia. Indian J Afr Stud D.U. XVIII(1 & 2):151–162
Pandit VN (2001) Structural Modeling Under Challenge, Working Paper No. 98, CDE, DSE
Pandit VN (2002) Sustainable economic growth for India: an exercise in marcro economic scenario building. Presidential Address, Indian Econometrics Society, Madras School of Economics, Madras, January 14–16, 2002
Patrick H (1966) Financial & economic growth in under development countries. Econ Dev Cult Chang 14:174–198
Rakshit M (1994) Issues in financial liberalization. Econ Polit Wkly 2547–2552
Rangarajan C (2002) Financial development and economic growth. Soc Sci Res J. ICSSR. New Delhi (June 2002)
Schumpeter JA (1911) The theory of economic development. Harvard University Press, Cambridge
Shaw ES (1973) Financial deepening in economic development. Oxford University Press, New York
Stiglitz JE (1994) The role of state in financial markets. Annual Conference on Development Economics (Proceedings of the World Bank) Supplement to the World Bank Economic Review and The World Bank Research Observer, Washington, D.C.: 204330, 19–57
Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. Amer Econ Rev 71:393–410
World Bank (1989) World Development Report

Chapter 25
Dynamics of the Indian Stock Market

Sadhan Kumar Chattopadhyay

1 Introduction

One of the most important national policy decisions during the late twentieth century and forepart of this century has been the financial liberalization of equity markets across the world. This has led to increased financial integration among the stock markets across the globe. This motivates international investors to look for new investment opportunities in order to improve risk-adjusted returns for their portfolios. While investors play in the domestic stocks, they are exposed to various domestic factors that originate within the country. But when they get into the international markets in order to diversify their investments, they are subjected to the exposure of various international factors which may not be similar to the domestic factors. Modern portfolio theory suggests that diversification of investments benefits investors only when the correlations between the assets included in the portfolio are low. Thus, incentives for investing into the international markets arise from lower correlations between asset returns as compared with that of domestic assets (Grubel 1968; Levy and Sarnat 1970). As world markets integrated and, hence, correlations between asset returns of the developed markets are increasing, international investors are also looking at the emerging economies for exploiting the benefit of international diversification with the belief that the correlation between the returns of developed markets and developing markets will be lower. This was supported by the earlier research by Solnik (1991), Divecha et al. (1992), Wilcox (1992) and recently by Driessen and Laeven (2007), Chang et al. (2008) and Gupta and Donleavy (2009).

Under this backdrop, the main objective of this chapter is to investigate the issue of stock market integration in India in the light of financial liberalization. Following the global trend, financial liberalization has also started in India since 1992. Increasing globalisation of the world economy should obviously have an impact on the behaviour of domestic stock markets (Cerny 2004). The relaxation of all types of

S. K. Chattopadhyay (✉)
M - 240, RBI Staff Quarters, North Avenue Santacruz (West),
Mumbai 400 054, India
e-mail: sadhanchattopadhyay@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_25, © Springer India 2014

economic barriers and developments in information technologies are, among others, expected to induce stronger stock market integration as opposed to stock market fragmentation. As well-developed and large financial markets contribute significantly to economic growth (see Arestis et al. 2001 and Beck et al. 2000), the development and integration of Indian financial markets is of particular importance. Further, the nature and extent of equity market integration is of importance for corporate managers as it influences the cost of capital and for investors as it influences international asset allocation and diversification benefits (e.g. Sentana 2002). Since the work of Grubel (1968) on expounding the benefits from international portfolio diversification, the relationship among national stock markets has been widely studied. Hence, the relationship among different stock markets has a great influence on investment because the diversification theory assumes that prices of different stock markets do not move together so that investors could buy shares in foreign as well as domestic markets and seek to reduce risk through global diversification. Under this backdrop, it is worth examining whether the Indian stock market has really integrated with the world markets. The chapter tries to contribute to the empirical literature by analysing the existence of a long-run relationship between Indian stock markets and several developed markets, viz. USA, UK, Japan, Singapore and Hong Kong. We select these stock markets because they are the most important in terms of size at the Asian level and USA and UK at the global level. As Table 25.1 indicates, these stock markets are large in comparison with their national economies as shown by the ratio of market capitalization to gross domestic product (GDP). On average, Hong Kong and Singapore have the largest capital markets relative to their own economy.[1] Stock to GDP ratio measures the development of stock markets. For Hong Kong, this ratio on average is 369.1 %, which is the highest amongst the sample of study. Turnover ratio measures value of stock transaction relative to the size of the markets, which is frequently measured as market liquidity. It may be observed that based on this measure, the Indian stock market has been the most liquid one during 2000–2011.

There is debate amongst the theorists about the factors which trigger financial market integration among countries. Chambet and Gibson (2008) argue that increasing openness contribute positively to financial market integration. Pretorious (2002), on the other hand, argues that a strong bilateral relationship may increase the level of interdependence between stock markets. Chen and Zhang (1997) opine that countries with strong economic ties in terms of trade tend to have financial markets that move together. Chinn and Forbes (2004) show that bilateral trade flows are an important determinant of cross-country linkages in stock markets. All countries considered in the present study are the most important trade partners in the international trade with India which is discussed later. Since the 2000s, trade among India and these countries increased significantly in terms of value (Table 25.2). However, the trade

[1] Freixas, Hartmann and Mayer (2008) points out that the size of capital market with respect to the economy is important given that economies with larger overall capital markets are able to provide easier financing for real investment.

Table 25.1 Feature of stock markets. (Source: World Development Indicators, World Bank)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Average
CAP/GDP ratio													
India	31.2	22.4	25.1	45.2	53.7	66.3	86.3	146.9	52.7	86.6	95.9	54.9	63.9
Hong Kong	368.6	303.8	282.7	347.6	401.0	390.1	471.4	561.4	617.0	437.6	481.0	365.1	418.9
Japan	66.7	54.1	53.4	70.7	79.0	103.6	108.5	102.2	66.4	67.1	74.7	60.3	75.6
Singapore	159.3	128.7	112.5	245.6	253.4	256.4	198.6	209.9	107.9	176.6	173.6	128.6	179.3
UK	174.5	147.2	115.7	132.2	127.9	134.1	155.2	137.2	70.3	128.8	138.0	49.4	125.9
USA	152.6	135.4	104.8	128.6	138.4	135.1	145.9	142.9	82.5	108.8	118.6	103.6	124.8
Stock/GDP													
India	107.4	50.6	37.7	46.1	52.5	52.0	67.3	89.4	85.8	80.0	62.7	40.1	64.3
Hong Kong	223.4	117.9	128.6	153.6	169.7	165.4	212.6	442.8	755.1	711.8	711.7	636.8	369.1
Japan	56.9	43.9	39.5	52.8	73.7	109.3	143.5	149.1	121.2	83.3	78.0	70.9	85.2
Singapore	95.4	69.5	62.0	94.1	74.4	97.0	132.5	228.1	162.4	143.4	132.4	105.9	116.4
UK	124.2	126.6	118.5	118.9	168.4	182.7	173.5	367.0	246.1	156.7	133.5	122.2	169.9
USA	321.9	283.8	239.6	140.2	164.1	171.2	249.9	305.2	450.2	337.1	210.8	203.7	256.5
Turnover													
India	306.5	192.9	163.3	138.9	113.7	92.2	93.1	84.0	85.2	119.3	75.6	56.3	126.7
Hong Kong	61.3	34.8	43.5	48.0	46.3	43.3	50.8	89.1	130.5	132.7	160.1	157.6	83.2
Japan	69.9	67.5	71.9	88.0	102.1	118.8	132.1	141.6	153.2	127.1	114.5	108.9	108.0
Singapore	52.1	46.9	51.2	53.1	32.1	40.4	62.2	122.0	101.6	102.8	82.9	74.8	68.5
UK	66.6	78.5	94.8	102.3	140.5	141.9	123.8	269.8	227.2	146.4	101.9	137.9	136.0
USA	200.8	200.6	203.4	122.6	126.5	129.2	182.8	216.5	404.1	348.6	189.1	187.6	209.3

CAP/GDP is the market capitalization of listed companies as percentage of GDP. Stock/GDP is the total values of shares traded as percentage of GDP. Turnover ratio is the total value of shares traded divided by the average market capitalization

Table 25.2 India's international trade, 2000–2010. (Source: Handbook of Statistics on Indian Economy 2011)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
India's exports to											
Hong Kong	2,640.9	2,366.4	2,613.3	3,261.8	3,691.8	4,471.3	4,680.6	6,305.2	6,607.6	7,862.1	11,419.8
Japan	1,794.5	1,510.4	1,864	1,709.3	2,127.9	2,481.3	2,862.7	3,853.8	3,002.1	3,613.3	5,216.5
Singapore	877.1	972.3	1,421.6	2,124.8	4,000.6	5,425.3	6,068.9	7,367.5	8,209.2	7,577.1	10,600.9
UK	2,298.7	2,160.9	2,496.4	3,023.2	3,681.1	5,059.3	5,618	6,698.2	6,597.6	6,213	7,181.3
USA	9,305.1	8,513.3	10,895.8	11,490	13,765.7	17,353.1	18,866.1	20,712	20,972.3	19,479.4	25,596
India's imports from											
Hong Kong	852.1	728.9	972.6	1,492.7	1,730.1	2,207	2,483.8	2,699.2	6,464.5	4,703.9	8,504.9
Japan	1,842.2	2,146.4	1,836.3	2,667.7	3,235.1	4,061.1	4,595.6	6,323.2	7,790.9	6,722.5	8,146.4
Singapore	1,463.9	1,304.1	1,434.8	2,085.4	2,651.4	3,353.8	5,489.5	8,117.6	7,514.5	6,454.7	6,693.7
UK	3,167.9	2,563.2	2,777	3,234.3	3,566.2	3,930.3	4,174.5	4,953.1	5,819.9	4,452.8	5,109.1
USA	3,015	3,149.6	4,443.6	5,034.8	7,001.4	9,454.7	11,736.1	21,019.3	18,441.5	16,985.4	18,531.2

Both export and import values are in millions of US dollars

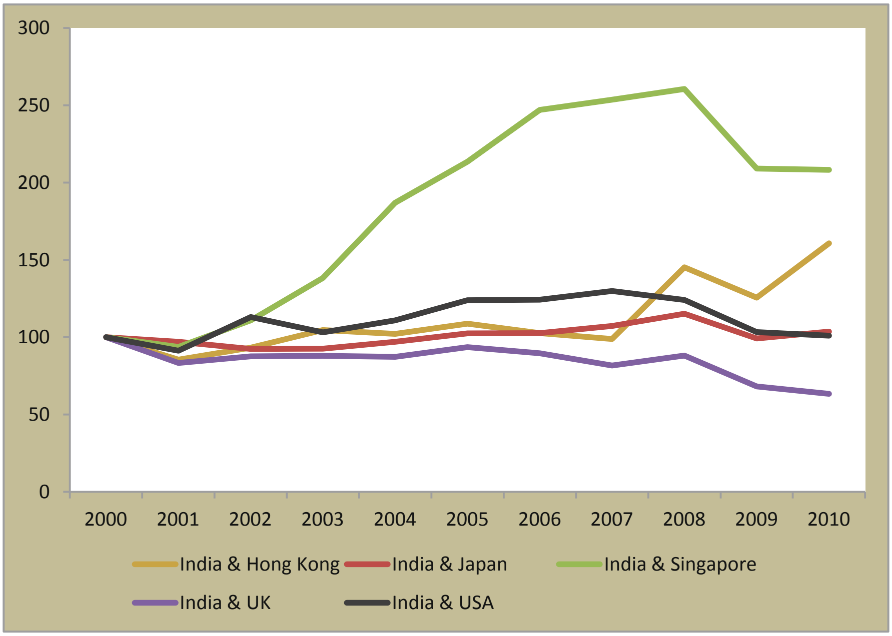


Fig. 25.1 Trade integration index, 2000–2010

integration[2] among India and countries considered in the present study do not seem to have a clear trend (Fig. 25.1). This could mean that we may expect a low level of integration among these countries.

The rest of the chapter is organised as follows. As a prelude to our statistical investigation, Sect. 2 explains the liberalization measures adopted in Indian stock markets since 1991 and the development of the stock market, which has taken place so far. While Sect. 3 deals with the survey of literature on stock market integration, Sect. 4 describes the methodology and data. Findings of the study are discussed in Sect. 5. Finally, Sect. 6 concludes the study.

2 Stock Market Reform in India and Its Aftermath

The Indian stock market is one of the earliest in Asia being in operation since 1875, but remained largely outside the global integration process until the late 1980s. A number of developing countries with the initiative of the International Finance Corporation and the World Bank started the reform process in the stock markets in order to mobilize finance in an effective way. In line with the global trend, the Indian

[2] Following Baele and Inghelbrecht (2009), trade integration is calculated as the ratio of imports plus exports over GDP. We then transform these values as indices starting at 100 at the beginning of the sample.

stock market also initiated the reform process in the financial market in general and stock market in particular. However, the critics argue that the stock market reform of the 1990s in India is an offshoot of the crisis erupted in 1992 owing to the infamous stock market scam (Shah and Thomas 2001). Thus, it is claimed that although reform process in India started with the establishment of the Securities and Exchange Board of India (SEBI) in 1988 to frame rules and guidelines for various operations of the stock exchange in India, it was not that active as it became after the post-scam period. Over the decade of the 1990s, a series of measures in the stock markets were taken. The stock markets introduced the best possible systems practised in advanced stock markets, viz., electronic trading system; dematerialization of shares, replacement of the Indian carry-forward trading system called '*badla*' by the index-based and scrip-based futures and options, rolling settlement in place of the account period settlement, adoption of risk management through 'novation' at the clearing corporation, etc. With the introduction of these advanced practices, transparency has also increased in the stock market. Further, among the significant measures of opening up of capital market, portfolio investment by foreign institutional investors (FIIs) such as pension funds, mutual funds, investments trusts, asset management companies, nominee companies and incorporated portfolio managers allowed since September 1992 have been the turning point for Indian stock markets. As of now, India is allowed to invest in all categories of securities traded in the primary and secondary segments and in the derivative segment. On the other hand, the ceiling on aggregate equity of FIIs including non-resident Indians (NRIs) and overseas corporate bodies (OCBs) in a company engaged in activities other than agriculture and plantation has been enhanced in phases. Further, with the financial sector reforms initiated in 1991, not only FIIs and NRIs are allowed to invest in Indian stock markets but Indian corporate have also been allowed to tap the global market with global depository receipts (GDRs), American depository receipts (ADRs) and foreign currency convertible bonds (FCCBs) since 1993. However, the company with a good track record is required to obtain prior permission from the Government of India in order to issue GDR/ADR/FCCB.[3]

With the automation and liberalization of Indian stock markets, there has been a perceptible change in the Indian stock market towards the latter part of the 1990s and forepart of the current decade. The trading system in Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) has, no doubt, reached a global standard. It has created a nationwide trading system that provides equal access to all investors irrespective of their geographical location. In that sense, technology has brought about equality among the investors across the country. This has resulted in phenomenal growth of the Indian stock market during the post-liberalization period. The number of shareholders and investors in mutual funds rose from 2 million in 1980 to 40 million in 1993 (Biswal and Kamaiah 2001). This makes the Indian investors' population the second largest in the world next to USA and largest in terms of companies listed, with nearly 7,985 companies listed by the end of 1995 (Misra 1997). Besides, the BSE is reported to have the highest density of transactions in the world

[3] For details of the stock market development, see Patil 2000.

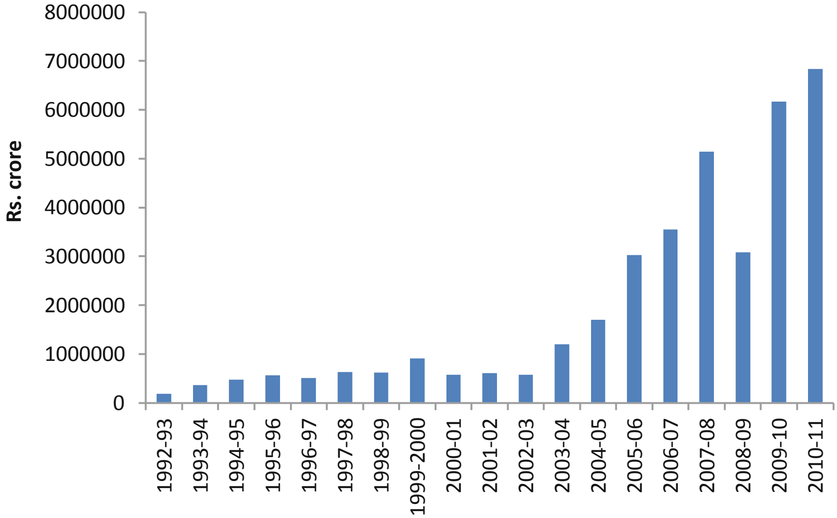


Fig. 25.2 Market capitalization of BSE (as at the end of March)

behind only Taiwan (Biswal and Kamaiah 2001). The daily turnover of shares in BSE increased substantially from Rs. 13 crore in 1980–1981 to Rs. 4,333 crore in 2010–2011. Due to the policy changes listed above, the market capitalization increased from Rs. 368,071 crore in 1993–1994 to Rs. 6,839,084 crore in 2010–2011 (Fig. 25.2). The total turnover that reflects the volume of business has also increased gradually over the years (Fig. 25.3).

Further, the market capitalization ratio, which is considered to be an important indicator of stock market size, gradually increased from 42.8 % in 1993–1994 to 146.9 % in 2007–2008, though it went down to 54.9 % in 2011–2012[4] (Table 25.1). Further, the value-traded ratio, the second development indicator which acts as a measure of liquidity of the stock market, also increased from 9.8 % in 1993–1994 to 89.4 % in 2007–2008, though it went down to 40.1 % in 2011–2012. Another important indicator of the stock market development is the turnover ratio, which complements the value-traded ratio in measuring the stock market liquidity, increasing from 23.0 % in 1993–1994 to 119.3 % in 2009–2010, but going down to 56.3 % in 2011–2012. The average daily trading volume on the Bombay stock market in the early 1990s was about the same as that in London—about 45,000 trades a day. The number of FIIs registered with SEBI increased from only 10 in January 1993 to 350 by the end of January 1996 and by the end of March 2011, the number increased to 1,722. Consequently, the liberal policies have led to increasing inflow of foreign investment in India in terms of portfolio investment increasing from Rs. 4.3 crore in 1992–1993 to Rs. 93,725.5 crore in 2011–2012 (Fig. 25.4). On the other

[4] In terms of economic significance, market capitalization as a proxy for market size is positively related to the ability to mobilize capital and diversify risk.

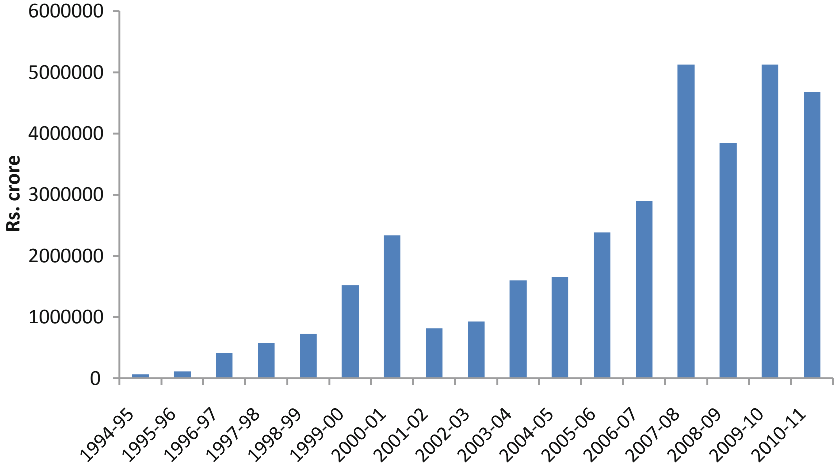


Fig. 25.3 Total turnover in BSE and NSE

hand, the process of integration received impetus further when the Indian companies were allowed to raise funds by issuing euro issues. As a result, starting with issue of Reliance in 1992, around 100 companies have so far taken advantage of the global market by raising funds of Rs. 133,691 crore as at the end of June 2011 (Table 25.3).

From the above analysis it is evident that the stock market in India has witnessed a phenomenal but uneven growth during the post-liberalization period. In other words, the deregulation and market liberalization measures and the increasing activities of multinational companies have accelerated the growth of the Indian stock market. Thus, given the newfound interest in Indian stock markets during the liberalization period, an intriguing question may obviously arise in one's mind as to how far India has gone down the road towards international stock market integration and whether any linkages have taken place among the stock indices of India and world's major stock indices. To answer these questions, we examine the interrelationship between Indian stock markets and major developed stock markets and study the underlying mechanism through which the Indian stock indices interact with international stock indices by analysing empirically the long-run pairwise, and multiple cointegration relationship and short-run dynamic Granger causality linkages between the Indian stock market and the major developed markets, viz., USA, UK, Japan, Singapore and Hong Kong.

3 Stock Market Integration: A Select Review of Literature

Although the study of financial integration dates back to the late 1970s, the number of studies was scanty during that time due to conservativeness of the stock markets. However, the financial markets, especially the stock markets, for developing

Table 25.3 Number and quantum of Euro issues. (Source: Handbook of Statistics on Indian Economy, Reserve Bank of India (RBI), 2011)

Year/Month	(Rs. crore)												
	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	Total
1992–1993	–	–	–	–	–	–	–	–	–	–	–	–	702.32 (2)
1993–1994	–	–	–	–	–	–	–	–	–	–	–	–	7,897.82 (27)
1994–1995	279.98 (1)	221.48 (2)	625.54 (3)	1,113.64 (4)	936.42 (5)	0.00	529.79 (2)	958.82 (6)	1,636.72 (7)	2.35	6.30	432.19 (1)	6,743.23 (31)
1995–1996	4.49 (1)	0.00	277.20 (1)	0.00	0.00	0.00	0.00	0.00	0.00	105.00 (1)	0.00	910.00 (2)	1,296.69 (5)
1996–1997	612.50 (2)	52.5	125.60 (2)	402.50 (1)	700.00 (1)	455.00 (2)	945.39 (2)	1,425.99 (2)	150.25 (2)	112.04 (1)	0.00	612.50 (1)	5,594.27 (16)
1997–1998	1,842.94 (1)	385.00 (2)	0.00	0.00	0.00	40.18 (1)	0.00	0.00	1,614.04 (2)	127.30 (1)	0.00	0.00	4,009.46 (7)
1998–1999	0.00	0.00	0.00	63.10 (1)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1,084.68 (2)	1,147.78 (3)
1999–2000	0.00	7.28 (1)	0.00	0.00	0.00	1,373.28 (1)	375.24 (1)	0.00	130.47 (1)	0.00	948.84 (1)	652.10 (1)	3,487.21 (6)
2000–2001	1,200.65 (2)	649.35 (2)	0.00	774.86 (2)	348.08 (1)	52.40 (1)	80.03 (1)	0.00	0.00	13.46 (1)	462.84 (1)	615.40 (2)	4,197.07 (13)
2001–2002	0.00	0.00	1,480.31 (3)	0.00	813.33 (1)	91.17 (1)	0.00	0.00	0.00	0.00	0.00	0.00	2,384.81 (5)
2002–2003	0.00	99.53 (1)	0.00	0.00	0.00	0.00	1,147.31 (4)	0.00	1,921.20 (2)	145.52 (2)	0.00	112.86 (2)	3,426.42 (11)
2003–2004	71.01 (1)	0.00	153.42 (1)	128.04 (2)	1,309.06 (1)	157.54 (2)	0.00	578.95 (3)	434.93 (3)	64.53 (1)	200.07 (4)	0.00	3,097.55 (18)
2004–2005	155.34 (1)	614.50 (3)	0.00	0.00	596.78 (2)	0.00	0.00	1,003.29 (2)	0.00	235.87 (3)	0.00	747.47 (4)	3,353.25 (15)

Table 25.3 (continued)

Year/Month	(Rs. crore)												
	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	Total
2005–2006	59.00 (2)	1,514.23 (3)	261.06 (1)	292.66 (4)	374.34 (2)	1,327.13 (5)	2,510.44 (8)	1,320.84 (2)	1,164.82 (7)	701.40 (6)	572.61 (3)	1,258.99 (6)	11,357.52 (49)
2006–2007	1,957.84 (10)	2,656.41 (5)	1,171.96 (6)	1,332.06 (6)	0.00	798.25 (2)	234.52 (3)	344.41 (2)	344.99 (1)	6,968.77 (1)	1,086.92 (2)	108.98 (2)	17,005.11 (40)
2007–2008	45.40 (1)	18.33 (1)	1,187.42 (1)	8,200.86 (2)	1,694.74 (3)	137.60 (2)	2,388.12 (1)	626.74 (3)	10,672.54 (3)	982.11 (4)	346.90 (1)	255.63 (4)	26,556.39 (26)
2008–2009	2,151.41 (4)	1,901.30 (3)	2.82 (1)	29.71 (1)	566.72 (1)	0.00	34.50 (2)	0.00	0.00	0.00	0.00	101.90 (1)	4,788.36 (13)
2009–2010	166.81 (1)	0.00	47.73 (1)	48.48 (1)	4,618.31 (4)	7,763.48 (1)	446.44 (1)	1,774.13 (2)	298.60 (2)	348.40 (4)	0.00	454.99 (1)	15,967.37 (18)
2010–2011	694.86 (3)	581.51 (4)	3,567.26 (12)	2,010.00 (7)	0.00	590.00 (4)	284.00 (2)	540.50 (4)	222.40 (2)	526.45 (2)	0.00	424.60 (2)	9,441.58 (42)
2011–2012	719.67 (6)	300.83 (5)	216.69 (2)	.	.	.	.	.	.	.	.	.	1,237.19 (13)

1) ‘–’ indicates nil
2) Data for the year 2011–2012 are provisional
3) Data for January 1995, February 1995 and May 1996 represent the amount of warrants exercised by the investors attached to original GDRs issued and not to a fresh GDR issue
4) Figures in brackets indicate number of issues

and developed markets have now become more closely interlinked despite the uniqueness of the specific market and country profile. This has happened specifically due to financial liberalization adopted by most of the countries around the world, technological advancement in communications and trading systems, introduction of innovative financial products and creating more opportunities for international portfolio investments. This has intensified the curiosity among the academics in exploring international market linkages.

Earlier studies by Ripley (1973), Lessard (1976) and Hillard (1979) found low correlation between national stock markets supporting the benefits of international diversification. Applying the vector autoregression (VAR) models, Eun and Shim (1989) found the evidence of co-movements between the US market and other world equity markets. Cheung and Ng (1992) examined the dynamic properties of stock returns in Tokyo and New York and found that the US market is an important global factor from January 1985 to December 1989. Lee and Kim (1994) examined the effect of the October 1987 crash, concluded that national stock markets became more interrelated after the crash and found that the co-movements among national stock markets were stronger when the US stock market was more volatile. Applying the VAR approach and the impulse response function analysis, Jeon and Von-Furstenberg (1990) show that the degree of international co-movement in stock price indices has increased significantly since the 1987 crash. On the other hand, Koop (1994) used Bayesian methods and concluded that there are no common trends in stock prices across countries. Further, Corhay et al. (1995) studied the stock markets of Australia, Japan, Hong Kong, New Zealand and Singapore and found no evidence of a single stochastic trend for these countries. Syriopoulos and Roumpis (2009) examined both short- and long-run relationships among central Europe (CE) emerging markets (Poland, the Czech Republic, Hungary and Slovakia) and several developed markets (viz., Germany and USA) during the period 1999–2003. Results show the presence of long-run relationship among these markets. Economic reforms, the impact of the European Monetary Union (EMU) and consistent foreign direct investment inflows in the CE economies can be considered relevant factors explaining the presence of a cointegration relationship among CE and developed stock markets. In a study of Sri Lanka and Asian developed markets for the period 1989–1994, Elyasiani, Perera and Puri (1998) found that there was no interdependence between the Sri Lankan and other stock markets.

Although there is no dearth of literature on financial integration, there are only a few studies existing on this area in the case of India. Not only that; most of the studies are very old and were carried out during the time when Indian stock markets were not open to the world. For instance, Sharma and Kennedy (1977) examined the price behaviour of the Indian market with the UK and US markets and concluded that the behaviour of the Indian market is statistically indistinguishable from that of the US and UK markets and found no evidence of systematic cyclical component or periodicity for these markets. Applying cross-spectral analysis, Rao and Naik (1990) found that the relation between the Indian stock market and international markets is weak. Ignatius (1992) compared returns on the BSE Sensex with those on the New York Stock Exchange (NYSE) Standard & Poor's (S&P) 500 Index and found no evidence of integration. Agarwal (2000), with a correlation coefficient of 0.01 between Indian and developed markets, concluded that there is a lot of scope for the Indian stock market to integrate with the world market. Hansda and Ray (2002) found that Nasdaq and other technology-oriented indices of the NYSE have their influence on the domestic stock prices. By using the BSE 200 data, Wong et al. (2005) have found that the Indian stock market is integrated with the matured markets of the world. As mentioned above, some of the studies are age-old and have lost relevance especially after the opening up of the economy to the rest of the world since the early 1990s, from which the relationship between the Indian stock market and international markets may have changed. Some other studies except Wong et al. (2005) which are relatively new have not done any cointegration analysis to examine the long-run relationship. Although Wong et al. (2005) have studied the stock market integration, they have taken BSE 200 data and also have dealt with monthly data which have their own limitation.[5] Hence, our chapter revisits the issue of nature of co-movement between the developed and emerging markets.

4 Methodology

To test for Granger causality and cointegration, we use the standard methodology proposed by Granger (1969, 1968) and Engle and Granger as described in Enders (2004). All tests are performed on natural logarithm of the indices' time series using the ordinary least squares (OLS) estimation procedure.

In order to test for Granger causality among stock market indices y_t^I and y_t^f, we estimate the equation

$$\Delta y_t^I = \alpha_0 + \sum_{i=1}^{n} \alpha_{1i} \Delta y_{t-i}^I + \sum_{i=1}^{m} \alpha_{2i} \Delta y_{t-1}^f + \varepsilon_{1t}$$

$$\Delta y_t^f = \beta_0 + \sum_{i=1}^{n} \beta_{1i} \Delta y_{t-i}^f + \sum_{i=1}^{m} \beta_{2i} \Delta y_{t-1}^I + \varepsilon_{2t}$$

and perform an F test for joint insignificance of the coefficients. The null hypothesis claims that y_t^f does not Granger cause y_t^I or vice versa. Therefore, a rejection of the null hypothesis indicates the presence of Granger causality. For each pair of stock market indices, we perform two Granger causality tests so that we can decide whether y_t^f Granger causes y_t^I or y_t^I Granger causes y_t^f or both, or none.

In order to examine the co-movement between the Indian stock market and the developed markets, we strictly follow the standard methodology available in the literature. We first study the relationship between Indian stock markets and foreign markets by the simple regression

$$y_t^I = \alpha + \beta y_t^f + e_t, \tag{25.1}$$

[5] The main limitation is that BSE 200 data are not representative of the Indian stock market.

where the endogenous variable y_t^I represents India's stock index; the exogenous variable y_t^f is the stock index of the foreign markets; and e_t is the error term. In order to examine the joint effect of all important markets on the Indian market, we study the multiple regression

$$y_t^1 = \alpha + \beta_1 y_t^{f1} + \beta_2 y_t^{f2} + \beta_3 y_t^{f3} + \beta_4 y_t^{f4} + \beta_5 y_t^{f5}, \tag{25.2}$$

where y_t^{fi} are the stock indices for the USA, the UK, Japan, Hong Kong and Singapore, for $i = 1, 2, 3, 4$ and 5, respectively.

The validity and reliability of the regression relationship require the examination of the trend characteristics of the variables and cointegration test as the presence of unit root processes in the stock indices results in the spurious regression problem. Before testing for cointegration, we need to go for stationary test. In order to do so, we apply the Dickey and Fuller (1979, 1981) (DF) and augmented Dickey–Fuller (ADF) unit root tests based on the following regression:

$$\Delta y_t = b_0 + a_0 t + a_1 y_{t-1} + \sum_{i=1}^{p} b_i \Delta y_{t-1} + \varepsilon_t, \tag{25.3}$$

where $\Delta y_t = y_t - y_{t-1}andy_t$ can be y_t^I, y_t^f ory_t^{fi}, ε_t is the error term. Regression (25.3) includes a drift term (b_0) and a deterministic trend (a_0t).

In addition, we apply the PP test developed by Phillips and Perron (1988) to detect the presence of a unit root. The PP test is non-parametric with respect to nuisance parameters and, thereby, is suitable for a very wide class of weakly dependent and possibly heterogeneously distributed data.

If both $y_t^I, y_t^f (y_t^{fi})$ are of the same order, say $I(d)$, $d > 0$, we then estimate the cointegrating parameter in (25.1) or (25.2) by OLS regression. If the residuals are stationary, the series, $y_t^I$and$y_t^f (y_t^{fi})$, are said to be cointegrated. Otherwise, $y_t^I$and$y_t^f (y_t^{fi})$ are not cointegrated.

Cointegration exits for variables means despite variables are individually non-stationary, a linear combination of two or more time series can be stationary and there is a long-run equilibrium relationship between these variables. If the error term in (25.1) or (25.2) is stationary while the regressors are individually trending, there may be some transitory correlation between the individual regressors and error term. However, in the long run, the correlation must be zero because of the fact that trending variables must eventually diverge from stationary ones. Thus, the regression on the level of the variables is meaningful and not spurious.

The most common tests for stationarity of estimated residuals are Dickey–Fuller (CRDF) and augmented Dickey–Fuller (CRADF) tests based on the regression

$$\Delta \hat{e}_t = \gamma \hat{e}_{t-1} + \sum_{i=1}^{p} \gamma \Delta \hat{e}_{t-1} + \xi_t, \tag{25.4}$$

where $\hat{e}_t$are residuals from the cointegrating regression (25.1) or (25.2) and p is chosen to achieve empirical white noise residuals for CRADF and set to zero for CRDF test.

We further apply the multivariate cointegrated system developed by Johansen (1988a, b). Assume each component $y_{i,t}$ $i = 1, \ldots ., k$, of a vector time series process y_t is a unit root process, but there exists a $k \times r$ matrix β with rank $r < k$ such that $\beta' y_t$ is stationary. Clive Granger has shown that under regularity conditions we can write cointegrated process y_t as a vector error correction model (VECM):

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \cdots + \Gamma_{p-1} \Delta y_{t-(p-1)} - \Pi y_{t-p} + \varepsilon_t, \tag{25.5}$$

where $\varepsilon_t \sim$ iid(0, Ω). The basic idea of the Johansen procedure is simply to decompose Π into two matrices α and β, both of which are $k \times r$ such that $\Pi = \alpha\beta'$ and so the rows of β may be defined as the distinct cointegrating vectors. Then a valid cointegrating vector will produce a significantly non-zero eigenvalue and the estimate of the cointegrating eigenvector. Johansen proposes a trace test for determining the cointegrating rank r, such that

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i) \quad r = 0,1,2,\ldots\ldots\ldots.\text{n} - 1, \tag{25.6}$$

and proposes another likelihood ratio test whether there is a maximum of r cointegrating vectors against $r + 1$ such that

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}), \tag{25.7}$$

with critical values given in Johansen (1995).

4.1 Data

We have taken daily BSE Sensitive Index (Sensex) comprising 30 most sensitive scrips. BSE Sensex is considered as the 'core barometer' of the Indian stock market for a number of reasons, viz., (1) oldest stock exchange in Asia, (2) it is the premier bourse with the largest listing, (3) it attracts a major chunk of the foreign institutional investment and (4) popularity (Hansda and Ray 2002). We have used daily data in order to capture potential interactions, for example, impulse responses, because a month or even a week may be long enough to obscure interactions that may last only a few days (Cotter 2004).[6] Our sample covers the period from September 1, 1999, to August 3, 2012, a total of 3,373 observations. The sample consists of daily closing stock indices of India (BSE 30), Hong Kong (Hang Seng), Japan (Nikkei 225), USA (S&P 100), UK (Financial Times and the London Stock Exchange, FTSE) and Singapore (Straits Time Index, STI). All indices have been obtained from Datastream Direct of Thomson Reuters and they are in domestic currency in order to avoid problems associated with transformation due to fluctuations in exchange rates. Table 25.4 shows that during the sample period, the BSE index had the highest average rate of

[6] Cotter, John (2004): 'International Equity Market Integration in a Small Open Economy: Ireland January 1990-December 2000', International Review of Financial Analysis, 13 (2004) 669–685.

Table 25.4 Summary statistics for equity index returns

	BSE 30	Hang Seng	Nikkei 225	S&P 100	FTSE	STI
N. obs	3,372	3,372	3,372	3,372	3,372	3,372
Mean	0.0377	0.0111	– 0.0217	– 0.00255	– 0.00295	0.00991
Maximum	15.9900	13.4068	13.2346	10.6551	9.3843	7.5305
Minimum	– 11.8092	– 13.5820	– 12.1110	– 9.1862	– 9.2656	– 8.6960
Std. Dev.	1.6399	1.5983	1.5236	1.3308	1.2849	1.2674
Skewness	– 0.13861	– 0.05874	–0.40696	– 0.1023	– 0.1429	– 0.24993
Kurtosis	9.469152	10.71456	10.24791	10.04763	8.862367	7.496898
Jarque–Bera test	5,890.713	8,363.718	7,473.849	6,984.395	4,840.089	2,876.308
Probability	0.00	0.00	0.00	0.00	0.00	0.00

All daily returns were calculated as log differences using daily closing prices

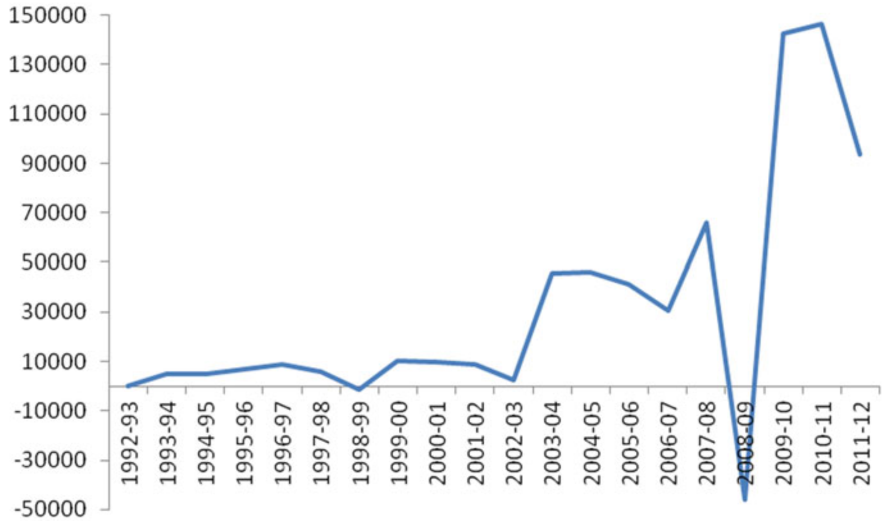


Fig. 25.4 Net foreign institutional investor (FII) investment in India

returns followed by the Hang Seng index. The standard deviation of returns on BSE is higher than that in Hang Seng, Nikkei, S&P100, FTSE or STI. All returns have negative skewness implying that the distribution has a long right tail, while kurtosis values are high in all cases implying that the distributions are peaked relative to normal. The Jarque–Bera test statistic indicates that returns of the stock markets are not normally distributed for the sample period used in the study.

Figure 25.5 plots both the index values and returns for India, Hong Kong, Japan, Singapore, USA and UK. Almost all the indices including BSE Sensex observed a steep fall during the period 2000–2003. From 2003 to 2007, an upward trend is common across all markets. From the second half of 2007, we observe a large decline of stock prices across all markets, whereas some rise is observed during the second quarter of 2009. Asian as well as US and UK stock market returns show evidence of

Table 25.5 Correlations of stock index returns (September 1, 1999, to August 3, 2012)

	BSE 30	STI	Hang Seng	Nikkei	S&P 100	FTSE
Panel A: 1999–2012						
BSE 30	1					
STI	0.479	1				
Hang Seng	0.459	0.696	1			
Nikkei	0.315	0.524	0.578	1		
S&P 100	0.191	0.223	0.19	0.116	1	
FTSE	0.317	0.408	0.307	0.291	0.52	1
Panel B: 1999–2006						
BSE 30	1					
STI	0.305	1				
Hang Seng	0.289	0.588	1			
Nikkei	0.237	0.430	0.489	1		
S&P 100	0.056	0.155	0.118	0.113	1	
FTSE	0.153	0.303	0.289	0.207	0.425	1
Panel C: 2006–2012						
BSE 30	1					
STI	0.604	1				
Hang Seng	0.565	0.765	1			
Nikkei	0.370	0.59	0.633	1		
S&P 100	0.288	0.271	0.237	0.119	1	
FTSE	0.433	0.48	0.414	0.349	0.587	1

volatility clustering, that is, as pointed out by Campbell et al. (1997), small returns are followed by more small returns, called low volatility periods, and large returns tend to be followed by more large returns, called high volatility periods. The 2007–2009 financial crisis appears to be a longer and more intense period of high volatility during the decade reflecting large uncertainty and loss of confidence among market participants. We also observe that India being a developing country was equally affected by the crisis which is evident from the high volatility of returns during the period negating the so-called decoupling theory. As we can see in Fig. 25.5, almost all stock market returns reached the highest level of volatility in September 2008 when the US government decided to put two government-sponsored enterprises (GSEs), i.e. Fannie Mae and Fannie Mac, into the control of the Federal Housing Finance Agency (FHFA) (Gupta et al. 2012).

Table 25.5 reports the correlation of the returns of the four Asian markets along with US and UK markets. It is observed that the highest correlation is between the STI and Hang Seng (more than 69 %) while, surprisingly, the lowest correlation was observed between S&P and Nikkei (about 12 %).[7] Partitioning the period into two sub-periods, we observe that correlations among stock markets change over time. It is observed that correlations among India and all the markets have increased during the second half of the period of our study (Table 25.5, panel C).

[7] Roll (1992) argues that the lowest correlations among international stock markets may be due to reasons like indices construction, differences in the industrial structure as well as in the conduct of national monetary policies.

Table 25.6 Results of unit root tests

	Lag length p	ADF	P value[a]	Bandwidth	PP	P value[a]
Variables in log levels						
BSE 30	1	– 0.5934	0.8697	11	– 0.552997	0.8783
STI	0	– 1.0768	0.7271	4	– 1.129319	0.7063
Hang Seng	0	– 1.5593	0.5034	2	– 1.543337	0.5116
Nikkei	0	– 1.8990	0.3331	12	– 1.785179	0.3883
S&P 100	2	– 2.1929	0.2090	10	– 2.144358	0.2273
FTSE	5	– 2.1138	0.2393	0	– 2.477625	0.1211
Variables in first log difference						
BSE 30	0	– 55.1778	0.0001	14	– 55.14555	0.0001
STI	0	– 56.5663	0.0001	2	– 56.56761	0.0001
Hang Seng	0	– 58.9900	0.0001	1	– 58.99009	0.0001
Nikkei	0	– 58.8373	0.0001	13	– 59.02482	0.0001
S&P 100	1	– 45.2396	0.0001	12	– 64.48869	0.0001
FTSE	4	– 28.1828	0.0000	4	– 61.02846	0.0001

5 Empirical Results

A necessary condition to perform a cointegration test is that the order of integration of variable has to be the same. In order to detect the order of integration we employed two unit root tests, that is the ADF test (Dickey und Fuller 1979) and Phillips–Perron (PP) test (Phillips und Perron 1988). Unit root tests results are shown in Table 25.6. The null hypothesis of a unit root is not rejected for all indices in log levels, whereas it is rejected when they are taken in their log first differences.

5.1 Engle Granger Cointegration Test

If stock markets have a cointegration relationship, the residual error series of each of the equations estimated in the first step should have stationarity, which have been discussed above in the methodology section. Results in Table 25.7 show that there is no long-run relationship between BSE and other Asian equity markets except the Singapore stock market. This is also not found between BSE and US and UK markets. This is done by using the specification of the ADF with intercept but without trend. While ADF test with trend and intercept is used, weak evidence of cointegrating relationship between BSE and Hong Kong stock and also Singapore stock markets is found. Further, when ADF test without trend and intercept is used, a weak cointegrating relationship between BSE and Hong Kong stock market is observed, while a strong long-run relationship is observed between BSE and Singapore stock market. However, no evidence of long-run relationship between Indian and developed markets, viz., the Japanese, US or UK stock markets, was found. Since the ADF test is performed on residual from regression equation, there is no need to include an intercept term; we consider the results of ADF test without trend and intercept more

Table 25.7 ADF test results on Engle–Granger cointegration test residuals

Stock markets	ADF test statistics with intercept	ADF test statistics with trend and intercept	ADF test statistics without trend and intercept
BSE 30 and Hang Seng	– 2.3955	– 3.3402**	– 2.3959**
BSE 30 and STI	– 2.6740*	– 3.0576**	– 2.6743***
BSE 30 and Nekkei	– 0.6084	– 1.9910	– 0.6089
BSE 30 and S&P	– 0.5555	– 1.9953	– 0.5559
BSE 30 and FTSE	– 1.0833	– 2.4356	– 1.0830
BSE 30, Hang Seng, Nikkei, STI, S&P 100 and FTSE	– 4.5045*	– 4.6777*	– 4.5053*

The critical value for both the ADF and PP t-statistics are – 3.43, – 2.86 and – 2.56 at 1, 5 and 10 % levels of significance, respectively. For both tests, a constant term was included. For the ADF test, the optimal lag lengths are determined by using Schwarz Baysian Criterion (SBC) with a maximum lag of 28. For the PP test, the spectral estimation method is Bartlett kernel, while bandwidth is the Newey–West.[a]MacKinnon (1996) one-sided *P* values In the ADF test, critical values are – 3.432, – 2.862 and – 2.567 on models with intercept, – 3.961, – 3.411 and – 3.1247 on models with trend and intercept and – 2.5656, – 1.9409 and – 1.6166 on models with neither trend nor intercept for 1, 5 and 10 % levels of significance. ***, ** and * indicate rejection of the null hypothesis at the 1, 5 and 10 % levels

reliable.[8] We infer that there is a long-run relationship between BSE and Asian stock markets except Japan, while no relationship exists between BSE and the US or UK stock markets. Although CRDF and CRADF are significant at 1 % level of significance for all variables taken together, this may not be justified to conclude that they are integrated. This is because the cointegration test is based on the ADF test, which is known to have a low power (Cerny 2004).[9] Not only that, testing cointegration with the help of ADF test with more than two variables may not give the correct result.

5.2 *Johansen Cointegration Test*

In order to further investigate the cointegration results, we employ Johansen's cointegration test, which is considered to be the best measure for cointegration. As Johansen (1988) is a powerful way of analysing complex interaction of causality and structure among variables in a system, this process is further applied to determine whether any cointegration relationship exists among the Indian, US, UK, Japanese, Singapore and Hong Kong stock markets as all the markets are integrated of order one (Table 25.5). Johansen's procedure requires estimating a VAR (p). In order to estimate the optimal number of lag p of the VAR, we used both the Akaike Information Criterion (AIC) and the Schwarz Criterion (SC). When we estimated the VAR with BSE and Nikkei in log

[8] See Enders (2004), pp. 336.

[9] Cerny, Alexandr (August 2004): 'Stock Market Integration and the Speed of Information Transmission', Working paper series (ISSN 1211–3298), Center for Economic Reasearch and Graduate Education, Academy of Sciences of the Czech Republic, Economic Institute.

Table 25.8 Tests for the cointegrating vectors

	λ_{trace}	Critical value 5 %	λ_{max}	Critical value 5 %
Panel A: Bivariate Johansen cointegration results				
BSE 30 and Nikkei 225 market indices				
$r = 0$	3.593	15.495	3.400	14.265
$r \leq 1$	0.194	3.841	0.194	3.841
BSE 30 and Hang Seng market indices				
$r = 0$	10.092	15.495	9.737	14.265
$r \leq 1$	0.355	3.841	0.355	3.841
BSE 30 and STI market indices				
$r = 0$	8.554	15.495	8.035	14.265
$r \leq 1$	0.519	3.841	0.519	3.841
BSE 30 and S&P 100 market indices				
$r = 0$	6.786	15.495	5.660	14.265
$r \leq 1$	1.126	3.841	1.126	3.841
BSE 30 and FTSE market indices				
$r = 0$	7.320	15.495	6.528	14.265
$r \leq 1$	0.792	3.841	0.792	3.841
Panel B: Bivariate Johansen cointegration results				
BSE 30, Hang Seng, Nikkei, STI, S&P100 and FTSE				
$r = 0$	71.450	95.754	29.294	40.078
$r \leq 1$	42.156	69.819	18.088	33.877
$r \leq 2$	24.069	47.856	13.643	27.584
$r \leq 3$	10.426	29.797	5.933	21.132
$r \leq 4$	4.493	15.495	3.866	14.265
$r \leq 5$	0.626	3.841	0.626	3.841

The 5 % critical values provided by MacKinnon, Haug and Michelis (1999) indicate no cointegration

form, the AIC selects a VAR with 7 lags, while the SIC selects a model with 2 lags.[10] We estimated the VAR (2) model selected by the SC given that it is the more parsimonious in terms of coefficients to estimate. However, checking for the serial correlation of the residual series through the autocorrelation Lagrange multiplier (LM) test, we reject the null hypothesis of no serial correlation. In order to eliminate the serial correlation, we estimated a VAR (4). Checking for the serial correlation, we were not able to reject the null hypothesis of no serial correlation of the residual series. We also find that BSE and STI indices seem to be best represented by a VAR of order 4. Similarly, a VAR with 4 lags was estimated for BSE and Hang Seng equity indices. For BSE and S&P 100, VAR (5) did not reject null hypothesis of no correlation in the residual series using LM test. In the case of BSE and FTSE indices, VAR (8) based on AIC gives a more robust result with no serial correlation in the residual series. Further, in the case of BSE with all the equity indices, VAR (10) was found to be more appropriate with respect to serial correlation. After estimating the VAR models with the optimal number of lags, we were able to conduct the Johansen cointegration test at both bivariate and multivariate levels. Empirical findings depicted in Panel A and Panel B of Table 25.8

[10] Theoretical rationale for differing lag length is the different weight assigned to the penalty term for the number of parameters (Lütkepohl 2005).

Table 25.9 Granger causality test for stock returns, 1999–2012

	F-statistic	Probability
BSE 30 does not Granger cause Hang Seng market	20.06	0.00
Hang Seng Market does not Granger cause BSE 30 market	4.01	0.02
BSE 30 does not Granger cause Nikkei market	53.97	0.00
Nikkei does not Granger cause BSE 30 market	1.34	0.26
BSE 30 does not Granger cause STI market	6.62	0.00
STI does not Granger cause BSE 30 market	7.39	0.00
BSE 30 does not Granger cause S&P 100 market	0.58	0.56
S&P 100 does not Granger cause BSE 30	69.51	0.00
BSE 30 does not Granger cause FTSE market	0.77	0.46
FTSE does not Granger cause BSE 30	31.01	0.00

does not support the presence of cointegrating vector in the BSE and Nikkei stock markets. The null hypothesis that BSE and Nikkei are not cointegrated ($r = 0$) against the alternative of one cointegrating vector $r \leq 1$ is not rejected, since both the λ_{trace} and λ_{max} statistics do not exceed the critical values with 5 % level of significance.

We found no evidence of cointegration on a bivariate basis between the Indian and other developed markets. We also test if these markets, as a group, could be integrated. A multivariate Johansen test was carried out to test if these markets as a group are cointegrated. The results (Panel B of Table 25.8) indicate that there is no long-term relationship among the stock markets. The absence of long-run relationship shows an evidence of limited financial integration among these markets. As pointed out by Chinn and Forbes (2004) where economic and industrial structures in countries differ, then the degree of financial integration between markets can also be different. Further, absence of a cointegrating relationship suggests that in the long run, stock prices are not driven by a common international risk factor in all markets.

5.3 Granger Causality Test

Given that the indices are difference stationary and the cointegration results do not show clear evidence of robust cointegration between them, we use a further methodology. Égert and Kočenda (2007) argue that given the lack of cointegration among markets, a valid tool remains the well-known Granger causality test in order to identify the relationship among these markets. Results (Table 25.9) show that there is bilateral or feedback causality between BSE and Hang Seng stock market. Bidirectional causality is also observed between BSE and STI stock market. The direction of causality is from BSE to Nikkei; however, there is no reverse causation from Nikkei to the BSE stock index. We found that there is unidirectional causality between BSE and S&P 100 index, which reveals that the latter Granger causes the former. This is also true for BSE and FTSE, where it is found that FTSE Granger causes BSE but not the reverse.

We have not done any exercise on structural breaks in the series of stock indices. This is because if at all any break is observed in the series the break dates will

Table 25.10 Granger causality test for stock returns (before and after the global financial crisis)

	Before crisis		After crisis	
	F-statistic	Probability	F-statistic	Probability
BSE 30 does not Granger cause Hang Seng market	2.30	0.100	26.09	0.00
Hang Seng Market does not Granger cause BSE 30 market	5.12	0.006	1.62	0.20
BSE 30 does not Granger cause Nikkei market	14.75	0.000	47.92	0.00
Nikkei does not Granger cause BSE 30 market	1.25	0.287	0.57	0.57
BSE 30 does not Granger cause STI market	2.344	0.096	5.89	0.00
STI does not Granger cause BSE 30 market	3.131	0.044	6.29	0.00
BSE 30 does not Granger cause S&P 100 market	0.92	0.40	2.12	0.12
S&P 100 does not Granger cause BSE 30	54.31	0.00	21.60	0.00
BSE 30 does not Granger cause FTSE market	1.42	0.24	4.94	0.01
FTSE does not Granger cause BSE 30	27.88	0.00	12.34	0.00

be heterogeneous, which can be seen from Gupta et al. (2012). However, we test whether the recent global financial crisis has affected causality between India's stock market return and those of other developed markets of Asia, Europe and USA. Kenc and Dibooglu (2010) has pointed out that the 2007–2009 global crisis started in USA on July 2007 as a consequence of the collapse of two Bear Stearns hedge funds. The financial crisis caused a slowing of the growth in large emerging economies in Asia including China and India in early 2008 (Fidrmuc and Korhonen 2010). However, as indicated by Bartram and Bodnar (2009), the global equity market crisis can be dated around September 15, 2008, the day of the Lehamn Brothers bankruptcy. In order to test whether crisis has affected the causality due to global crisis, we divide our sample into two subsamples. The first one is from September 1, 1999, to September 14, 2008, while the second one is from September 15, 2008, to August 3, 2012. The Granger causality test for both samples show that before the crisis BSE market did not cause Hang Seng, while after the crisis BSE did cause Hang Seng, which affected BSE before the crisis but not after the crisis. It is also observed that STI market did not cause BSE before the crisis, while after the crisis it does cause, indicating bidirectional causality between the two indices. Overall, during the recent global financial crisis, the Granger causality results indicate that not only has the influence of Asian markets as well as the American and UK markets on the Indian stock market remained significant as during the period before the crisis, but also the BSE market has had a large influence on the other developed markets except Japan after the crisis (Table 25.10). Analysing the interdependence among Indian and international stock markets during the period 2000–2007, Dicle et al. (2010) have shown strong causal dependence with international stock markets confirming our results coming from whole sample as well as for the two subsamples we considered. Thus, it goes against the findings of Gupta et al. (2012), who observed that the influence of Asian markets as well as the US markets on Indian markets after the crisis has disappeared. Rather, our results show that after the crisis the influence of the Indian market on other developed markets has become much stronger because of increasing importance of Indian economy at both regional and global level. In fact, after the crisis Indian economy has shown its resilience before the world during the crisis period, which helped in building confidence among the foreign investors.

The results between the US and Indian stock markets are obvious since the US market is the world's foremost securities market and has a heavy influence on other stock markets. Hence, one may not be surprised that the US stock market Granger causes the Indian stock market in the short run (25.9 and 25.10). More rationally, several macroeconomic factors may give a good explanation of the causal relationship between two stock markets. They include economic connection, regulatory structures similarity, exchange rate policy and trade flows. Coinciding with the start of the liberalization of the Indian economy, there has been a steady improvement in India–USA trade relations during the last decade. The US government has identified India as one of the 10 major emerging markets (Wong et al. 2005). The volume of India–USA bilateral trade also started growing at a steady pace with the export from India to the USA growing from US$ 1,209.5 million in 1980 to US$ 25,596 million in 2010.

On the other hand, the India–USA trade volume still remains a small fraction of USA's global trade. While USA's export to India accounts for more than 10 % of India's non-oil imports and USA is the destination of one fifth of India's exports, USA's trade turnover with India constitutes less than 1 % of its global trade. India's percentage share in US imports has remained stable over the last few years; it was 0.7 % during 2010. In 2000, India ranked 21st among countries that export to the USA. These figures show that US economy is very important to the Indian economy, though the reverse is not true. This seems to be consistent with our result of unidirectional causality from S&P 100 to BSE Sensex.

The results in Table 25.9 also reveal the evidence of short-run impact of the UK stock market on the Indian stock market. It may be noted that after the opening up of the Indian economy since 1991 the bilateral trade between India and UK has been constantly increasing. UK continues to be India's other important trading partner and continues to be the largest cumulative investor in India and third largest investor in the post-1991 period. As Indian economy is linked with UK's economy closely, it is not surprising that the UK stock market does have an impact on the Indian stock market.

However, no evidence of short-run impact of the Japanese stock market on the Indian stock market can be found from Table 25.9, although the Indian stock market appears to have an influence on the Japanese market. It may be mentioned that although there has been an increase in the volume of trade between India and Japan in absolute terms in percentage, it has gone down.

On the other hand, the share of Hong Kong in export has gone up from 2.1 % in 1980 to 4.5 % in 2010. The evidence of short-run impact can be found from Table 25.9. The Hong Kong stock market is found to have an influence on the Indian stock market, and the reverse is also true.

Table 25.11 Variance decompositions for the Indian stock market

Days ahead	*Explained by innovations in*											
	S&P 100		FTSE		NIKKEI		HANG SENG		STI		SENSEX	
	I	II	I	II	I	II	I	II	I	II	I	II
1	5.1	0.0	4.5	0.0	3.6	0.0	8.5	0.0	3.2	0.0	75.2	100.0
2	9.1	2.5	4.3	1.6	3.6	0.1	8.2	0.2	3.1	0.2	71.6	95.4
5	9.8	3.0	4.3	1.9	3.6	0.1	8.2	0.6	3.2	0.4	70.9	94.0
10	9.7	3.2	4.6	2.1	3.7	0.2	8.4	0.8	3.2	0.5	70.4	93.2
15	10.0	3.6	4.6	2.1	4.0	0.5	8.4	1.0	3.7	1.0	69.3	91.8
20	10.3	3.8	4.6	2.2	4.1	0.6	8.5	1.3	3.9	1.1	68.7	91.0

Finally, the share export of India to Singapore has gone up from 1.6 % in 1988 to 4.2 % in 2010. The evidence of short-run impact can be found in Table 25.9 and it is bidirectional.

Since there are no linear combinations of the stock indices that are stationary, there is no error correction representation. This brings us to the issue of a dynamic relationship between the Indian equity markets and the other market. The dynamic relationship is broken into two areas of investigation, viz. variance decomposition and impulse response functions. Table 25.11 gives variance decompositions for the Sensex returns equation of the VAR for 1, 2, 5, 10 and 20 steps ahead for the two variable orderings:[11]

Order I: S&P 100, FTSE, NIKKEI, HANG SENG, STI, SENSEX
Order II: SENSEX, STI, HANG SENG, NIKKEI, FTSE, S&P 100

The results show that the ordering of the variables is important in the decomposition. Thus, two orderings are applied, which are the exact opposite of one another, and the sensitivity of the results is considered. It is clear that by the 2-year forecasting horizon, the variable ordering has become almost irrelevant in most cases. The variance decompositions which show the proportion of the movements in the dependent variables that are due to their own shocks, versus shocks to the other variable, seem to suggest that the US, UK, Japanese, Singapore and Hong Kong markets are to a certain extent exogenous in the system. An interesting feature of the result is that shocks to the US and Hong Kong markets together account for only 5–19 % of the variance of the Indian stock market. All the stock markets together account for only 9–31 % of the variance of the Sensex (Table 25.11).

Turning to the impulse response estimates, Fig. 25.6 provides normalized responses for the Sensex for a typical shock to and from the Indian market. These responses represent unit shocks measured standard deviations. As can be seen from the results, innovations to S&P 100 have a positive impact on the Sensex up to the fifth day, since the impulse response is positive up to this period, but the effect of the shock dies down after 5 days. However, innovations to the rest of the stock markets,

[11] As a consequence of the effect of ordering on the variance decomposition, when theory does not suggest an obvious ordering of the series, some sensitivity analysis should be undertaken (Brooks 2002; pp. 358). In this case, the exercise has been repeated by reversing the order of the series.

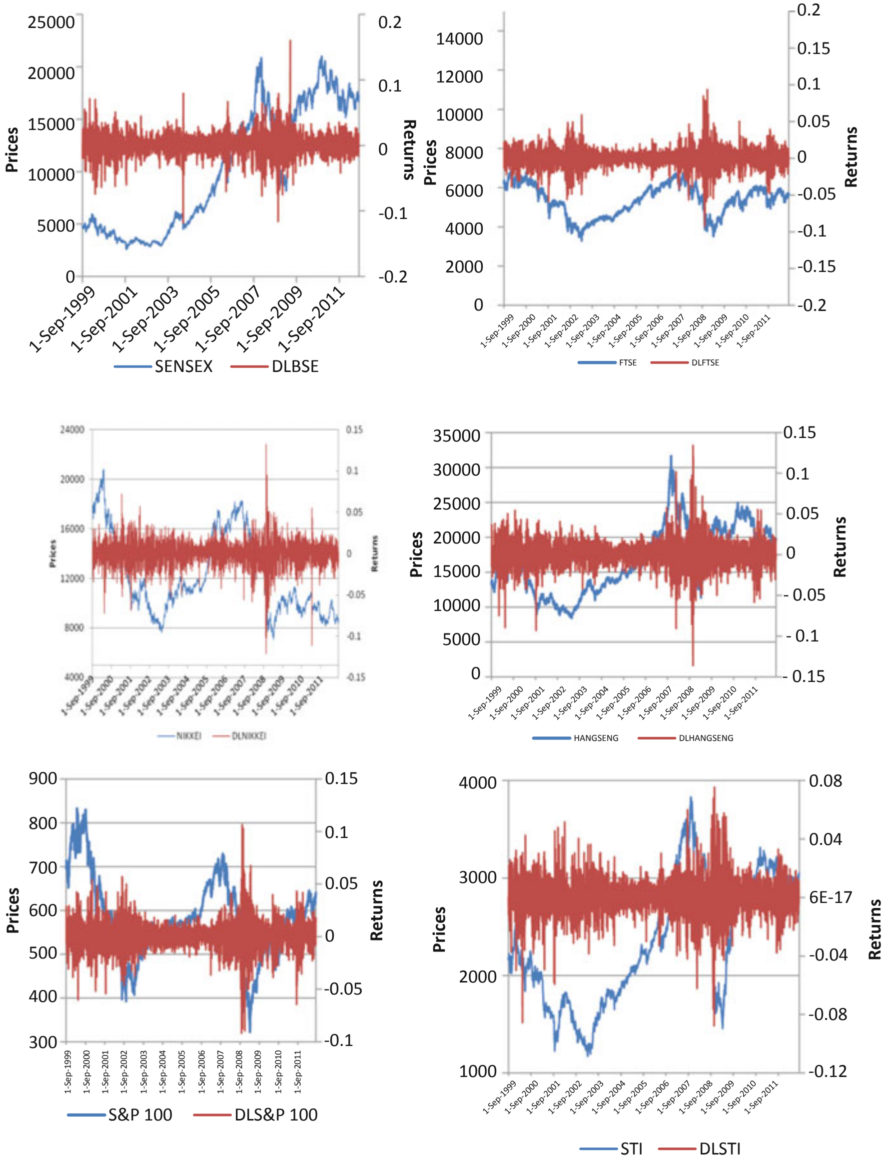


Fig. 25.5 Daily prices and returns of sample stock indices

except Hang Seng and Nikkei, have a positive impact on the Indian stock markets, *albeit* short; the effects of the shocks die down on the second day itself. Only in the case of Hang Seng and Nikkei, effects remained positive at the initial stage but subsequently became negative, but die down after the second period. Thus, in a nutshell, it is observed that the shock in one market does have an impact on the Indian stock

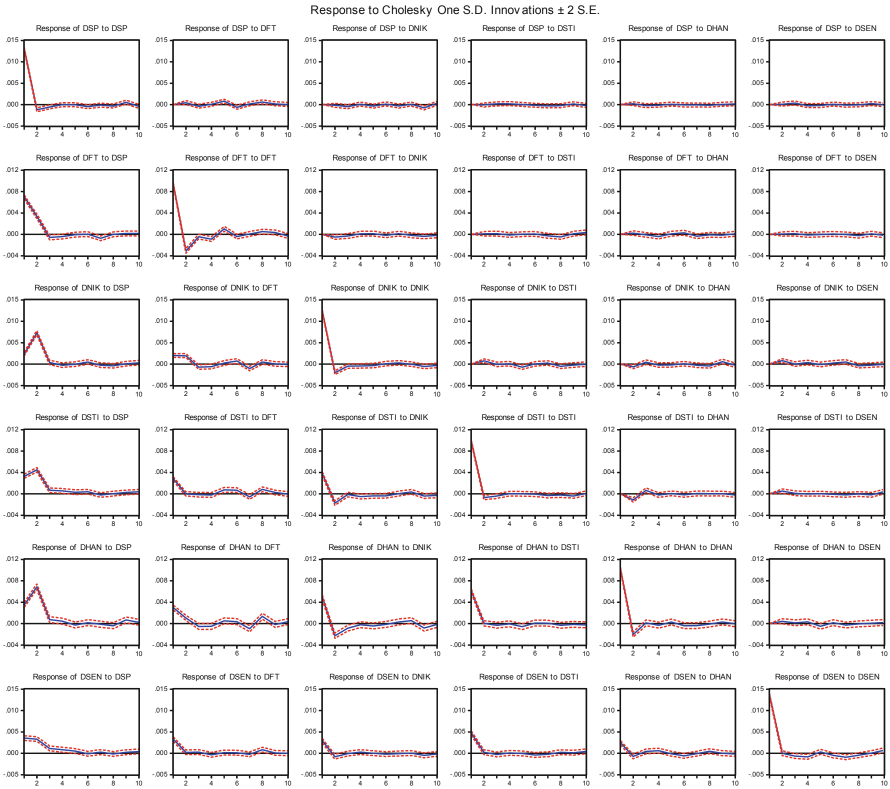


Fig. 25.6 Impulse responses for the SENSEX

market. Such a result implies that the possibility of making excess returns by trading in one market on the basis of 'old news' from another market appears to be unlikely.

Thus, the results show the absence of a stable cointegrating relationship among the stock markets considered in the study. This has had a number of implications which are the following:

1. The absence of a stable relationship can be recognized or incorporated by the investors or analysts in the formulation of their models of stock markets behaviour (Gupta et al. 2012).
2. The Indian market can be an ideal destination for the international investors, since Indian stock markets are partially immune to external shocks, which was tested during the global financial crisis in 2008 (Goel et al. 2010). It was observed that there was no first-generation impact of global crisis on Indian economy.
3. The length of low-volatility period is longer in India than the other markets, which implies that adverse events may not cause an abrupt outflow of capital from India.

Perhaps, transparency in the stock markets and clarity of rules governing stock markets play a major role for stabilizing the stock markets. This may provide further incentive to investors for considering the stock market of India as a potential market for inclusion in their diversified portfolios. This also provides an indication that the monetary and fiscal policies as implemented during the recent past have worked well. In fact, India's wait-and-watch policy has strengthened the resilience of the country. This, in a way, has helped the foreign investors' confidence in the Indian stock market, which in turn, has made the stock market remain more buoyant.

6 Concluding Observations

This chapter investigated the relationship between Indian and Asian, US and UK equity markets over the period 1999–2012. By applying the unit root test, we find that all stock market series are nonstationary. Engle and Granger cointegration tests do not find evidence of a cointegrating relationship among these stock markets. Johansen cointegration tests also confirm that a long-run relationship between these markets does not exist. India is one of the emerging economies, which has witnessed significant development in the stock markets during the recent periods due to the liberalization policy initiated by the government. It is generally believed that due to the liberalization policy and the consequent development of Indian stock markets, the latter might have integrated with the developed markets. One may argue that due to this integration, which appears to have taken place after liberalization, the Indian stock market would mainly be governed by a common factor as in the case of the developed markets. However, our study does not support this view. Rather, it finds that the Indian stock market is not integrated with the world markets. Hence, we may conclude that the Indian stock market is not influenced by other markets. Of course, some short-term sentiment in the world market does have an impact but this is short lived. That means the prerequisites, which are required for a long-run relationship, have not been achieved by India so far. Further, since the Indian stock market is not affected by the common factors, this may act as a good destination for the foreign investors to park their funds in the Indian stock market during times of high volatility or major crisis. Of course, the major prerequisite for the Indian stock market to remain vibrant is the strong fundamentals of the overall economy.

Views expressed here are those of the author and do not necessarily reflect the views of the organisation he is working with.

References

Agarwal RN (2000) Financial integration and capital markets in developing countries: a study of growth, volatility and efficiency in the Indian capital market. Institute of Economic Growth, Delhi

Arestis Philip, Demetriades Panicos O, Luintel Kul B (2001). Financial development and economic growth: the role of stock markets. J Money Credit Bank 33(1):16–41

Baele L, Inghelbrecht K (2009) Time-varying integration and inteofrnational diversification strategies. J Empirical Finance 16:368–387

Beck T, Levive R, Loeysa N (2000) Finance and the sources growth. J Financ Econ 58(1-2):261–300

Bartram SM, Bodnar GM (2009) No place to hide: the global crisis in equity markets in 2008/2009. J Int Money Finance 28:1246–1292

Biswal PC, Kamaiah B (2001) Stock market development in India. Econ Polit Wkly 36:(04), 27 January 2001.

Brooks C (2002) Introductory econometrics on finance. Cambridge University Press, UK

Cerny A (2004) Stock market integration and the speed of information transmission. CERGE-EI Working Papers wp 242. The Center for Economic Research and Graduate Education - Economic Institute, Prague

Chambet A, Gibson R (2008) Financial Integration, Economic Instability and Trade Structure in Emerging Markets. J Int Money Finance 27:654–675

Campbell JY, Lo AW, MacKinlay AC (1997) The econometrics of financial markets. Princeton University Press, Princeton, NJ

Chang CP, Chen PF, Lee CC. (2008) Money demand function vs. monetary integration: revisiting panel cointegration among GCC countries. Math Comput Simul 79(1):85–93

Chen N, Zhang F (1997) Correlations, trades and stock returns of the pacific-basin markets. Pacific-Basin Finance J 5:559–577

Cheung Y-W, Ng LK (1992) Stock Price dynamics and firm size: an empirical investigation. J Finance 47:1985–1997

Chinn M, Forbes K (2004) A decomposition of gobal linkages in financial markets over time. Rev Econ Stat 86(3):705–722

Corhay A, Rad A, Urbain J (1995) Long-run behaviour of pacific-basin stock prices. Appl Finan Econ 5:11–18

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

Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49(4) (July 1981)

Dicle MF, Beyhan A, Yao LJ (2010) Market efficiency and international diversification: evidence from India. Int Rev Econ Finance 19:313–339

Divecha AB, Drach J, Stefak D (1992) Emerging markets: a quantitative perspective. J Portfol Manage 18(1):30–32

Driessen J, Laeven L (2007) International portfolio diversification benefits: cross-country evidence from a local perspective. J Banking Finance 31(6):1693–1712

Égert B, Kočenda E (2007) Interdependence between Eastern and Western European Stock Markets: Evidence from intraday data. Economic Systems

Enders W (2004) Applied Econometric Time Series. Wiley Series.

Elyasiani E, Perera P, Puri TN (1998) Interdependence and dynamic linkages between stock markets of Sri Lanka and its trading partners. J Multinat Finan Manage 8:89–101

Eun CS, Shim S (1989) International Transmission of Stock Market Movements. J Finan Quant Anal 24:41–56

Fidrmuc J, Korhonen I (2010) The impact of the global financial crisis on business cycles in Asian emerging economies. J Asian Econ 21:293–303

Freixas X, Hartmann P, Mayer C (2008) Handbook of European financial markets and institutions. Oxford University Press, Oxford

Goel R, Chattopadhyay SK, Rawat PS, Prasad G, Kumar N (2010) Impact of global crisis on indian economy. Reserve Bank of India Staff Studies.

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438

Granger CWJ (1986) Developments in the study of cointegrated economic variables. Oxford Bull Econ Statist 48:213–228

Grubel HG (1968) Internationally diversified portfolios: welfare gains and capital loss. Amer Econ Rev 58(5):1299–1314

Gupta R, Guidi F (2012) Cointegration relationship and time varying co-movements among Indian and Asian developed stock markets. Int Rev Finan Anal 21:10–22

Gupta R, Donleavy GD (2009) Benefits of diversifying investments into emerging markets with time-varying corelations: an Australian perspective. J Multinat Finan Manage 19(2):160–177

Hansda SK, Ray P (2002) BSE and Nasdaq - globalisation, information technology and stock prices. Economic and Political Weekly. February 2:2002
Hillard J (1979) The relationship between equity indices on world exchanges. J Finance (March):103–114
Ignatius R (1992) The Bombay stock exchange: seasonalities and investment opportunities. Indian Economic Review 28(2):223–227
Jeon B, Furstenberg Von (1990) Growing international co-movement in stock price indexes. Q Rev Econ Finance 30(30):17–30
Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dynam Control 12:231–254
Johansen S (1995) Likelihood-inference in cointegration vector auto-regressive models. Oxford University Press, Oxford
Kenc T, Dibooglu S (2010) The 2007-2009 financial crisis, global imbalances and capital flows: implications for reforms. Econ Systems 34:2–31
Koop G (1994) An objective bayesian analysis of common stochastic trends in international stock prices and exchange rates. J Empirical Finance 1:343–364
Lee SB, Kim KJ (1994) Does the October 1987 Crash Strengthen the Co-movement in Stock Price Indexes. Quart Rev Econ Bus 3(1-2):89–102
Lessard DR (1976) World, country and industry factors in equity returns: implications for risks reduction through international diversification. Financ Anal J 32:32–38
Levy H, Sarnat M (1970) International diversification of investment portfolios. Amer Econ Rev 60(4):668–675
Lütkepohl H (2005) New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin Heidelberg
MacKinnon JG (1996) Numerical distribution functions for unit root and cointegration tests. J Appl Econometrics 11:601–618
MacKinnon JG, AA Haug, L Michelis (1999) Numericla distribution functions of likelihood ratio tests for cointegration. J Applied Economterics 14:563–577
Misra BM (1997) Fifty years of indian capital market. Reserve Bank of India occasional papers, 18, 2 & 3.
Patil RH (2000) The capital market in 21st Century. Econ Polit Wkly 35(47):18–24
Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75: 335–346
Pretorious E (2002) Economic determinants of emerging stock market interdependence. Emerg Mark Rev 3:84–105
Rao BSR, Naik Umesh (1990) Inter-relatedness of stock market spectral investigation of USA, Japan and Indian markets note. Artha Vignana 32(3,4):309–321
Ripley DM (1973) Systematic elements in the linkage of national stock market indices. Rev Econ Statist 55(3):356–361
Roll R (1992) Industrial structure and the comparative behaviour of international stock market Indices. J Finance 47:3–42
Sentana E (2002) Did the EMS reduce the cost of capital? Econ J 112(October):786–809 (Royal Economic Society).
Shah A, Thomas S (2001) The evolution of the securities markets in India in the 1990s. Technical Report, IGIDR, September.
Sharma JL, Kennedy RE (1977) Comparative analysis of stock price behaviour on the Bombay, London & New York stock exchanges. J Financ Quant Anal. September.
Solnik B (1991) International investments. Addison-Wesley, Workingham
Syriopoulos T, Roumpis E (2009) Dynamic correlations and volatility effects in the balkan equity markets. J Int Financ Mark Institutions Money 19(4):565–587
Wilcox JW (1992) Taming frontier markets. J Portfol Manage 18(1):51–56
Wong W, Agarwal A, Du J (2005) Financial integration for indian stock market: a fractional cointegration approach. National University of Singapore, Dept. of Economics, Working Paper No. 050.

Chapter 26
Analysis of Revenue Efficiency: Empirical Study of Indian Non-Life Insurance Companies

Anirban Dutta and Partha Pratim Sengupta

1 Introduction

The proponents of liberalised insurance regime advocated that there was a wide gap in terms of market potential and its exploitation by the nationalised companies. The consumer did not benefit in the absence of competition in terms of wider choice and competitive pricing and the reach of the nationalised companies were limited, the range of products offered were restricted, and the service to the consumers inadequate. It was felt in the 1990s that the scale of economic activity attained in the mid-1980s, the momentum generated through the reforms process in other sectors of the economy cannot be sustained by state-controlled insurance industry, and the insurance penetration and enlargement of the market can be accomplished only when a large number of companies compete with each other. It was also realised that the objectives of nationalisation of the industry could largely be accomplished through appropriate regulatory measures and a state monopoly was no longer necessary. The initiators of change finally prevailed and the Insurance Regulatory and Development Authority Act (IRDA Act) was notified in April 2000.

The issue of efficiency is a key concern of policy makers for two reasons. First, state insurance firms have to be reformed so that they can provide better services to people and become competitive after the domestic market is fully opened for competition. Second, the market has to be more diversified, so that more insurers will be allowed to run businesses, reducing the monopolistic or oligopolistic powers of existing players. The long-term objective will be the expansion of the insurance market and encourage wider participation so as to maximise the social benefits of insurance for the entire population by increasing the insurance penetration.

A. Dutta (✉)
NSHM Business School, Durgapur, West Bengal, India
e-mail: anirbandutta3@rediffmail.com

P. P. Sengupta
National Institute of Technology (NITD), Durgapur, West Bengal, India
e-mail: pps42003@yahoo.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_26,

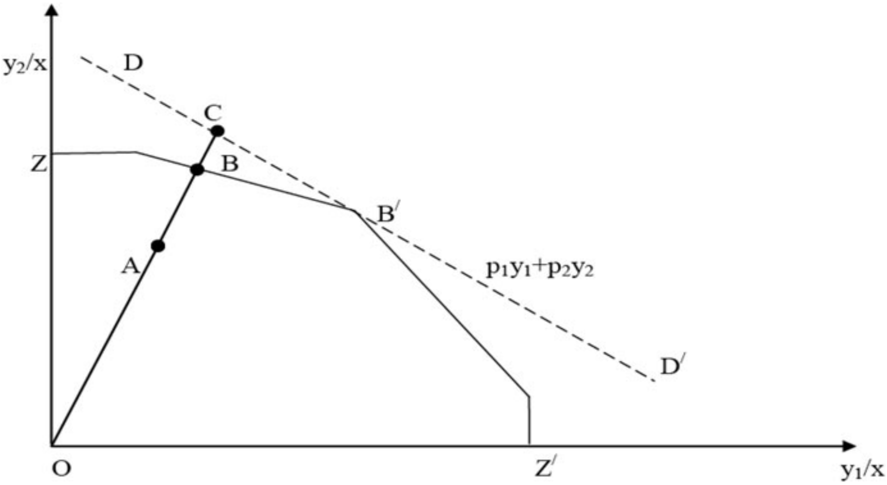


Fig. 26.1 Measuring output-oriented technical, allocative, and revenue efficiency

Profitability, return on investment (ROI), and other financial ratios are highly relevant as performance measures; however, they are not sufficient to evaluate operating efficiencies. Hence, more insightful methods are desired. The more recently developed frontier techniques provide more comprehensive and reliable information than using a set of operating ratios and profit measures. In utilising these methods, management is within a framework that supports the planning, decision-making, and control processes. The frontier methodologies measure firm performance relative to 'best practice' frontiers derived from efficient firms in the industry. Such methods are considered to be superior in comparison to traditional techniques such as financial ratio analysis, since they summarise firm performance in a single statistic that controls differences among firms using a sophisticated multidimensional framework. Frontiers have been estimated to measure firm success in maximising revenue (revenue efficiency).

2 Concept of Revenue Efficiency[1]

Let us consider output-oriented measures where production involves two outputs (y_1 and y_2) and a single input (x). This is depicted in Fig. 26.1, where ZZ′ represents unit piecewise production possibility set and the point A corresponds to an inefficient firm.

The Farrell output-oriented technical efficiency measure (Färe et al. 1985) is defined as the ratio:

$$TE_o = OA/OB.$$

[1]The discussion is based on Coelli et al. (2005).

If we have the information on output prices (p_1 and p_2), then we can draw the iso-revenue line, DD′, and define the *revenue efficiency* (RE) as the ratio:

$$\mathrm{RE} = \mathrm{OA}/\mathrm{OC}.$$

DD′ represents the highest revenue which can be achieved by keeping input level same. Revenue efficiency can be decomposed as:

$$\mathrm{RE} = \mathrm{OA}/\mathrm{OC} = \mathrm{OA}/\mathrm{OB} \times \mathrm{OB}/\mathrm{OC} = \mathrm{TE_o} \times \mathrm{AE_o},$$

which has a revenue efficiency interpretation. Again, the values of all these three measures are bounded within zero and one.

3 Methodology

Most of the input and output data related to insurance sector are price based, for example, premium earned, commission paid, investment, etc. Now, because of the unavailability of different individual price data and suitable proxies in the Indian context, it is really a complicated job to consider the unit price of each input or output items that has been considered for this research. Therefore, to measure the 'revenue efficiency', here we introduce the most appropriate model, i.e. new cost-based data envelopment analysis (DEA) measurement as suggested by Tone (2002).

The concepts dealing with 'allocative efficiency' were introduced by Farrell (1957) and Debreu (1951) with the common unit prices assumption. This was transformed into linear programming formulation by Färe et al. (1985).

Tone (2002) identified the serious shortcoming in the traditional Farrell-Debreu cost and allocative efficiency measure and suggests a new scheme for evaluating cost and revenue efficiency under different unit prices that is explained below.

Let us consider a cost-based production possibility set P_c as:

$$Pc = \{(\bar{x}, y) \mid \bar{x} \geq \bar{X}\lambda,\ y \leq Y\lambda,\ \lambda \geq 0\}, \tag{26.1}$$

where $\bar{X} = (\bar{x}_1, \ldots\ldots, \bar{x}_n)$ with $\bar{x}_j = (c_{1j}x_{1j}, \ldots\ldots\ldots, c_{mj}x_{mj})^T$.

Here, we assume that the elements of $\bar{x}_{ij} = (c_{ij}x_{ij})\ \forall(i, j)$ are denominated in homogeneous units like rupee or dollar, so that adding up the elements of $\bar{x}_{ij}$ has a well-defined meaning.

The new revenue efficiency $\bar{\rho}_\mathrm{o}^*$ is defined as :

$$\bar{\rho}_\mathrm{o}^* = e\bar{y}_\mathrm{o}/e\bar{y}_\mathrm{o}^* \tag{26.2}$$

The optimal solution $\bar{y}_\mathrm{o}^*$ can be obtained from the following LP problem:

$$\begin{array}{ll} [N\ \text{Revenue}] & e\bar{y}_\mathrm{o}^* = \max\limits_{\bar{y},\lambda} e\bar{y} \\ \textit{subject to} & x_\mathrm{o} \geq X\lambda \\ & \bar{y} \leq \bar{Y}\lambda \\ & L \leq e\lambda \leq U \\ & \lambda \geq 0, \end{array} \tag{26.3}$$

where price-based output $\bar{Y} = (\bar{y}_1, \ldots\ldots, \bar{y}_n)$ with $\bar{y}_j = (p_{1j}y_{1j}, \ldots\ldots, p_{sj}y_{sj})$. The value of revenue efficiency is also bounded with one and zero.

4 Specification of Input, Output, and Sample Selection

In the financial service sector, three principal approaches have been used to measure outputs: the *asset* or *intermediation approach*; the *user-cost approach*; and the *value-added approach* (Berger and Humphrey 1992). The *intermediation approach* views the insurance company as a financial intermediary that manages a reservoir of assets, borrowing funds from policyholders, investing them on capital markets, and paying out claims, taxes, and costs (Brocket et al. 1998). The *user-cost method* differentiates between input and output based on the net contribution to revenue. The *value-added approach* counts output as important if they contribute a significant benefit, based on operating cost allocations (Berger et al. 2000). Usually, several types of outputs are defined, representing the single lines of business under review. This study uses *premiums earned* (y_1) as the output for risk bearing/risk pooling service following the value-added approach and *income from investment* (y_2) as the intermediation output. There are three main insurance inputs: *labour*, *business service and materials,* and *capital*. We have used *expenses related to labour* (x_1) as a close proxy of labour, *expenses related to business service and materials* (x_2) as a close proxy of business services and materials, and *total investment* (x_3) as a close proxy of equity and debt capital which are representing major three inputs. The detailed description of the output and input is mentioned in Table 26.1.

The sample of this research comprises 12 non-life insurers within the period of 2005–2006 to 2009–2010. The list of the non-life insurers is given in Table 26.2. Data were collected from IRDA annual reports.

5 Result and Analysis

To find out the result, we have used the new revenue efficiency DEA model by Tone (2002) as explained above in both situations of constant returns to scale (CRS) and variable returns to scale (VRS). Table 26.3 represents the year wise and overall revenue efficiency for non-life insurance business over the study period of 2005–2006 to 2009–2010. Some interesting findings are as follows:

1. The overall revenue efficiency for the life insurance sector is 86 and 96.4 % under the CRS and VRS assumption, respectively. This implies that there is a scope of 14 and 3.6 % improvement of revenue under both the assumptions, respectively.
2. Under CRS assumption, the overall revenue efficiency of public non-life insurers (0.756) is considerably lower than private non-life insurers (0.912) for the entire study period, but these scores (0.956 and 0.969, respectively) marginally differ under VRS case.

Table 26.1 Description of input and output variables

Non-life insurance business		
Outputs		
1. Premiums (y_1)	Gross premium from direct business written[a]	Risk pooling/bearing service followed by value-added approach
2. Investment income (y_2)	Gross interest, dividends and rent + profit on sale/redemption of investments of the policy holder and share holder accounts	Intermediary services followed by intermediation approach
Inputs		
1. Expenses related to labour (x_1)	Gross direct commission paid[b] to the agents/intermediaries plus employee's remuneration and welfare fund	Labour (agents + employee)
2. Expenses related to business service and materials (x_2)	Gross direct claims incurred[c] + operating expenses related to insurance business excluding employee's remuneration and welfare fund	Business service and materials
3. Total investment (x_3)	Investment of policy holders and share holders fund	Capital (equity and debt)

[a] A huge amount has to be spent as premium on reinsurance ceded and adjustment for change in reserve for unexpired risks in case of non-life insurance business, so we have taken gross amount to represent actual output

[b] Because of the high amount of reinsurance cede in non-life insurance business, sometimes net commission turns into negative value (as per the available data), which is not logical as input in a business

[c] High amount of reinsurance cede to claims paid reduce the value of claims cost (net). Therefore, we have taken gross amount of claims incurred

Table 26.2 List of the insurers considered for the research work

Non-life insurance companies
Public insurers
1. New India Assurance Co. Ltd.
2. National Insurance Co. Ltd.
3. United India Insurance Co. Ltd.
4. The Oriental Insurance Co. Ltd.
Private insurers
1. Royal Sundaram Alliance Insurance
2. IFFCO-TOKIO General Insurance Co.
3. Bajaj Allianz General Insurance Co.
4. ICICI Lombard General Insurance Co.
5. Reliance General Insurance Co.
6. TATA AIG General Insurance Co. Ltd.
7. Cholamandalam MS General Insurance Co.
8. HDFC ERGO General Insurance Co. (Earlier HDFC General Insurance Co. from 27.9.2000 to 5.4.2008

Table 26.3 Revenue efficiency of non-life insurers under CRS and VRS assumptions. (Source: Author's calculation)

No.	Non-life insurers	Revenue efficiency										Overall revenue efficiency	
		2005–2006		2006–2007		2007–2008		2008–2009		2009–2010		2005–2010	
		CRS	VRS	CRS	VRS	CRS	VRS	CRS	VRS	CRS	VRS	CRS	VRS
1	New India	1	1	0.570	1	0.879	1	0.906	1	0.746	1	0.820	1
2	Oriental	0.867	0.992	0.548	0.988	0.837	0.921	0.813	0.967	0.679	0.870	0.749	0.948
3	National	0.746	0.960	0.471	0.894	0.764	0.891	0.785	0.971	0.702	0.900	0.694	0.923
4	United	0.784	0.965	0.485	0.883	0.860	0.948	0.899	0.970	0.779	1	0.761	0.953
5	Royal Sundaram	0.894	0.894	0.757	0.791	0.866	0.917	0.927	0.940	0.835	0.861	0.856	0.880
6	Bajaj Allianz	1	1	0.829	1	1	1	1	1	0.823	0.944	0.931	0.989
7	IFFCO-TOKIO	1	1	1	1	1	1	1	1	1	1	1	1
8	ICICI Lombard	1	1	0.997	1	0.999	1	0.956	1	0.844	1	0.959	1
9	TATA AIG	0.840	0.866	0.792	0.805	0.925	1	0.910	0.912	0.870	0.898	0.867	0.896
10	Reliance	1	1	1	1	1	1	0.932	1	1	1	0.986	1
11	Cholamandalam	0.812	0.981	0.722	0.935	0.931	1	0.943	1	0.872	1	0.856	0.983
12	HDFC ERGO	0.824	1	0.624	1	0.769	1	1	1	1	1	0.843	1
Avg. revenue efficiency		**0.897**	**0.971**	**0.733**	**0.941**	**0.902**	**0.973**	**0.923**	**0.980**	**0.846**	**0.956**	**0.860**	**0.964**
No. of revenue efficient firms		**5**	**6**	**2**	**6**	**3**	**8**	**3**	**5**	**3**	**7**		
No. of revenue inefficient firms		**7**	**6**	**10**	**6**	**9**	**4**	**9**	**7**	**9**	**5**		

CRS constant returns to scale, *VRS* variable returns to scale

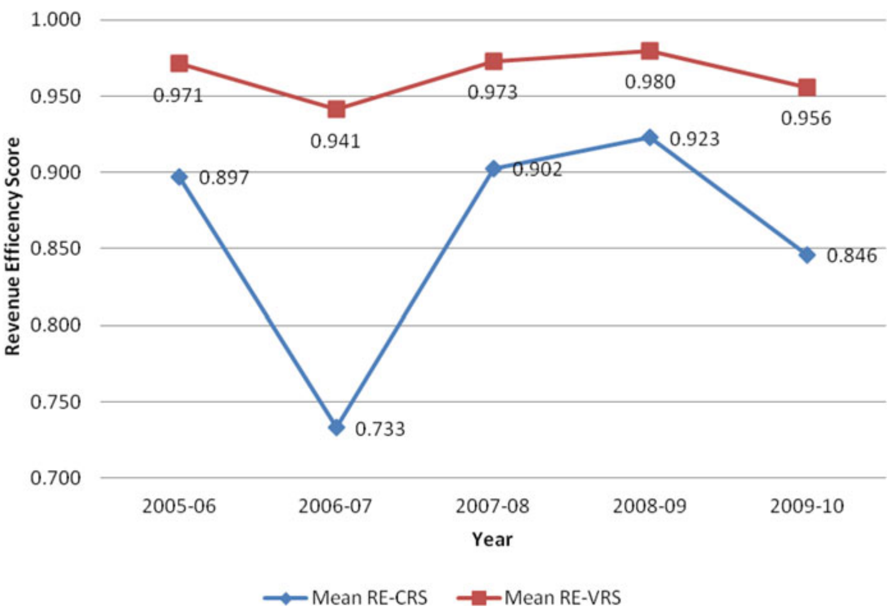


Fig. 26.2 Mean annual revenue efficiency for non-life insurers under CRS and VRS

3. In the non-life insurance segment, NATIONAL (0.694) and ROYAL SUNDARAM (0.88) are the lowest revenue efficient under CRS and VRS assumption, respectively over the entire study period.
4. IFFCO-TOKIO is the only revenue efficient non-life insurer under both CRS and VRS but NEW INDIA, ICICI LOMBARD, RELIANCE, and HDFC ERGO are revenue efficient for VRS case only.
5. From Fig. 26.2, it is clear that the variability in annual average revenue efficiency under CRS situation is higher than VRS assumption.

6 Conclusion

From the comparison of revenue efficiencies within two different scales of assumption, CRS and VRS, we found considerable changes in scores because of scale error. Under the unrestricted situation of scale, both the efficiency scores have gone up in a considerable dimension, for example, in the case of non-life insurance business, overall cost efficiency under CRS is 86 % which is increased to 96.4 % under VRS. This might be the reason for most of the private insurers in the Indian market are new. They have to bear a huge establishment cost at their initial stages of business and, therefore, scale of operation has a considerable impact over revenue efficiency but still they are competitive performers, especially in the non-life insurance business compared to public players who have been in business since 1972.

References

Berger AN, Humphrey DB (1992) Measurement and efficiency issues in commercial banking. In: Griliches Z (ed) Output measurement in the services sectors. The University of Chicago Press, Chicago

Berger AN, Cummins JD, Weiss MA, Zi H (2000) Conglomeration versus strategic focus: evidence from the insurance industry. J Financ Intermed 9(4):323–362

Brockett PL, Cooper WW, Golden LL, Rousseau JJ, Wang Y (1998) DEA evaluations of the efficiency of organizational forms and distribution systems in the US property and liability insurance industry. Int J Syst Sci 29(11):1235–1247

Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis. Springer, New York

Debreu G (1951) The coefficient of resource utilization. Econometrica 19(3):273–292

Farrell MJ (1957) The measurement of productive efficiency. J Roy Stat Soc Series A (General) 120(3): 253–281

Färe R, Grosskopf S, Lovel CAK (1985) The measurement of efficiency of production. Kluwer Nijhoff, Boston

IRDA Annual Reports of 2005–2006, 2006–2007, 2007–2008, 2008–2009, 2009–2010

Tone K (2002) A strange case of the cost and allocative efficiencies in DEA. J Oper Res Soc 53(2002):1225–1231

Chapter 27
Empirics on Fiscal Smoothing: Some Econometric Evidence for the Indian Economy

Narain Sinha and R. C. Sharma

1 Introduction

Fiscal smoothening plays an important role in macroeconomic policies. Fiscal smoothening includes both tax smoothing and revenue smoothing in an economy. A basic result in fiscal theory is that the economic costs of raising taxes are minimized by striking an average tax rate over the period. This is the "tax-smoothing hypothesis." Optimal collection of taxes has been an area of interest. Assuming that the taxes are distorting and using a dynamic optimal control problem to be solved by the government, Barro (1979) propounded the tax-smoothing hypothesis which suggests that tax rates should be smoothed over time. Employing similar framework, Mankiw (1987) derived the revenue-smoothing hypothesis as a part of positive theory of monetary and fiscal policies. The underlying basic principle is that an increase in government revenue requires the use of both fiscal and monetary policies. Several attempts have been made in the past to test the tax-smoothing hypothesis at the national and subnational levels. Using the contemporaneous single equation ordinary least squares (OLS), Mankiw (1987) and Poterba and Rotemberg (1990) find a support for the revenue-smoothing hypothesis. Employing more advanced theory of econometric analysis of time series, Trehan and Walsh (1990), Froyen and Waud (1995), Ghosh (1995), and Evans and Amey (1996) generally reject the revenue smoothing. Tax-smoothing hypothesis is generally supported for the federal governments and rejected for the state and local levels (Strazicich 1997).

In this chapter, an attempt is made to test the tax- and revenue-smoothing hypotheses for the Indian economy implying that the tax rates are martingale. The analysis

R. C. Sharma (✉)
Department of Management, Central University of Rajasthan,
Kishangarh, Rajasthan 305 801, India
e-mail: rameshsharma@curaj.ac.in

N. Sinha
Department of Economics, University of Botswana,
Gaborone, Botswana
e-mail: narainsinha@hotmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_27, © Springer India 2014

is organized as follows. Section 2 reviews the tax- and revenue-smoothing models and presents their testable theoretical propositions. Section 3 reports the empirical results and Sect. 4 concludes the chapter.

2 Review of Earlier Studies

Ashworth and Evans (1998) test the extended tax-smoothing model given by Barro (1979) for a sample of 32 developing countries. Importantly, the testable implications employed relax the assumption of constant money velocity. Although seigniorage is an important source of revenue in developing countries, all the evidence indicates that the principles of optimal taxation have not been used when developing countries raise revenue from inflation. The hypothesis that cooperation between fiscal and monetary authorities to minimize the distortionary costs of financing an exogenous stream of government expenditures implies a long-run relationship between inflation and tax rates is called the revenue-smoothing hypothesis. Using the marginal tax rate as a tax measure, Akhand and Marshall (1996) test a hierarchy of hypotheses implied by the revenue-smoothing model. Gogas and Serletis (2005) test the Mankiw revenue-smoothing hypothesis, that the inflation rate moves one-for-one with the marginal tax rate in the long run, using the new average marginal tax rate series and the long-horizon regression approach. It reports considerable evidence against revenue smoothing.

3 Theoretical Foundations

Tax-smoothing hypothesis: It is assumed that the taxes are distorting and that the government minimizes the distortion by allocating taxes across time. Assuming that the monetary policy is not used to generate revenues, Barro (1979) argues that a tax collection sequence $\{T_T\}_{t=0}^{\infty}$ is chosen by the government that minimizes the loss function given as

$$\mathrm{L} = E_t \sum \beta^j L(T_{t+j}, Y_{t+j}), \tag{27.1}$$

where Y_t is the exogenously determined real national income and T_t is the total real tax revenue at time t. If τ is the tax ratio, then

$$L = E_t \sum \beta^j L(\tau_{t+j})Y \text{ with } \mathrm{L}_1 > 0, \text{ and } \mathrm{L}_{11} > 0. \tag{27.2}$$

The objective function (27.1) is minimized with respect to $\{T_t, B_{t+1}\}_{t=0}^{\infty}$ subject to a sequence of budget constraints:

$$B_{t+1} = (1 + r)B_t + G_t - T_t, \tag{27.3}$$

where B_0 is given, $(1+r)$ is the constant gross real rate of interest, B_t is the level of real interest-bearing government debt at time t, and G_t is real net-of-interest government expenditures assumed to follow an exogenous stochastic process:

$$\text{If } \lim_{t\to\infty} \frac{B_t}{(1+r)^t} = 0. \tag{27.4}$$

On solving Eqs. (27.3) and (27.4), we get the inter-temporal budget constraint given as follows:

$$\sum_{j=0} \frac{1}{(1+r)^j} E_t T_{t+j} = B_t + \sum_{j=0}^{\infty} \frac{1}{(1+r)^j} E_t G_{t+j}. \tag{27.5}$$

Minimizing Eq. (27.1) subject to Eq. (27.5) yields the Euler equation as given below:

$$\frac{L'(T_t)}{E_t \beta L'(T_{t+1})} = 1 + r. \tag{27.6}$$

If $\beta = \frac{1}{1+r}$, then Eq. (27.6) reduces to

$$E_t L'\left(T_{t+1}\right) = L'\left(T_t\right). \tag{27.7}$$

This implies that the marginal social costs of taxation have to be equated over years. This suggests that the marginal social cost of taxation is a martingale. For a loss function which is quadratic in tax levels, the inter-temporal first-order condition (FOC) becomes

$$E_t(T_{t+1}) = T_t. \tag{27.8}$$

If the social cost function is homogeneous, then the results mentioned earlier hold for the average tax rate also. That is, $E_t(\tau_{t+1}) = \tau_t$. In the parlance of time series, this implies a unit root in the tax rates (τ_t), but it should be observed that it is a necessary condition and not sufficient condition for

$$E_t(\tau_{t+1}) = \tau_t, \tag{27.9}$$

where $T = \tau Y$. Thus, the tax-smoothing hypothesis holds for the Indian economy if the tax ratios follow the unit roots.

Revenue-smoothing hypothesis: In Barro's analysis, it is assumed that the government abstains from inflationary finance. Extending the work by Phelps (1973) and Barro (1979), Mankiw (1987) developed a positive theory of monetary and fiscal policy. Abstracting from the possibility of government borrowing, he shows that if both fiscal and monetary policies are used to optimally finance government expenditures, inflation and nominal interest rates and inflation (implicit tax) will move together over time.

Following Mankiw (1987), let Y be the exogenous level of real output and τ the tax rate on output. The revenue raised by this tax is τY. If the government is assumed to finance the expenditure in excess of taxes from seigniorage, the demand for money is described by the quantity equation $M/P = kY$, and the real revenue from seigniorage is

$$\frac{\dot{M}}{P} = \frac{\dot{M}}{M} \cdot \frac{M}{P} = \left(\frac{\dot{P}}{P} + \frac{\dot{Y}}{Y}\right) kY = (\pi + g)kY. \tag{27.10}$$

Here, π is the inflation rate and g is the growth rate of real output which is also exogenous in nature. Total real tax revenue (T) consists of two components, that is, the receipts from direct taxation (τY) and seigniorage $\{(\pi + g)\ kY\}$.

Thus,

$$T = \tau Y + (\pi + g)kY. \tag{27.11}$$

The social costs of taxation and seigniorage are assumed homogeneous in output and, hence, can be expressed as $f\ (\tau)$. Y and $h\ (\pi)$. Y, respectively, such that $f' > 0$, $h' > 0$ and $f'' > 0$, $h'' > 0$. Thus, the loss function to be minimized by the government with respect to π and τ is expressed in terms of the present value of the social losses due to taxation, and seigniorage is given by

$$E_t \sum_{j=0}^{\infty} \beta^j [f(\tau_{t+j}) + h(\pi_{t+j})k]Y \tag{27.12}$$

subject to the present value of the inter-temporal budget constraint expressed as follows:

$$\sum_{j=0}^{\infty} \beta^j EG_{t+j} + B_t = \sum_{j=0}^{\infty} \beta^j ET_{t+j}. \tag{27.13}$$

The FOCs for an optimal inter-temporal monetary and fiscal policy are obtained as follows:

$$E_t\{f'(\tau_{t+j})\} = f'(\tau_t) \tag{27.14}$$

$$E_t\{h'(\pi_{t+j})\} = h'(\pi_t) \tag{27.15}$$

$$f'(\tau_t) = kh'(\pi_t). \tag{27.16}$$

First two FOCs which are similar to those obtained in Barro suggest that marginal social cost of inflation has to be equated as do the marginal social cost of taxation (Barro 1979). If f (.) and h (.) are assumed to be quadratic, then the marginal social costs of taxation and inflation are the same as tax rate and inflation rate, respectively. Thus, the tax rate and inflation rate are martingale individually. In this sense, the tax rate and inflation rate are said to be smoothed over time. The third FOC is static in nature and equates the marginal social cost of raising revenue through taxation to that of raising revenue through seigniorage and, thus, relates the tax rate to the rate of inflation.

Another crucial implication of this result from the theory of optimal seigniorage is that any increase in government revenue requires an increase in the use of both instruments of economic policy, namely taxation and inflation. In other words, if both fiscal and monetary policies are employed to optimally finance the government expenditures, tax rates and inflation will vary together over time. This is the augmented version of Barro's tax-smoothing hypothesis and in known as revenue-smoothing hypothesis. This results in the empirical testing of the revenue-smoothing hypothesis by estimating the following model:

$$\tau_t = \alpha + \beta \pi_t. \tag{27.17}$$

This suggests that the tax and inflation rates move together.

4 Empirical Results

Using the annual time series on tax to gross domestic product (GDP) ratios for the total tax revenue (Total Tax Rev), revenue from central taxes (Cen Tax), states' share in central taxes (States Share Cen Tax), states' own tax revenue (States Own Tax), and state tax revenue (State Tax Rev), the revenue-smoothing hypothesis has been tested separately for direct taxes (indicated by the suffix d) and indirect taxes (indicated by the suffix ind) for the period 1970–1971 through 2000–2001. Inflation has been measured here in terms of consumer price index for urban nonmanual workers. Data on tax ratios have been taken from Government of India (2001). Both Augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) unit root tests have been employed for testing the tax-smoothing hypothesis employing the model with intercept, and the results are presented in Table 27.1 (Figure 27.1).

The results indicate that the ADF statistical value in each case is smaller in absolute terms than the critical value and, hence, we do not reject the null hypothesis of unit root and, thus, conclude that the tax ratios are stationary (Asteriou and Hall 2007).

None of the series in out of the ten tax rates considered here and inflation rate is stationary, but these series are I (1) if the model in difference with intercept is employed for the testing. Hence, the tax rates in Indian economy are *martingale* and, hence, the tax-smoothing hypothesis holds. Our results are robust for the methods employed for testing namely ADF and PP tests. The CPI_{unm} also has unit root, but the inflation rate measured as logarithm of CPI_{unm} is I (0). These data are given in the Appendix.

In this chapter, we have employed the single equation OLS method for testing the revenue-smoothing hypothesis at the macrolevel for various components of tax revenue in Indian economy. The inflation has been measured in terms of the change in (that is, log of) the consumer price index for the urban nonmanual workers (CPI_{unm}) and data have been taken from Reserve Bank of India (2000). Since the tax rates are martingale and are integrated of order one (i.e., I (1)), we applied cointegration analysis. The consumer price index is found to be I (2), but the inflation measured as logarithm of CPI is I (1). These results suggest that the revenue-smoothing hypothesis

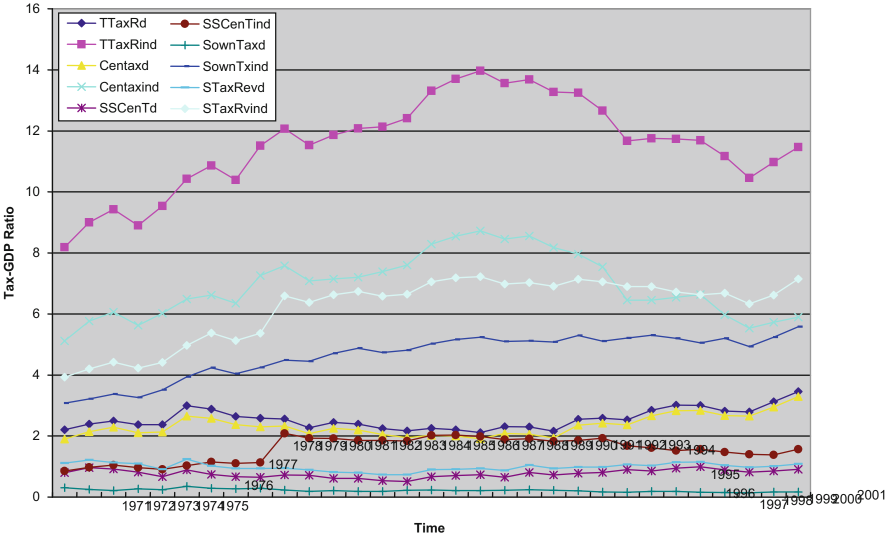


Fig. 27.1 Tax GDP Ratio over time

Table 27.1 H_0: Tax smoothing, i.e., presence of unit root

S.No.	Dependent variable	Test statistic	
		ADF	PP
1	Tot Tax Rev_d	– 0.75	– 0.82
2	Tot Tax Rev_{ind}	– 2.43	– 2.40
3	Cen Tax_d	– 0.53	– 0.53
4	Cen Tax_{ind}	– 1.65	– 1.74
5	States Share Cen Tax_d	– 2.03	– 1.88
6	States Share Cen Tax_{ind}	– 2.01	– 2.03
7	States Own Tax_d	– 0.80	– 2.51
8	State Own Tax_{ind}	– 1.72	– 2.76
9	State Tax Rev_d	– 2.59	– 2.59
10	State Tax Rev_{ind}	– 2.13	– 2.27
11	CPI_{unm}	– 0.20	– 1.82
12	$lnCPI_{unm}$	– 4.32[a]	– 4.46[a]
	1 % Critical value[a]	**– 3.6752**	**– 3.6661**
	5 % Critical value	**– 2.9665**	**– 2.9627**
	10 % Critical value	**– 2.6220**	**– 2.6200**

[a]MacKinnon critical values for rejection of hypothesis of a unit root

holds in all but two cases. These two cases, where the revenue-smoothing hypothesis is not supported, are states' share in central taxes from direct taxes and states' tax revenue from direct taxes in the Indian economy during the period 1970–1971 through 2000–2001. The cointegration analysis has supported this finding. Results from the cointegration are given in the Appendix.

Table 27.2 Revenue-smoothing hypothesis

S.No.	Dependent variable	Intercept	π	R^2	H_0: No cointegration
1	Tot Tax Rev_d	1.199905	0.212007*	0.231951	Do not reject
		0.01414	0.006084		
2	Tot Tax Rev_{ind}	5.336848	0.97833*	0.238379	Reject
		0.015865	0.005327		
3	Cen Tax_d	0.667118	0.261657*	0.326874	Do not reject
		0.146617	0.00078		
4	Cen Tax_{ind}	6.014781	0.143497*	0.011232	Reject
		0.000784	0.000784		
5	States Share Cen Tax_d	0.457115	0.049008	0.087419	Reject
		0.021839	0.106335		
6	States Share Cen Tax_{ind}	0.040107	0.23805*	0.205622	Reject
		0.943127	0.010408		
7	States Own Tax_d	0.537965	– 0.05031*	0.598888	Reject
		7.78×10^{-12}	3.29×10^{-07}		
8	State Own Tax_{ind}	– 0.67941	0.834811*	0.774191	Reject
		0.215912	7.04×10^{-11}		
9	State Tax Rev_d	0.990471	– 0.00068	1.65×10^{-05}	Do not reject
		2.82×10^{-05}	0.982677		
10	State Tax Rev_{ind}	– 0.64769	1.074279*	0.596437	Reject
		0.543142	3.6×10^{-07}		

Values below the coefficient are the *p*-values
* indicates significance

The empirical results are shown in Table 27.2. Our results strongly support revenue-smoothing hypothesis for the tax revenues in the Indian economy except in the case of state tax revenue from direct taxes. The series considered are stationary and the cointegration is rejected only in the case of the direct taxes. In most of the recent studies, the advances in the theory of nonstationary regressors and integration and cointegration analysis of the series have been employed. Since a meaningful testing of the hypothesis critically depends on such properties, the revenue-smoothing hypothesis cannot be rejected except in three cases, namely total tax revenue from direct taxes, central tax revenue from direct taxes, and state tax revenue from direct taxes. For the purpose of cointegration,[1] we used the model with intercept without the linear trend. The null hypothesis of no cointegration was tested on the basis of the trace statistics at 5 % level of significance (Asteriou and Hall 2007).

The use of cointegration tests and unit root tests based on the single equation approach in which the tax rate is treated as predetermined has been employed here. In most of the earlier studies using these techniques, the revenue-smoothing hypothesis has been rejected, because the time series properties are imposed on the data. However, in the present study, revenue smoothing is rejected only in the case of state revenue from direct taxes. In all other cases, the cointegration has supported the revenue-smoothing hypothesis in India.

[1] The results from cointegration analysis can be obtained from the corresponding author.

5 Conclusion

The revenue-smoothing hypothesis has been empirically tested for various components of the tax revenue in Indian economy using the annual data for the period 1970–1971 to 2000–2001. As in Mankiw (1987) and Poterba and Rotemberg (1990), the empirical evidence for the Indian economy has supported the tax-smoothing hypothesis for most of the components of the tax revenue. The revenue smoothing is not supported for state tax revenue from direct taxes. Comparison with these two studies is tenable because in the present chapter, we have used single equation OLS method of estimation. Unlike most of the earlier studies based on value at risk (VAR) and cointegration techniques rejecting this hypothesis, in the present study the fiscal-smoothing hypotheses have been found to hold true in the case of India.

Appendix

Year	$TTaxR_d$	$TTaxR_{ind}$	$Centax_d$	$Centax_{ind}$	$SSCenT_d$	$SSCenT_{ind}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1971	2.21	8.19	1.90	5.12	0.80	0.85
1972	2.39	9.00	2.14	5.77	0.96	0.97
1973	2.49	9.43	2.29	6.06	0.92	1.05
1974	2.37	8.90	2.10	5.63	0.82	0.96
1975	2.37	9.54	2.13	6.03	0.67	0.91
1976	2.99	10.43	2.65	6.49	0.89	1.03
1977	2.88	10.86	2.58	6.62	0.74	1.15
1978	2.64	10.39	2.37	6.35	0.67	1.10
1979	2.59	11.51	2.30	7.26	0.65	1.13
1980	2.56	12.07	2.33	7.58	0.72	2.09
1981	2.27	11.53	2.08	7.08	0.71	1.93
1982	2.45	11.87	2.25	7.15	0.61	1.92
1983	2.39	12.08	2.20	7.20	0.61	1.85
1984	2.24	12.13	2.05	7.39	0.54	1.85
1985	2.17	12.41	1.95	7.60	0.51	1.84
1986	2.25	13.31	2.02	8.29	0.67	2.02
1987	2.21	13.70	2.00	8.55	0.70	2.03
1988	2.11	13.97	1.91	8.72	0.73	1.98
1989	2.31	13.56	2.09	8.45	0.65	1.88
1990	2.30	13.68	2.06	8.56	0.81	1.91
1991	2.16	13.27	1.94	8.18	0.72	1.83
1992	2.55	13.25	2.35	7.96	0.78	1.85
1993	2.59	12.66	2.42	7.55	0.81	1.93
1994	2.53	11.67	2.36	6.45	0.90	1.68
1995	2.85	11.75	2.66	6.45	0.85	1.61
1996	3.01	11.74	2.83	6.54	0.95	1.52
1997	3.00	11.69	2.84	6.64	0.99	1.57
1998	2.82	11.17	2.67	5.97	0.89	1.47
1999	2.79	10.46	2.65	5.53	0.82	1.40
2000	3.12	10.98	2.95	5.73	0.85	1.38
2001	3.46	11.47	3.29	5.90	0.91	1.57
Mean	2.5002	9.326	2.14	5.671	0.669	1.108
Variance	0.0953	10.12	0.16	3.516	0.027	0.439

Year	$SownTax_d$	$SownTx_{ind}$	$STaxRev_d$	$STaxRv_{ind}$	CPI_{unm}
(1)	(8)	(9)	(10)	(11)	(12)
1971	0.31	3.08	1.11	3.93	174
1972	0.25	3.22	1.21	4.2	180
1973	0.21	3.37	1.13	4.42	192
1974	0.27	3.26	1.09	4.23	221
1975	0.24	3.51	0.91	4.41	270
1976	0.35	3.94	1.24	4.97	277
1977	0.29	4.24	1.02	5.38	277
1978	0.27	4.04	0.94	5.13	296
1979	0.29	4.25	0.94	5.37	306
1980	0.23	4.49	0.95	6.59	330
1981	0.19	4.45	0.89	6.38	369
1982	0.21	4.71	0.82	6.63	413
1983	0.19	4.88	0.80	6.74	446
1984	0.19	4.74	0.73	6.58	492
1985	0.22	4.81	0.73	6.65	535
1986	0.23	5.02	0.90	7.05	572
1987	0.21	5.16	0.91	7.18	615
1988	0.21	5.24	0.94	7.22	674
1989	0.22	5.1	0.87	6.98	727
1990	0.24	5.12	1.05	7.03	775
1991	0.22	5.08	0.94	6.91	861
1992	0.2	5.29	0.98	7.14	979
1993	0.17	5.11	0.98	7.05	1,080
1994	0.16	5.21	1.07	6.90	1,155
1995	0.19	5.30	1.03	6.90	1,267
1996	0.19	5.20	1.14	6.72	1,385
1997	0.16	5.05	1.15	6.63	1,513
1998	0.15	5.20	1.04	6.68	1,615
1999	0.14	4.93	0.97	6.33	1,802
2000	0.17	5.24	1.01	6.62	1,882
2001	0.17	5.58	1.08	7.15	1,984
Mean	0.364	3.654	1.033	4.7584	763.43
Variance	0.044	1.921	0.022	4.1129	309,102

References

Ashworth J, Lynne E (1998) Seigniorage and tax smoothing in developing countries. J Econ Studies 25(6):486–495

Asteriou D, Hall SG (2007) Applied econometrics: a modern approach. Macmillan

Akhand HA, Marshall JB (1996) Inter-temporal revenue-smoothing in the postwar United States and Canada. Working papers—Department of Economics, Regina

Barro RJ (1979) On the determination of the public debts. J Polit Econ 87:940–971

Evans JL, Amey MC (1996) Seigniorage and tax smoothing: testing the extended tax-smoothing model. J Macroecon 18:111–125

Froyen RT, Waud RN (1995) Optimal seigniorage versus interest rate smoothing. J Macroecon 17:111–129

Ghosh AR (1995) Inter-temporal tax smoothing and the government budget surplus: Canada and the United States. J Money Credit Bank 27:1033–1045

Gogas P, Serletis A (2005) Revenue smoothing in an ARIMA framework: evidence from the United States. In: Darné O, Diebolt C, Kyrtsou C (eds) New Trends in Macroeconomics. Springer

Government of India (2001) Indian public finance statistics, 2000–2001. Ministry of Finance, Department of Economic Affairs, Economic Division, New Delhi

Mankiw NG (1987) The optimal collection of seigniorage: theory and evidence. J Monet Econ 200:327–341

Phelps ES (1973) Inflation in the theory of public finance. Swed J Econ 75:67–82

Poterba JM, Rotemberg JJ (1990) Inflation and taxation with optimizing governments. J Money Credit Bank 22:1–18

Reserve Bank of India (2000) Handbook of statistics on the Indian economy. Reserve Bank of India, Mumbai

Strazicich MC (1997) Does tax smoothing differ by the level of government? Time series evidence from Canada and the United States'. J Macroecon 19:305–326

Trehan B, Walsh CE (1990) Seigniorage and tax smoothing in the United States: 1914–1986. J Monet Econ 25:97–112

Chapter 28
Index of Financial Inclusion: Some Empirical Results

Ram Pratap Sinha

1 Introduction

Financial inclusion may be defined as the process of ensuring access to financial services (including savings, loans, insurance, payments, remittance facilities, etc.) and the provision of adequate credit to the socially vulnerable groups at affordable rates. Financial inclusion can therefore be considered as an important prerequisite for just and equitable growth of a modern society. The (Rangarajan) Committee on Financial Inclusion (2008) thus pointed out that: 'Access to finance, especially by the poor and vulnerable groups is a prerequisite for employment, economic growth, poverty reduction and social cohesion. Further, access to finance will empower the vulnerable groups by giving them an opportunity to have a bank account, to save and invest, to insure their homes or to partake of credit, thereby facilitating them to break the chain of poverty'.

In the Indian context, the process of financial inclusion in the banking sector was initiated by the public sector banks since the end of the 1960s through the process of branch banking which provided the rural people with the opportunity to have bank accounts in their names even with small amounts of deposits. The competition scenario in the commercial banking sector changed consequently with the entry of new private sector banks in the market and in recent times some of the leading private sector banks are making their presence felt in the semi-urban areas. This is not withstanding the fact that the distribution of financial services in India is highly skewed. Therefore, the unevenness in the provision of financial services in India constitutes an important research agenda.

R. P. Sinha (✉)
Government College of Engineering and Leather Technology, Kolkata 700098, India
e-mail: rampratapsinha39@gmail.com, rp1153@rediffmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_28,

2 Financial Inclusion in the Banking and Insurance Sector

The importance of financial inclusion in the context of Indian banking sector has long been understood by the government and the market regulator (the Reserve Bank of India, RBI) which took a proactive role in the growth of mass banking in the rural sector with the help of the public sector commercial banks and the entities like the regional rural banks and cooperative credit societies. The governmental policy tried to ensure the expansion of the base of institutional credit in the rural sector and also aimed at providing directed credit to the disadvantaged borrowers at concessional rates of interest. However, with the passage of time, the delivery system/mechanism underwent important changes. In this context, Fisher and Shriram (2006) identified three distinct phases of credit delivery: In the first phase (1950–1965), cooperatives were the institutional vehicles of choice; in the second phase (1970–1989), attention shifted to commercial banks and regional rural banks; and in the third phase (the reform period since the early 1990s which involved the restructuring of the banking system), the emergence of self-help groups (SHGs) and a growing number of micro-finance institutions.

In November 2005, banks were advised by the RBI to make available a basic banking 'no-frills' account with low or nil minimum stipulated balances as well as charges to expand the outreach of such accounts to vast sections of the population. By 31 March 2006, about five lakh no-frill accounts have been opened, of which about two-thirds are with the public sector and one-third with the private sector banks. Further, in January 2006, banks were permitted to utilise the services of non-governmental organisations (NGOs/SHGs), micro-finance institutions, etc. as intermediaries in providing financial and banking services through the use of business facilitator and business correspondent models.

Prior to the formation of Insurance Regulatory and Development Authority (IRDA), there has not been any conscious attempt to develop the rural insurance market in India. The public sector commercial banks bundled insurance with directed lending to the rural poor primarily with the objective of hedging their own risk. However, the IRDA introduced social and rural sector obligations for the insurance companies in 2002. The rural sector was defined by the IRDA as comprising of (a) a population $< 5{,}000$, (b) density of population $< 400/\mathrm{km}^2$, and (c) more than 25 % of the male working population is engaged in agricultural activity (cultivation, agricultural labour, livestock farming, forestry, fishing, hunting and plantations, orchards, and allied activities). The social sector is defined to comprise of (1) unorganized sector, (2) informal sector, (3) economically vulnerable or backward classes, and (4) other categories of persons, in both rural and urban areas. The IRDA targets corresponding to the rural/social sector for the life insurance companies are provided in Table 28.1.

Table 28.1 IRDA rural/social sector obligation guideline for the life insurance companies (2002). (Source: Adapted from UNDP Report 2007)

Particulars	Year 1	Year 2	Year 3	Year 4	Year 5
Per cent policies (rural sector) written in the year	7	9	12	14	16
Number of persons covered (social sector)	5,000	7,500	10,000	15,000	20,000

Table 28.2 The rural–urban gap in the provision/access of banking services (as on 31 March 2005). (Source: Mohan 2006)

Particulars	Offices (% to total)	No. of deposit accounts (% of number of households)	Deposits (% to total)	No. of credit accounts (% of number of households)	Credits (% to total)
Rural	68	181.8	29.1	32.2	20.8
Urban	32	335.4	70.9	50.2	79.2
Total	100	517.2	100	82.4	100

Table 28.3 Regional variation in banking penetration. (Source: RBI 2008: Statistical Tables Relating to Banks in India, 2007–2008, www.rbi.org.in.)

Region	Deposit mobilisation per office (Rs. crore)	Credit disbursement per office (Rs. crore)
Western	88.46	78.59
Northern	58.11	38.28
Southern	33.8	30.19
Eastern	28.02	14.35
Central	24.57	11.59
Northeastern	24.8	9.68

3 Disparity in Financial Inclusion: The Indian Scenario

In spite of the government and regulatory initiatives to have balanced and inclusive growth in India, substantial disparity does exist in the provision of financial services. The disparity can exist either as a rural–urban gap or in terms of inter-region differences. To provide a snapshot picture of the disparity, we present three tables here. Table 28.2 provides a very brief account of the rural–urban disparity in the provision of deposits and credit. Tables 28.3 and 28.4 provide inputs on the state-wise distribution of bank and insurance offices and the mobilisation of deposits and premia as well as credit delivery as on 31 March 2008.

4 Index of Financial Inclusion: The Received Literature

The earliest attempt towards the construction of an index of financial inclusion (IFI) was by Sarma (2008). She pointed out that while there is a widespread recognition of the importance of financial inclusion, one does not find any comprehensive measure

Table 28.4 Regional variation in insurance penetration. (Source: IRDA 2008: Annual Report, 2007–2008, www.irdaindia.org)

Region	New policies mobilised per office	Insurance premia mobilised per office (Rs. crore)
Western	5,492.67	9.24
Northern	4,031.35	8.91
Southern	5,683.17	9.19
Eastern	8,138.09	9.37
Central	5,994.21	7.93
Northeastern	5,594.61	7.46

of financial inclusion in the literature which may be used to measure the extent of financial inclusion across economies. In her paper she presented an approach for the construction of an IFI and used the same to compare 100 countries in respect of financial inclusion. Sarma (2008) proposed a multidimensional approach for the construction of an IFI. The approach involves the computation of a dimension index for each dimension of financial inclusion. The dimension index for the *i*th dimension, d_i, is computed by the following formula:

$$d_i = (A_i - m_i)/(M_i - m_i), \tag{28.1}$$

where

A_i Actual value of dimension *i*,
m_i Minimum value of dimension *i*, and
M_i Maximum value of dimension *i*.

Formula (28.1) ensures that $0 \leq d_i \leq 1$. Higher the value of d_i, higher the country's achievement in dimension *i*. If *n* dimensions of financial inclusion are considered, then a country *i* will be represented by a point $Di = (d1, d2, d3, \ldots .dn)$ on the *n* dimensional Cartesian space.

Then the IFI for entity *i* is computed as:

$$IFI_i = 1 - \sqrt{\left[(1-d_1)^2 + (1-d_2)^2 + ------ + (1-d_n)^2\right]}/\sqrt{n}. \tag{28.2}$$

Sarma made use of the IFI to make a cross-country comparison of financial inclusion for the year 2004 using three dimensions of financial inclusion: banking penetration (BP) as measured by bank accounts/population, availability of the banking services (BS) as measured by the number of bank branches per 1,000 population and usage of the banking system (BU) as measured by the volume of deposit and credit to gross domestic product (GDP).

Goyal (2009) used the multidimensional IFI to compare the state of financial inclusion of the northeastern states for the year 2005 using the same dimensions used by Sarma. However, the index computed by Goyal used per capita credit and deposit instead of the volume of credit and deposit as proportion of the country's GDP used by Sarma.

Kuri and Laha (2011) compared 31 Indian states and union territories (UTs) in respect of IFI in terms of Euclidean distance function approach using the dimension indices suggested by Sarma. They also computed Human Development Indices (HDIs) for the same states and UTs by considering three dimensions: average monthly per capita expenditure, longevity index and a composite indicator of adult literacy rate, and school attendance rate. The study also indicated that most of the components of IFI and HDI exhibited high correlation.

5 Comparison of State Performance in Financial Inclusion

This chapter makes use of two approaches for measuring state performance in financial inclusion: Euclidean and Shepherd's distance function. Before describing the two approaches, we provide a conceptual framework for the construction of the index.

Let us consider a state which provides financial service infrastructure to its population. There are several indicators of financial service infrastructure denoted by $I(I_1, I_2, \ldots., I_m)$. The indicators depend on the population level P. Thus, in functional form we may relate them as

$$I(I_1, I_2, \ldots., I_m) = f(P). \tag{28.3}$$

The technology used by the state is defined by the production possibility set (PPS).

$$\text{PPS} = \{(P, I) : I \text{ can be produced from } P\}$$

An input–output combination (P^0, I^0) is feasible if and only if $(P^0, I^0) \in$ PPS.

We assume that the observed states are output maximisers, i.e. they try to produce the maximum possible level of output given the level of income.

5.1 Construction of IFI: The Euclidean Distance Function Approach

The Euclidean distance between points m and n is the length of the line segment connecting them. In the n-dimensional case, $d(p.q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + - - - - + (q_n - p_n)^2} = \sqrt{\sum (q_i - p_i)^2}$.

In the present case, if we have considered m indicators of financial inclusion the performance of a state is measured by

$$IFI = 1 - \sqrt{[(1-d_1)^2 + (1-d_2)^2 + - - + (1-d_m)^2]}/\sqrt{m}. \tag{28.4}$$

5.2 Construction of IFI: Output Distance Function Approach

In the context of multi-criteria portfolio evaluation, Shephard's (1953, 1970) distance function approach provides a sound conceptual basis for the derivation of evaluation criteria. The idea is invoked from a multi-input–multi-output production system where distance function provides a functional characterisation of the structure of production technology. The input set of the production technology is characterised by the input distance function while the output set is characterised by the output distance function.

In order to elaborate the concept of distance function, we consider a technology T using a non-negative vector of inputs $X = (x_1, x_2, \ldots\ldots, x_n) \in R^n{}_+$ to produce a non-negative vector of outputs $Y = (y_1, y_2, \ldots\ldots, y_m) \in R^m{}_+$. In functional terms, they can be related as: $Y = P(X)$ and $X = L(Y)$.

Given this, an input distance function can be defined as $D_{\text{input}} = \text{Max}[\lambda : X/\lambda \in L(Y)]$. Intuitively speaking, an input distance function gives the maximum amount by which the producer's input vector can be radially contracted and yet remain feasible for the output vector it produces. The reciprocal of the input distance function can be considered as the radial measure of input-oriented technical efficiency.

In an analogous fashion, the output distance function is defined as: $D_{\text{output}} = \text{Min}[\mu : Y/\mu \in P(X)]$. Intuitively speaking, an output distance function gives the minimum amount by which the producer's output vector can be deflated and yet remain feasible for a given input vector.

In the present context, the problem for a state is:

$$Max\phi \tag{28.5}$$

$$s.t.\phi I^0 \leq \lambda I, P^0 \geq \lambda P \tag{28.6}$$

$$\lambda j \geq 0, \sum \lambda = 1, \tag{28.7}$$

where (P^0, I^0) refer to the observed input–output bundle of the state. $\sum \lambda_j = 1$ implies that the state is operating under variable returns to scale.

The IFI computed from the optimization exercise is equal to $1/\phi$

5.3 Selection of Input and Output

In both the approaches, three output indicators (deposit mobilised, credit disbursed, and insurance premium mobilised) have been used and the state-wise population has been used as the input indicator. The output–input relationship used in this model is thus:

$$\text{Output (Deposit Mobilised, Credit Disbursed, Insurance Premium Collected)}$$
$$= \text{f (State Population)}.$$

The projected population data used for the study have been taken from census data. The remaining data have been obtained from the Statistical Tables Relating to Banks in India and the IRDA Annual Reports.

5.4 Results

(a) Descriptive Statistics

Table 28.5 provides the information about the descriptive statistics of the indices computed.

Table 28.5 Descriptive statistics of index of financial inclusion. (Source: Calculated)

Particulars	Euclidean distance function	Output distance function
No. of states	35	35
Mean index of financial inclusion	0.0850	0.4275
Standard deviation	0.1129	0.3109
Max.	0.0003	1
Min.	0.5120	0.0561

(b) Returns to Scale

Table 28.6 provides the summary information about the returns to scale of the Indian states/UTs under the output distance function approach.

Table 28.6 Returns to scale exhibited by Indian states. (Source: Calculated)

Returns to scale	Efficient	Projected	Total
No. of states/UTs exhibiting IRS	0	0	0
No. of states/UTs exhibiting CRS	2	7	9
No. of states/UTs exhibiting DRS	3	23	26
Total	5	30	35

IRS increasing returns to scale, *CRS* constant returns to scale, *DRS* decreasing returns to scale

6 Conclusion

In this chapter, two models (Euclidean distance function and output distance function approach) have been applied to compare interstate performance in financial inclusion (for 2005–2006). The regional variations have been captured in Table 28.7 and the state/UT-wise indices, rankings, and returns to scale (for the second model) are presented in appendix tables A1–A9. Kindly note that unlike Sarma (2008), Goyal (2009), and Kuri and Laha (2011), the present study does not consider the income level of the state for the computation of the indices. Consequently, the regional variable which is apparent in the present study is partially due to income variations across the states and UTs (Fig. 28.1).

Table 28.7 Index of financial inclusion: regional variations. (Source: Calculated)

Region	Euclidean distance function (EDF)	Rank (EDF)	Output distance function (ODF)	Rank (ODF)
Northern	0.1781	1	0.5787	2
Western	0.1324	2	0.5729	3
Southern	0.1038	3	0.7296	1
Central	0.0264	5	0.3323	4
Eastern	0.0326	4	0.2584	5
Northeastern	0.0205	6	0.1128	6
All states	0.0850	–	0.4275	–

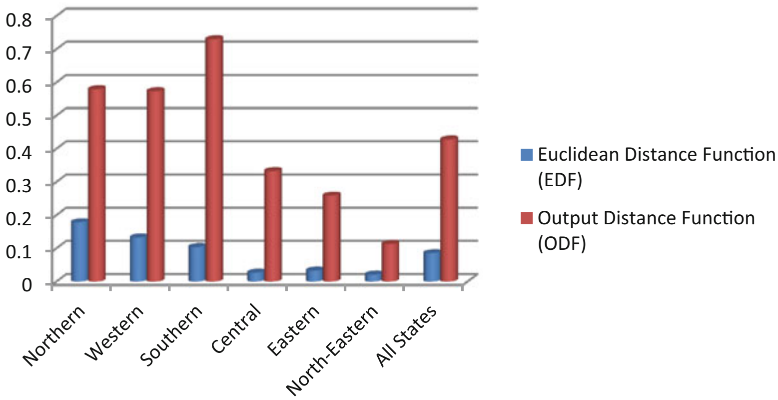


Fig. 28.1 Index of financial inclusion: Euclidean and output distance functions

Appendix

Table A.1 Indices of financial inclusion. (Source: Calculated)

State/UTs	Euclidean distance function	Output distance function
Andaman and Nicobar	0.0499	0.1386
Andhra Pradesh	0.0512	0.5948
Arunachal Pradesh	0.0522	0.1518
Assam	0.0100	0.2022
Bihar	0.0003	0.2203
Chandigarh	0.5120	1
Chhattisgarh	0.0171	0.1473
Dadra and Nagar Haveli	0.0790	0.3416
Daman and Diu	0.0734	0.3094
Delhi	0.4480	1
Goa	0.2472	0.6527
Gujarat	0.0580	0.5610
Haryana	0.0652	0.3393
Himachal Pradesh	0.0632	0.5019

Table A.1 (continued)

State/UTs	Euclidean distance function	Output distance function
Jammu and Kashmir	0.0455	0.2173
Jharkhand	0.0192	0.2253
Karnataka	0.0886	0.6324
Kerala	0.0868	1
Lakshadweep	0.2171	1
Madhya Pradesh	0.0175	0.3025
Maharashtra	0.2042	1
Manipur	0.0048	0.0877
Meghalaya	0.0262	0.0770
Mizoram	0.0229	0.0679
Nagaland	0.0100	0.0561
Orissa	0.0192	0.2849
Pondicherry	0.0857	0.1772
Punjab	0.0921	0.5857
Rajasthan	0.0206	0.407
Sikkim	0.0650	0.1445
Tamil Nadu	0.0931	0.9729
Tripura	0.0171	0.1467
Uttar Pradesh	0.0129	0.6088
Uttarakhand	0.0581	0.2706
West Bengal	0.0418	0.5366

Table A.2 Ranking of states based on indices of financial inclusion. (Source: Calculated)

State	Euclidean distance function	Output distance function
Andaman and Nicobar	20	31
Andhra Pradesh	19	10
Arunachal Pradesh	18	27
Assam	33	25
Bihar	35	23
Chandigarh	1	1
Chhattisgarh	29	28
Dadra and Nagar Haveli	11	16
Daman and Diu	12	18
Delhi	2	1
Goa	3	7
Gujarat	17	12
Haryana	13	17
Himachal Pradesh	15	14
Jammu and Kashmir	21	24
Jharkhand	26	22
Karnataka	8	8
Kerala	9	1
Lakshadweep	4	1
Madhya Pradesh	28	19
Maharashtra	5	1
Manipur	34	32
Meghalaya	23	33

Table A.2 (continued)

State	Euclidean distance function	Output distance function
Mizoram	24	34
Nagaland	32	35
Orissa	27	20
Pondicherry	10	26
Punjab	7	11
Rajasthan	25	15
Sikkim	14	30
Tamil Nadu	6	6
Tripura	30	29
Uttar Pradesh	31	9
Uttarakhand	16	21
West Bengal	22	13

Table A.3 Returns to scale exhibited by the Indian states. (Source: Calculated)

States	Index of financial inclusion (output distance function)	Returns to scale
Andaman and Nicobar	0.1386	Constant
Andhra Pradesh	0.5948	Decreasing
Arunachal Pradesh	0.1518	Constant
Assam	0.2022	Decreasing
Bihar	0.2203	Decreasing
Chandigarh	1	Constant
Chhattisgarh	0.1473	Decreasing
Dadra and Nagar Haveli	0.3416	Constant
Daman and Diu	0.3094	Constant
Delhi	1	Decreasing
Goa	0.6527	Decreasing
Gujarat	0.5610	Decreasing
Haryana	0.3393	Decreasing
Himachal Pradesh	0.5019	Decreasing
Jammu and Kashmir	0.2173	Decreasing
Jharkhand	0.2253	Decreasing
Karnataka	0.6324	Decreasing
Kerala	1	Decreasing
Lakshadweep	1	Constant
Madhya Pradesh	0.3025	Decreasing
Maharashtra	1	Decreasing
Manipur	0.0877	Decreasing
Meghalaya	0.0770	Decreasing
Mizoram	0.0679	Constant
Nagaland	0.0561	Decreasing
Orissa	0.2849	Decreasing
Pondicherry	0.1772	Constant
Punjab	0.5857	Decreasing
Rajasthan	0.407	Decreasing
Sikkim	0.1445	Constant
Tamil Nadu	0.9729	Decreasing
Tripura	0.1467	Decreasing
Uttar Pradesh	0.6088	Decreasing
Uttarakhand	0.2706	Decreasing
West Bengal	0.5366	Decreasing

Table A.4 Index of financial inclusion of the northern states and UTs

State/UT	Euclidean distance function	Output distance function
Chandigarh	0.512	1
Delhi	0.448	1
Haryana	0.0652	0.3393
Himachal Pradesh	0.0632	0.5019
Jammu and Kashmir	0.0455	0.2173
Punjab	0.0921	0.5857
Rajasthan	0.0206	0.407
Mean	0.1781	0.5787

Table A.5 Index of financial inclusion of the western states and UTs

State/UT	Euclidean distance function	Output distance function
Dadra and Nagar Haveli	0.079	0.3416
Daman and Diu	0.0734	0.3094
Goa	0.2472	0.6527
Gujarat	0.058	0.561
Maharashtra	0.2042	1
Mean	0.1324	0.5729

Table A.6 Index of financial inclusion of the southern states and UTs

State/UT	Euclidean distance function	Output distance function
Andhra Pradesh	0.0512	0.5948
Karnataka	0.0886	0.6324
Kerala	0.0868	1
Pondicherry	0.0857	0.1772
Tamil Nadu	0.0931	0.9729
Lakshadweep	0.2171	1
Mean	0.1038	0.7296

Table A.7 Index of financial inclusion of the central-region states. (Source: Calculated)

State/UT	Euclidean distance function	Output distance function
Uttar Pradesh	0.0129	0.6088
Madhya Pradesh	0.0175	0.3025
Chhattisgarh	0.0171	0.1473
Uttarakhand	0.0581	0.2706
Mean	0.0264	0.3323

Table A.8 Index of financial inclusion of the eastern states. (Source: Calculated)

State/UT	Euclidean distance function	Output distance function
Bihar	0.0003	0.2203
Jharkhand	0.0192	0.2253
Orissa	0.0192	0.2849
West Bengal	0.0418	0.5366
Sikkim	0.065	0.1445
Andaman and Nicobar	0.0499	0.1386
Mean	0.0326	0.2584

Table A.9 Financial inclusion efficiency of the northeastern states. (Source: Calculated)

State/UT	Euclidean distance function	Output distance function
Arunachal Pradesh	0.0522	0.1518
Assam	0.01	0.2022
Manipur	0.0048	0.0877
Meghalaya	0.0262	0.077
Mizoram	0.0229	0.0679
Nagaland	0.01	0.0561
Tripura	0.0171	0.1467
Mean	0.0205	0.1128

References

Fisher T, Sriram MS (2006) Beyond micro-credit: putting development back into micro-finance. Vistaar, New Delhi

Goyal C (2009) Measuring financial inclusion—a study of North East India. Paper presented at the 45th Annual Conference of the Indian Econometric Society (Guwahati)

IRDA (2008) Annual Reports, 2007–2008. www.irdaindia.org

Kuri PK, Laha A (2011) Financial inclusion and human development in India: an inter-state analysis. Indian J Hum Dev 5(1):61–73

Mohan R (2006) Economic growth, financial deepening and financial inclusion. Annual Banker's Conference 2006 at Hyderabad

RBI (2008) Report of the Committee on Financial Inclusion. RBI, Mumbai

Sarma M (2008) Index of Financial Inclusion. ICRIER Working Paper No 215

Shephard RW (1953) Cost and productions. Princeton University Press, Princeton

Shephard RW (1970) Theory of cost and productions. Princeton University Press, Princeton

UNDP (2007) Building capacity for the poor-potential and prospect for micro insurance in India. UNDP Regional Centre, Colombo

Chapter 29
The Causal Linkage Between FDI and Current Account Balance in India: An Econometric Study in the Presence of Endogenous Structural Breaks

Jaydeep Mukherjee, Debashis Chakraborty and Tanaya Sinha

1 Introduction

The deepening wave of globalization over the last decade has considerably influenced cross-border foreign direct investment (FDI) flows, which stood at $ 1.52 trillion in 2011 (UNCTAD 2012). The direction of FDI has undergone a transformation over the last two decades. While during the twentieth century, FDI flows from the developed countries both to their developed and developing counterparts has been the norm, nowadays investment flows across developing countries is being witnessed more frequently. This recent phenomena can be explained by the prevailing higher growth rate in these economies, especially in the post-2009 recession period. The growing attraction of the developing and transition economies is reflected from the fact that for the first time in 2010 these countries jointly accounted for more than 50 % of global FDI flows (UNCTAD 2011).

The existence of the growing North–South and South–South FDI flows has motivated considerable research interest in the interrelationship between FDI and economic growth. However, the empirical evidence does not indicate an unambiguous relationship between the two series. One branch of the empirical literature reveals a positive relationship between FDI and economic growth (Tang et al. 2008; Katircioglu 2009; Ayanwale 2007). On the other hand, another segment of the literature suggests that FDI inflow may not significantly influence the economic growth of the recipient country (Braunstein and Epstein 2002; Katerina et al. 2004). The stark contrast between the findings of the two schools could be explained by a third branch

J. Mukherjee (✉) · D. Chakraborty
Indian Institute of Foreign Trade (IIFT), New Delhi, India
e-mail: jaydeepm74@gmail.com

D. Chakraborty
e-mail: debchakra@gmail.com

T. Sinha
Amity University, New Delhi, India
e-mail: tanaya82@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_29, © Springer India 2014

of literature, which indicates that to benefit from FDI inflow a country must reach a minimum threshold level in terms of efficiency, human capital stock and factor endowment (Alfaro et al. 2006; Borensztein et al. 1998).

In addition to the growth consequences, the FDI inflows can also play a crucial role in determining external balance stability. FDI is a major component of capital account of the balance of payments (BOPs), and it may, in the short run, compensate for an existing current account deficit caused by import of consumption goods or capital goods (Krkoska 2001; Yalta 2011). However, in the long run, the FDI repercussions on current account balance (CAB) might occur through several channels. First, FDI inflow generally boosts exports through gross capital formation, transfer of technology, enhanced productivity and competitiveness, introduction of newer production methods and products, better managerial techniques, greater access to new markets, etc. (Borensztein et al. 1998; Dunning and Rugman 1985; Krkoska 2001; UNCTAD 2002), which improves the CAB. Second, the foreign firms entering the recipient country may decide to import key inputs from their established global suppliers or pay royalties to the parent corporation for technical know-how, leading to an increase in imports (Onwuka and Zoral 2009; Williams and Williams 1998). As a result, the CAB is likely to worsen. Finally, the profit repatriation of foreign investors appears in the current account of BOP and greater outflow on this front also worsens the CAB (Yalta 2011). The overall impact of FDI inflow on the CAB of a particular country is, therefore, a function of the relative strengths of these three effects.

After its independence in 1947, India adopted the policy of import substitution-led growth model for securing economic growth. The adverse experience during foreign rule led to a strong 'self-reliance' focus on Indian growth model, as a result of which the economic development in early phase was guided by the ideology of nationalism and democratic socialism (Tendulkar and Bhavani 2007; Bandyopadhyaya 2006). Given the lack of importance attached to the need to promote exports during 1960s and 1970s, no special emphasis was laid on attracting FDI in that period. The inadequacy of the aforesaid policy, however, became apparent in 1980s and several reform measures were undertaken (Tendulkar and Bhavani 2007). The need for reform intensified in the post-Gulf war period, with continuous worsening of the CAB. While the break-up of the Soviet Bloc countries, in the short run, adversely affected Indian exports; the growing oil import bill simultaneously put tremendous upward pressure on imports. Macroeconomic mismanagement during the late 1980s further aggravated the problems (Bajpai 2002; Joshi and Little 1996). Finally, in 1991, as per the recommendations of the International Monetary Fund (IMF), India followed a structural adjustment programme. The new economic philosophy shifted towards export-oriented growth model, where concerns on augmenting competition in the domestic market through reforms in licensing provisions and adoption of better technological capabilities through FDI collaborations played crucial roles (Tendulkar and Bhavani 2007).

The continuous stream of economic reforms has improved the outward-orientation of Indian economy in considerable manner. While in 1991–1992 the outward-orientation of the economy (expressed as a percentage of export plus import as a percentage of gross domestic products (GDPs)) stood at 8.36 %, the same has improved to 51.24 % in 2011–2012. During 1990–2010 Indian merchandise exports and imports have grown at an annual average growth rate of 14.03 and 15.34 %, respectively. As a result, India's share in global merchandise export has increased from 0.50 % in 1991 to 1.44 % in 2010, while the corresponding figures for imports stood at 0.56 and 2.12 %, in that order. The service trade has witnessed a similar growth paradigm, with India's service exports and imports witnessing an average annual growth rate of 18.58 and 16.22 %, respectively. India's share in global commercial services export and import currently stand at 3.34 and 3.31 % respectively, vis-a-vis the corresponding figures of 0.60 and 0.69 % in 1991, in that order (calculated from ITS data, WTO).

Over the last decade, the high economic growth in India, resulting from the reforms, has motivated massive FDI inflows into the country (Chakraborty and Mukherjee 2012). The continuous inflow has caused India's share in global FDI inward stock to increase from 0.08 % in 1990 to 0.22 % and 1.03 % in 2000 and in 2010, respectively. It is observed that the FDI flows expressed as percentage of gross fixed capital formation (GFCF) in India has increased from 3.0 to 9.6 % over 2002–2008 (UNCTAD 2011).

However, the improved FDI scenario in India has simultaneously witnessed a decline in CAB. While in 1991–1992 the CAB expressed as a percentage of GDP stood at 0.35 %, the same has declined to 12.00 % in 2011–2012. It has been observed that capital account balance is significantly influencing CAB in India (Chakraborty et al. 2012). In this background, the current chapter attempts to explore the underlying long-term relationship between FDI inflow in India and its performance on CAB front.

The present chapter is arranged along the following lines. First, a literature review on the interrelationship between FDI and CAB has been undertaken. Next, the FDI-inflow pattern in India has been briefly analysed. Third, the CAB scenario in India has been discussed briefly. In the subsequent section, an empirical analysis has been carried out to understand the FDI–CAB relationship and causality pattern in the Indian context. Finally, on the basis of the discussions, certain policy conclusions are drawn.

2 FDI and its Impact on Current Account: Evidence from Literature

Before moving to analyse the effect of FDI on CAB, a brief discussion on the influence of foreign capital on the three subcomponents discussed earlier will not be irrelevant here. First, the evidence of FDI on exports is mixed. A section of literature has reported a positive influence of FDI on boosting exports from recipient countries (Dritsaki et al. 2004; Hossain 2008; Pfaffermayer 1994; Yamawaki 1991; Vural and Zortuk 2011; Chavez and Dupuy 2010). Conversely, the other branch of the literature reports that the FDI–export relationship may not necessarily be positive (Jeon 1992; Svensson 1996; Türkan 2006). The difference in the findings of the literature can be explained by the quality of FDI. FDI can be either vertical or horizontal. In the case of a horizontal FDI, the operations of the subsidiaries in recipient countries are geared for servicing the local markets and, hence, export promotion is not facilitated. Moreover, transfer of low-level technologies by multi-national companies (MNCs), coupled with the existing inefficiencies prevalent in recipient country firms, may adversely affect exports (Zhang 1999). However, vertical FDI leads to specialization in particular stages of production in different countries in line with their comparative advantages, and, hence, export to the partners of integrated production network (IPN) partners can be promoted (Lipsey 2004).

The CAB might also get adversely affected through augmentation of imports in the post-FDI period (Onwuka and Zoral 2009; Williams and Williams 1998), though the strength of the effect may vary depending on the motivation. For instance, if the MNC insists on importing specialized machineries and materials for production on the ground of their nonavailability in host countries, then imports increase as a result of FDI (Alguacil and Orts 2003). However, local production might contain the adverse effect (Blonigen 2001). Kinoshita (2011) has noted that during 2000–2007, FDI inflow in 15 Eastern European countries has majorly entered into the nontradable sector, as a result of which domestic demand rather than supply increased at a considerable pace. The development led to huge imports, and, consequently, to high levels of CAB.

Finally, remittances influence CAB scenario in considerable manner (Salisu 2005). It has been observed that the outflow of profit remittances on FDI flows leads to worsening of CAB (Yalta 2011). The effect may get stronger in the time of economic crisis (Doraisami 2007).

A number of studies have analysed the influence of FDI on BOP, in general, and CAB, in particular. A negative influence of FDI on CAB has been reported by several studies (Bosworth et al. 1999; Doraisami 2007; Jansen 1995; Mencinger 2008; Seabra and Flach 2005; Siddiqui and Ahmad 2012). Interestingly, a number of studies report profit remittances and higher import intensity as the major underlying factors for the decline in CAB. Analysing the data for Turkey, Yalta (2011) noted that while FDI lead to decline in exports, they result in increase in imports and profit remittances outflow, as a result of which CAB is de-stabilized. Analysing the scenario in Barbados, Campbell (2003) also noted that the possible gains derived from FDI might get eroded by the import of goods and services from abroad and investment income payments to nonresidents. The analysis of Muwanga-Zake and Katamba (2005) on Uganda explained the decline in CAB with faster growth in imports vis-a-vis exports, which, in the long run, may lead to chronic imbalance. A similar conclusion has been reached by Higgins et al. (2005), who have noted the adverse implications of the US net-income payments and CAB scenario.

However, positive influence of FDI on CAB has been reported by the other side of the literature. Fry et al. (1995) has noted that FDI is independent of current account, and neutrality increases with rise in openness of the exchange system. The analysis of Ehimare (2011) with data for Nigeria reported a positive relationship between

FDI and CAB, given its abundance in natural resources and large population, which signifies a large market. Fry (1996) has reported a positive relationship between FDI and CAB for six Pacific Basin economies, and explained the result in terms of their effects on national savings and accelerated growth.

3 FDI Scenario in India: Inflow and Effects

The process of liberalization slowed down in the late 1990s, but deepened again in the new millennium. The FDI inflow in India suitably reflects the gaining of pace of the reform measures. Rao and Dhar (2011) noted that the average reported FDI equity inflows from 1991–1992 to 1999–2000 was US$1.72 billion, but the same increased to US$2.85 billion from 2000–2001 to 2004–2005. In line with further economic reforms and emergence of the Indo-centric regional trade agreements (RTAs), an unprecedented FDI inflow has been observed afterwards, taking the corresponding figure to US$19.73 billion during 2005–2006 to 2009–2010. Capital account reform measures have been undertaken over the period, which considerably liberalized FDI inflow into the country. However, owing to Indian indirect taxes and transportation infrastructure FDI flows has been lower vis-à-vis the Chinese experience (Shah and Patnaik 2007).

Table 29.1 depicts sectoral distribution of FDI inflow in India in percent terms for the last one decade (2000–2011). It is observed that primary sector consistently shows a very meagre share in total FDI inflow, coming down from 1.06 % in 2000–2008 to 0.79 % in 2011. Mining has been the most prominent subsector within the primary sector in terms of drawing FDI, though the proportional contribution has come down in recent years.

In comparison with the primary sector, manufacturing sector has enjoyed a better FDI attractiveness. The percentage share for this segment stood at 32.08 % in 2000–2008, but declined to 22.31 % in 2009. However, a reversal has been witnessed in 2011 with the contribution of the sector increasing to 38.3 % in 2011. Within the manufacturing sector, automobile industry attracted most of the FDIs (as a single industry), with a share of 4.05 % in 2000–2008, which increased to 7.14 % in 2011. Metallurgical industry has been the other sector attracting most of the FDIs in the manufacturing sector. Proportional FDI inflow in other manufacturing sectors, like electrical equipment and textiles, has been erratic in recent years.

However, keeping at par with the past trends service and the related sector remains the most attractive destination for FDI inflow in India, with its share of 66.92 % during the time span of 2000–2008. Although the proportional inflow in this sector increased to 76.26 % in 2009, the global recession limited the FDI flow in this segment from then on and the corresponding figure has come down to 60.55 % in 2011. Computer software and hardware, telecommunication and real estate are among the largest recipients of FDI on this front. But notwithstanding the expectation, the proportional share of FDI in computer software and hardware industry has declined considerably from 11.56 % in 2000–2008 to 3.39 % in 2011. Real estate sector also shows major

Table 29.1 Sectoral distribution of foreign direct investment inflow in India (2000–2011). (Source: SIA Newsletter (April 2011))

Sector	(Percent Share (%))				
	2000–2008	2009	2010	2011	January 2000–December 2011 (Cumulative)
1. Primary	1.06	0.93	0.93	0.79	1.00
(a) Mining	0.65	0.66	0.49	0.25	0.61
2. Manufacturing	32.08	22.31	35.09	38.30	30.54
(a) Miscellaneous industries	5.63	3.06	7.67	7.14	5.43
(b) Automobile industry	4.05	4.98	5.93	7.76	4.66
(c) Metallurgical industries	3.26	1.78	5.12	3.90	3.25
(d) Petroleum and natural gas	2.72	1.44	2.90	1.02	2.42
(e) Chemicals (other than fertilizers)	2.51	1.67	2.12	1.32	2.23
(f) Electrical equipment	1.73	2.88	0.52	2.86	1.82
(g) Cement and gypsum products	2.09	0.31	2.96	0.91	1.80
(h) Drugs and pharmaceuticals	1.90	0.75	1.05	0.90	1.47
(i) Industrial machinery	0.39	0.71	3.34	1.13	0.96
(j) Food processing industries	1.00	0.72	0.99	1.68	0.95
(k) Textiles (including dyed and printed)	0.78	0.77	0.39	1.61	0.73
3. Service	66.92	76.26	63.90	60.55	68.35
(a) Services sector	21.62	21.13	17.63	16.03	20.71
(b) Computer software and hardware	11.56	2.66	4.77	3.39	8.23
(c) Telecommunications	7.86	9.34	7.08	9.76	8.12
(d) Housing and real estate (including cineplex, multiplex, integrated townships, commercial complexes, etc.)	5.82	12.18	7.06	3.05	7.38
(e) Construction activities	6.18	9.02	7.46	6.36	7.03
(f) Power	3.58	6.02	5.68	6.56	4.55
(g) Trading	1.88	2.51	2.67	1.47	2.14
(h) Hotel and tourism	1.50	2.18	2.35	2.49	1.82
(i) Information and broadcasting (including print media)	1.15	2.79	1.93	1.80	1.66
(j) Consultancy services	1.38	1.53	1.20	1.86	1.40

fluctuation in its FDI share, whereas telecommunication has been one of the relatively stable destinations of FDI.

The impact of the FDI inflow on India's growth and exports has, however, been mixed so far. Chakraborty and Nunnenkamp (2008) reported that while the growth effects of FDI are strong and moderate in the manufacturing sector and the services sector, respectively, no causal relationship is observed in the primary sector. Sharma (2000) noted a statistically nonsignificant relationship between FDI and export performance although the coefficient of FDI has a positive sign. Nevertheless, FDI in India has been able to create significant backward and forward linkages. NCAER (2010) noted that the sectors with both strong backward and forward linkages include construction, fuels, chemicals, metallurgical industries, etc. The sectors with strong backward linkages are electrical equipment, drugs and pharmaceuticals, food processing, textiles, etc., while telecommunications, consultancy services, etc. exhibit strong forward linkages.

The weak linkage between FDI and trade in India can be explained on two counts. First, the secondary data analysis of Singh (2007) observed that FDI has come in the most capital-intensive sectors with the consequent limited growth repercussions in the economy. The home-market effect might have dominated in this setting. Second, Banga (2003) noted that while FDI has a significant effect on the export-intensity of industries in the nontraditional export sector, the same for traditional export sector has been missing. Moreover, while FDI from the US bears a positive and significant effect on the export-intensity of industries in nontraditional export sector, a similar effect for Japanese FDI was not observed. The characteristic difference in horizontal and vertical FDIs may have led to such outcome.

The recent Indo-centric RTAs have played a favourable role in ensuring FDI inflows in the country, with obvious trade repercussions (Chaisse et al. 2011; Chakraborty and Sengupta 2010). For instance, towards the end of the Indo- Association of Southeast Asian Nations (ASEAN) free-trade area (FTA) negotiations Honda Motors expressed their interest in sourcing components from India for its manufacturing sites at other locations, especially within the ASEAN market (Economic Times 2008a). The proposal was soon followed by Mitsubishi Motors' expression of interest in starting production in India to make it an export hub, and to link the operations with the existing ASEAN bases (Economic Times 2008b). Many investment proposals from Southeast and East Asian partners have materialized since then. However, export promotion from these initiatives is still below expectation, as India's integration with international production networks (IPNs) is still at the nascent stage (Anukoonwattaka and Mikic 2011).

4 Movements in CAB: Indian Experience

Given the emergence of inefficient industrial structure resulting from the import-competing development strategy, India almost always experienced a negative CAB in the pre-liberalization period. Table 29.2 illustrates India's current account scenario for the last two decade (in Rs. (crores)). While both merchandise import and export show steady growth over the period, the import growth rate has been higher vis-a-vis the corresponding export growth rate. Consequently, trade balance has steadily and increasingly run into higher deficit. While India's trade deficit was Rs. 16,934 crore in 1990–1991, the same increased to Rs. 56,737 crore in a decade's time in 2000–2001. Eventually, the figure has magnified to Rs. 595,600 crore in 2010–2011.

On the other hand, the country also faced an invisible deficit in 1990–1991, which soon changed in the post-liberalization period. India has witnessed a steady rise in service exports since late 1990s, which continued in the new millennium, as well (Acharya 2001). After the recent global recession, the service sector export growth rate has declined resulting into lower invisible surplus on 2010–2011. Though India's net invisible remains positive but the huge trade deficit prompted the over CAB to stay perennially at deficit. The total current account deficit in India was Rs. 17,366 crore in 1990–1991, which declined to Rs. 11,598 crore in 2000–2001 but reached a staggering Rs. 210,100 crore in 2010–2011.

Table 29.2 Current account balance scenario in India. (Source: Economic Survey (2012))

Category	(Rs. (crore))				
	1990–1991	2000–2001	2006–2007	2008–2009	2010–2011
Imports	50,086	264,589	862,833	1,405,400	1,735,100
Exports	33,153	207,852	582,871	858,000	1,139,500
Trade balance	– 16,934	– 56,737	– 279,962	– 547,400	– 595,600
Invisible receipts	13,396	147,778	517,146	770,400	902,500
Invisible payments	13,829	102,639	281,567	350,600	517,000
Net invisible	– 433	45,139	235,579	419,800	385,500
CAB	– 17,366	– 11,598	– 44,383	– 127,600	– 210,100

Mohan (1996) noted that for the sustainability of India's current account deficit at a further lower level, a wider export growth rate is required. India's export growth rate in the early years of the new millennium was high, which resulted in a current account surplus in 2003–2004. However, Shah and Patnaik (2007) noted that in the Indian case a current account deficit scenario has been considered more favourable for long-term growth vis-à-vis current account surplus, with obvious capital inflow considerations. In this context, the study observed that, "Many economists argued that the current account surplus in 2003–2004, of 1.7 % of GDP, implied a significant opportunity cost in terms of investment forgone and thus lower GDP growth (Lal et al. 2003)". In this background, the present chapter attempts to understand the extent to which the FDI inflow in India influences the CAB series.

5 Data and Methodology

5.1 Data

Quarterly data over the period 1990–1991:Q1 to 2010–11:Q4[1] are used to examine the long run equilibrium or the co-integrated relationship between CAB and FDI for the Indian economy. The data have been compiled from Handbook of Statistics on Indian Economy (2011–2012), published by the Reserve Bank of India. All the variables are calculated in home currency price (Rs. (crores)).

5.2 Unit Root Test

Traditionally, the stationary properties of variables are examined by using Augmented Dickey–Fuller (ADF) (1979) and Phillips–Perron (PP) (1988) unit root tests. However, as suggested by Perron (1989), the standard unit root tests are biased towards the nonrejection of null hypothesis in the presence of structural breaks. "Most

[1] The financial year for the Indian economy ranges from April (of the current calendar year) to March (of the next calendar year).

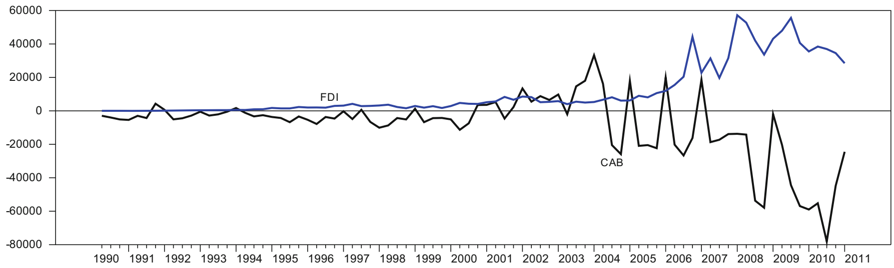


Fig. 29.1 Current account balance and foreign direct investment for the Indian economy (in Rs. (crores))

macroeconomic time series are not characterized by the presence of a unit root. Fluctuations are indeed stationary around a deterministic trend function. The only 'shocks' which have had persistent effects are the 1929 crash and the 1973 oil price shock" (Perron 1989, p. 1361). This is an important finding, especially because the span of time series in any empirical work is usually long enough to have had structural breaks.

A plot of the data on CAB and FDI indicates that there is a possible break in both the individual time-series, clustered around 2005–2006 (Fig. 29.1). Determining the accurate sequence of such break dates is a major task for researchers analysing time-series data on FDI and CAB. If the analysis is done oblivious of the existence of possible structural breaks, the empirical study using standard unit root may then yield confusing and spurious results. In other words, given the strong likelihood that the series under consideration are subject to structural breaks, the standard unit root tests for stationarity are likely to yield misleading conclusions.

There are a plethora of unit root tests in the presence of structural breaks to choose from. Perron's (1989) method of exogenous break-point treatment has been criticized by Christiano (1992) and later by Zivot and Andrews (1992) on the ground that the choice of break point is based on a pretest examination of the data, and, hence, is subject to the problem of 'data-mining'. Unit root test against a single-break stationary alternative was proposed by Zivot and Andrews (1992). Their test on endogenous structural break is a sequential test that uses a different dummy variable for each possible break date. It was extended to a two-break stationary alternative by Lumsdaine and Papell (1997) and up to a five-break alternative, with an a-priori unknown number of breaks, by Kapetianos (2005). However, these tests maintain the linearity assumption under the unit root null hypothesis and exhibit size distortions (over rejection of the null), as well as the wrong estimation of the break point (Nunes et al. 1997, Altinay 2005, Bec and Bassil 2009, etc.). To overcome these problems, Lee and Strazicich (2003, 2004) have proposed a Lagrange Multiplier (LM) test statistics-based unit root test that allows for (at most two) endogenous breaks both under the null and the alternative hypotheses. Thus, any conclusion on the rejection of unit root null based on this LM test provides quite a strong evidence of stationarity.

Let us consider the following data-generating process (DGP):

$$y_t = \delta' Z_t + e_t, \qquad e_t = \beta e_{t-1} + \varepsilon_t \tag{29.1}$$

where Z_t is a vector of exogenous variables, δ' is a vector of parameters and ε_t is a white noise process, such that $\varepsilon_t \sim$ NIID (0, σ2). First, the test with only one structural break is considered (Lee and Strazicich 2004). The 'crash' model that allows shift in level only is described by $Z_t = [1, t, D_t]'$, and the 'break' model that allows for changes in both level and trend is described as $Z_t = [1, t, D_t\ DT_t]'$, where D_t *and* DT_t are the two dummies defined as

$$D_t = 1, if\ t \ \geq\ T_B + 1;$$
$$= 0, otherwise$$

and

$$DT_t = t - T_B, if\, t \geq T_B + 1;$$
$$= 0,\ otherwise$$

where T_B is the time period of the break date.

Second, the framework that allows for two structural breaks are considered (Lee and Strazicich 2003). The crash model that allows two shifts in levels only is described by $Z_t = [1, t, D_{1t}\ D_{2t}]'$, and the break model that allows for two changes in both level and trend is described as $Z_t = [1, t, D_{1t}\ D_{1t}\ D_{2t}\ D_{2t}]'$, where D_{jt} and DT_{jt} *for* $j = 1{,}2$ are the appropriate dummies defined as above, viz.,

$$D_{jt} = 1,\ if\ t\ \geq\ T_{Bj} + 1;$$
$$= 0,\ otherwise$$

and

$$DT_{tj} = t - T_{Bj},\ if\ t\ \geq\ T_{Bj} + 1;$$
$$= 0,\ otherwise$$

Where T_{Bj} denotes the j-th break date.

This method estimates the following regression to obtain the LM unit root test statistics:

$$\Delta\ y_t = \delta'\ \Delta\ Z_t + \phi\ \tilde{S}_{t-1} + \sum_{i=1}^{k} \gamma_i\ \Delta\ \tilde{S}_{t-j} + u_t, \tag{29.2}$$

where $\tilde{S}_t = y_t - \widetilde{\Psi}_t - Z_t\tilde{\delta}, t = 2, \ldots., T$; $\tilde{\delta}$ denotes the regression coefficients of Δy_t *on* ΔZ_t and $\widetilde{\Psi}_t = y_1 - Z_1\tilde{\delta}$, y_1 and Z_1 being the first observations of y_t and Z_t respectively. The lagged terms $\Delta\tilde{S}_{t-j}$ are included to correct for likely serial correlation in errors. The null hypothesis of unit root ($\phi = 0$) is tested by the LM t-statistic using equation (2). The lag length k is selected by employing a general-to-specific (GTS) approach in all a-priori unknown break unit root tests[2]. The critical values are tabulated in Lee and Strazicich (2003, 2004) for the two-break and single-break cases, respectively.

[2] The lag length is also counter-checked using different lag selection criteria, like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), etc.

6 Empirical Results

Tables 29.3–29.5 contain the results for the LM test with one and two endogenous break(s) at level and first differences. The results suggest that the null hypothesis of a unit root can be rejected for both the series in levels for the crash model with both one and two endogenous structural breaks. In other words, both the CAB and FDI series are stationary for the crash model in the presence of endogenous structural break at 5 % level of significance. For the break model, the null hypothesis of a unit root can be rejected only with two structural breaks. However, if we take the first difference for the break model with one structural break, the unit root null for both the CAB and FDI series can be rejected, thereby suggesting that they are integrated of order 1. The results are in sharp contrast to the mixed orders of integration for the two series in the absence of endogenous structural breaks (Table 7 in the Appendix). As is evident from Tables 29.3–29.5, the break dates for both the series are concentrated around 2005–2006.

6.1 Co-integration

The tests for co-integration between two time-series can be conducted using a vector autoregressive model (VAR)-based framework. Under the Johansen (1988) approach, there are two tests. The first is the 'trace test'. It is a joint test where the null hypothesis is that the number of co-integrating vectors is greater than or equal to r, against an unspecified alternative that they are greater than r. The test statistic is formulated as:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i), \tag{29.3}$$

where r is the number of co-integrating vectors under the null hypothesis and $\hat{\lambda}_i$ is the estimated value for the i-th ordered Eigen value. Intuitively, the larger is $\hat{\lambda}_i$, the more large and negative will be $\ln(1 - \hat{\lambda}_i)$ and, hence, larger will be the test statistic. The second is the 'maximum Eigen value test' for co-integration which conducts separate tests on each Eigen value, and has as its null that the number of co-integrating vectors is r against an alternative of $r + 1$. The test statistic is formulated as:

$$\lambda_{\max}(r \mid r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{29.4}$$

for $r = 0, 1 \ldots k$-1. The critical values for the two statistics were provided by Johansen and Jesulius (1990).

In the present analysis, the co-integration test is performed to investigate whether the long-run equilibrium relationships among the two variables CAB and FDI exist. For the co-integration analysis, we consider the findings of the unit root test in the presence of two endogenous structural breaks, whereby the null hypothesis of a unit

Table 29.3 Unit root tests with one structural break (at level)

Series	Model	Break point	Optimal lags	Test statistics	Critical values at 1 %	Critical values at 5 %	Result
CAB	Crash (intercept only)	2005–2006:Q4	8	− 3.92	− 4.24	− 3.57	Reject null hypothesis of unit root at 5 % level
CAB	Break (intercept and trend)	2005–2006:Q4	8	− 3.78	(− 5.05 to − 5.11)	(− 4.45 to − 4.51)	Do not reject null hypothesis of unit root
FDI	Crash (intercept only)	2008–2009:Q2	7	− 3.59	− 4.24	− 3.57	Reject null hypothesis of unit root at 5 % level
FDI	Break (intercept and trend)	2006–2007:Q1	6	− 4.05	(− 5.05 to − 5.11)	(− 4.45 to − 4.51)	Do not reject null hypothesis of unit root

Note**:** Method applied is Lee and Strazicich's (2004)

Table 29.4 Unit root tests with one structural break (at first difference)

Series	Model	Break point	Optimal lags	Test statistics	Critical values at 1 %	Critical values at 5 %	Result
CAB	Break (intercept and trend)	2008–2009:Q3	6	− 9.85	(− 5.05 to − 5.11)	(− 4.45 to − 4.51)	Reject null hypothesis of unit root at 1 % level
FDI	Break (intercept and trend)	2005–2006:Q3	8	− 4.52	(− 5.05 to − 5.11)	(− 4.45 to − 4.51)	Reject null hypothesis of unit root at 5 % level

Note: Method applied is Lee and Strazicich's (2004)

Table 29.5 Unit root tests with two structural breaks (at level)

Series	Model	Break point (1)	Break point (2)	Optimal lags	Test statistics	Critical values at 1 %	Critical values at 5 %	Result
CAB	Crash (intercept)	2005–2006:Q4	2007–2008:Q4	8	– 4.27	– 4.55	– 3.84	Reject null hypothesis of unit root at 5 % level
CAB	Break (intercept and trend)	2002–2003:Q4	2006–2007:Q3	1	– 7.66	(– 6.16 to – 6.45)	(– 5.59 to – 5.74)	Reject null hypothesis of unit root at 1 % level
FDI	Crash (intercept)	2006–2007:Q1	2007–2008:Q4	7	– 3.88	– 4.55	– 3.84	Reject null hypothesis of unit root at 5 % level
FDI	Break (intercept and trend)	2005–2006:Q3	2007–2008:Q3	7	– 7.54	(– 6.16 to – 6.45)	(– 5.59 to – 5.74)	Reject null hypothesis of unit root at 1 % level

Note: Method applied is Lee and Strazicich's (2003)

Table 29.6 Results for Johansen test for co-integration

Test	Hypothesized number of co-integrating equation(s)	Eigen value	Test statistics	0.05 Critical value	Prob.**
Trace test	None*	0.18	17.53	15.49	0.02
	At most 1	0.01	1.08	3.84	0.30
Maximum Eigen-value test	None*	0.18	16.46	14.26	0.02
	At most 1	0.01	1.08	3.84	0.30

Notes: (1) Trace test and maximum Eigen-value test indicate 1 co-integrating equation at 0.05 level
(2) * denotes rejection of the hypothesis at the 0.05 level
(3) ** MacKinnon–Haug–Michelis (1999) *p*-values

Table 29.7 Results of Granger causality test procedure

Null hypothesis	F-statistic	*p*-values
CAB does not Granger cause FDI	1.21	0.30
FDI does not Granger cause CAB	14.46	0.000*

Notes: (1) Asterisk (*) denotes statistically significant at 1 % level

root has been rejected both for the crash and break models. The computed trace and maximum Eigen value statistics, as defined in Eqs. (29.3) and (29.4), are reported in Table 29.6, which indicates the presence of one co-integrating vector at 5 % level of significance (i.e. the null hypotheses of no co-integration is rejected for rank of zero only).[3] The result implies that there exists a unique long-run relationship among the two variables.

6.2 *Granger Causality Test*

Granger (1969) causality technique is commonly applied for identifying the direction of causal relationship between two variables.[4] The results of Granger causality test are depicted in Table 29.7. The analysis reveals a unidirectional causality from India's FDI to CAB at 5 % level. Our results are in conformity with similar findings for emerging countries (e.g. study by Siddiqui and Ahmed 2012 for Pakistan). Such one-way causality has important implications in the sense that FDI inflows may not necessarily contribute towards income-generating activities. On the contrary, they may increase conspicuous import-based consumption and repatriate the proceeds back home in the form of high returns. This has the potential to deteriorate the country's BOPs in the long run.

[3] The current analysis has considered the intercept and trend in co-integration test specification, i.e. a linear trend in Vector Autoregression (VAR).

[4] Hsiao's (1981) optimal lag length criteria for each variable in an equation based on a systematic autoregressive method has been applied in the analysis. This method combines Granger causality test and Akaike's final prediction error (FPE), defined as the (asymptotic) mean-square prediction error.

7 Policy Implications

The present analysis has identified the financial year 2005–2006 as the structural break year, which bears some significant implications vis-a-vis CAB and FDI due to several reasons. First, in 2005–2006 capital flows more than made up for the current account deficits of US$9.2 billion and resulted in reserve accretion. Second, India's impressive export growth in the first half of last decade mainly resulted from favourable external developments and domestic policy initiatives. Improved global growth and recovery in world trade since 2001 augmented the growth of Indian exports, while the gradual opening up of the economy and corporate restructuring strengthened the competitiveness of Indian industry. Third, improved domestic economic activity and the improvement of manufacturing sector provided a supporting base for strong sector-specific exports. Finally, the nominal effective exchange rate (NEER) measuring the value of country's currency relative to the currencies of principal trading partners depreciated on a yearly basis till 2004–2005 but appreciated in 2005–2006.

Overall, in 2005–2006, current receipts (including grants) grew by 27.6 % to US$197.4 billion. But, such receipts fell short of current payments (including grants) which grew by 31.4 % to US$206.6 billion. Current receipts covered 95.6 % of current payments in 2005–2006. During 2005–2006, for exports, while volume increased by a record 45.4 % (mainly in items like petroleum products, chemicals and related products and machinery and transport equipment), the unit value increased by 20.4 % (mainly in petroleum products, minerals and ores, machinery and transport equipment and footwear). The stable capital flows seamlessly financing the moderate levels of current account deficit caused primarily by the rise in international oil prices. On the other hand, merchandise imports grew by 33.8 % to US$149.2 billion in 2005–2006. Also, remittances outflow increased as all investments in India are freely repatriable after the payment of applicable taxes. Thus, despite the tremendous growth in export, trade deficit reached a record high of US$46 billion in 2005–2006.

On the capital flow front it is observed that external assistance and external commercial borrowing (ECB), the two major debt-creating flows picked up in 2004–2005. These debt flows, as a proportion of total capital flows, were 25 % in 2004–2005 and 18 % in 2005–2006. FDI inflows (net), which had declined from US$4.7 billion in 2001–2002 to US$2.4 billion in 2003–2004, continued its growth for the second consecutive year in 2005–2006 to climb back to US$4.7 billion again. FDI on a comparative net basis, year-on-year, exhibited a growth of 27.4 % in 2005–2006 reflecting the improved investment climate. FDI (net) in April–September 2006 at US$4.2 billion was almost twice its level in April–September 2005. Thus, the year 2005–2006 has been a trend-breaking one in terms of both current account scenario and FDI flow, thus, supporting the obtained structural break result.

The empirical result raises serious concerns regarding the sustainability of the current account deficit in India. The adverse effect of FDI on CAB, as observed from the empirical results implies that the export opportunities arising out of the foreign investment are being outweighed by the rising import volume and remittance payments leakages. The fact is indicative of two major shortcomings of the Indian economy: one, lesser competitiveness of Indian exports leading to failure to enhance the same in the world market and two, lower technological plane, which affects export prospects on one hand and inflates import bill on the other hand. There is a strong need to enhance the competitiveness as well as technology spillover effect in both manufacturing and service sectors. The recent policy measures being implemented (NMCC 2006, 2011; Prasad and Satish 2010) need to be further strengthened in light of the present findings.

Appendix: Analysis without structural breaks

Table A.1 Unit root tests

Variables	ADF		PP	
	Level	First difference	Level	First difference
CAB	– 1.23	– 10.75*	– 4.59*	–
FDI	– 4.31*	–	– 3.13	– 14.93*

ADF augmented Dickey–Fuller test
PP Philips–Perron test
Asterisk (*) denotes statistically significant at 1 % level
Results reported are those with drift and trend

References

Alfaro L, Areendam C, Sebnem K-O, Selin S (2006) How does foreign direct investment promote economic growth? exploring the effects of financial markets on linkages. Working paper, Harvard Business School

Alguacil MT, Orts V (2003) Inward foreign direct investment and imports in Spain. Int Econ J 17(3):19–38

Antilay G (2005) Structural breaks in long-term Turkish macroeconomic data, 1923–2003. Appl Econometrics Int Devel 5:117–130

Ayanwale AB (2007) FDI and economic growth: evidence from Nigeria. AERC Research Paper 165. African Economic Research Consortium, Nairobi

Bajpai N (2002) A decade of economic reforms in India: the unfinished agenda. Center for International Development Working Paper No. 89. Harvard University

Bandyopadhyaya J (2006) The making of India's foreign policy. Allied Publishers, New Delhi

Bec F, Bassil C (2009) Federal funds rate stationarity: new evidence. Econ Bull 29:867–872

Blonigen B (2001) In search of substitution between foreign production and exports. J Int Econ 53(1):81–104

Borensztein E, Gregoria J De, Lee J (1998) How does foreign investment affect economic growth? J Int Econ 45(1):115–135

Bosworth B, Collins S, Reinhart C (1999) Capital flows to developing economies: implications for saving and investment. Brookings Pap Econ Act 1:143–180

Braunstein E, Epstein G (2002) Bargaining power and foreign direct investment in China: can 1.3 billions consumers tame the multinationals? CEPA Working Paper 2002/13. Center for Economic Policy Analysis, New York

Campbell T (2003) Foreign direct investment inflows and their impact on barbados, current account: implications for policy. In: Harold C, Ronald C, Darrin D (eds) Facing globalisation—impact and challenges for barbados and the caribbean. Central Bank of Barbados, Bridgetown, pp 5–24
Chakraborty D, Mukherjee J (2012) Is there any relationship between foreign direct investment, domestic investment and economic growth in India? A time series analysis. Rev Market Integr 4(3) (forthcoming)
Chakraborty D, Mukherjee J, Sinha T (2012) Is there any long-run relationship between India's current and capital account balance? A time series analysis. Global Bus Rev 13(3):433–447
Chavez B, Dupuy J (2010) Inward FDI in Peru and its policy context. Vale Columbia Center on Sustainable International Investment. http://www.vcc.columbia.edu/files/vale/documents/Peru_IFDI_Aug_4_FINAL.pdf. Accessed 3 Oct 2012
Christiano LJ (1992) Searching for a break in GNP. J Bus Econ Statist 10:237–250
Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431
Doraisami AGG (2007) Financial crisis in Malaysia: did FDI flows contribute to vulnerability? J Int Devel 19:949–962
Dritsaki M, Dritsaki C, Adamopoulus A (2004) A causal relationship between trade, FDI and economic growth for Greece. Am J Appl Sci 1(3):230–235
Dunning JH, Rugman AM (1985) The influence of Hymer's dissertation on the theory of foreign direct investment. Am Econ Rev 75(2):228–232
Ehimare OA (2011) Foreign direct investment and its effect on the Nigerian economy. Bus Intell J 4(2):253–261
Fry M (1996) How foreign direct investment in pacific Asia improves the current account. Birmingham J Asian Econ 7(3):459–486
Fry M, Claessens S, Burridge P, Blanchet M (1995) Foreign direct investment, other capital flows and current account deficits: what causes what? World Bank, Policy Research Working Paper No.1527
Government of India (2012) Economic survey 2011–12. Ministry of Finance, New Delhi
Granger CWJ (1969) Investigating causal relation by econometric and cross-sectional method. Econometrica 37:424–438
Higgins M, Klitgaard T, Tille C (2006) Borrowing without debt? Understanding the U.S. international investment position. Staff Reports 271. Federal Reserve Bank of New York
Hossain MA (2008) Impact of foreign direct investment on Bangladesh's balance of payments: some policy implications. Policy Note No. 5. Bangladesh Bank, Dhaka
Jansen K (1995) The macroeconomic effects of foreign direct investment: the case of Thailand. World Devel 23(2):193–210
Jeon Y (1992) The determinants of Korean foreign direct investment in manufacturing industries. Weltwirtsch Arch 128:527–541
Johansen S (1988) Statistical analysis of co-integration vectors. J Econ Dynam Control 12:231–254
Johansen S, Jesulius K (1990) Maximum likelihood estimation and inference on co-integration with the application to the demand for money. Oxford Bull Econ Statist 52:169–210
Joshi V, Little IMD (1996) India's economic reforms: 1991–2001. Oxford University Press, New Delhi
Kapetanios G (2005) Unit root testing against the alternative hypothesis of m structural breaks. J Time Ser Anal 26:123–233
Katerina L, Papanastasiou J, Vamvakidis A (2004) Foreign direct investment and economic growth in transition economies. South-E Eur J Econ 1:97–110
Katircioglu S (2009) Foreıgn direct investment and economic growth ın Turkey: an empirical investigation by the bounds test for co-integration and causality tests. Econ Res 22(3):1–9
Kinoshita Y (2011) Sectoral composition of foreign direct investment and external vulnerability in eastern Europe. IMF Working Paper 11/123. International Monetary Fund
Krkoska L (2001) 'Foreign direct investment financing of capital formation in central and eastern Europe', Working Paper No. 67. European Bank for Reconstruction and Development, London

Lee J, Strazicich M (2003) Minimum LM unit root test with two structural breaks. Rev Econ Statist 85:1082–1089

Lee J, Strazicich M (2004) Minimum LM unit root test with one structural break. Working Paper. Department of Economics, Appalachian State University

Lipsey R (2004) Home- and host-country effects of foreign direct investment. In: Robert EB, Winters LA (eds) 'Challenges to globalization: analyzing the economics'. University of Chicago Press

Lipsey R, Weiss NY (1984) Foreign production and exports of individual firms. Rev Econ Statist 66:304–307

Lumsdaine R, Papell D (1997) Multiple trends breaks and the unit root hypothesis. Rev Econ Statist 79:212–218

Markusen JR (1984) Multinationals, multiplant economies, and the gains from trade. J Int Econ 16:205–226

Markusen JR, Venables AJ (1999) Foreign direct investment as a catalyst for industrial development. Europ Econ Rev 43(2):335–356

Mencinger J (2008) The "addiction" with FDI and current account balance. International Centre for Economic Research Working Papers No. 16

Muwanga-Zake ESK, Katamba PM (2005) Capital flows and current account sustainability—Uganda's experience: 1994–2004. Presented in the Workshop on "Capital flows and current account sustainability in African economies", Organized by Economic and Social Policy Division, United Nations Economic Commission for Africa, 21–22 September. Accra

Nunes L, Newbold P, Kuan CM (1997) Testing for unit roots with breaks: evidence on the great crash and the unit root hypothesis reconsidered. Oxford Bull Econ Statist 59:435–448

Onwuka K, Zoral KY (2009) Foreign direct investment and imports growth in Turkey. J Yasar Univ 4(15):2357–2380

Perron P (1989) The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57:1361–1401

Phillips RCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75(2): 335–346

Pfaffermayer M (1994) Foreign direct investment and exports: a time series approach. Appl Econ 25:335–351

Salisu M (2005) The role of capital flight and remittances in current account sustainability in Sub-Saharan Africa. Presented in the Workshop on "Capital flows and current account sustainability in African economies", Organized by Economic and Social Policy Division, United Nations Economic Commission for Africa, 21–22 September, Accra.

Seabra F, Flach L (2005) Foreign direct investment and profit outflows: a causality analysis for the Brazilian economy. Econ Bull 6(1):1–15

Siddiqui DA, Ahmad MH (2012) The causal relationship between foreign direct investment and current account: an empirical investigation for Pakistan economy. Eur J Econ Financ Adm Sci 44:107–116

Svensson R (1996) Effects of overseas production on home country exports: evidence based on Swedish multinationals. Weltwirtsch Arch 132:304–329

Tang S, Selvanathan EA, Selvanathan S (2008) Foreign direct investment, domestic investment, and economic growth in China. Research Paper No. 2008/19, UNU-Wider (United Nations University)

Tendulkar SD, Bhavani TA (2007) Understanding reforms: post 1991 India. Oxford University Press, New Delhi

Türkan K (2006) Foreign outward direct investment and intermediate goods exports: evidence from USA. http://www.etsg.org/ETSG2006/papers/Turkcan.pdf. Accessed 5 Oct 2012

United Nations Conference on Trade and Development (UNCTAD) (2012) World investment report 2012: towards a new generation of investment policies. Geneva

United Nations Conference on Trade and Development (UNCTAD) (2011) World investment report 2011: non-equity modes of international production and development. Geneva

United Nations Conference on Trade and Development (UNCTAD) (2002) World investment report 2002: transnational corporations and export competitiveness. Geneva
Vural İY, Zortuk M (2011) Foreign direct investment as a determining factor in Turkey's export performance. Eurasian J Bus Econ 4(7):13–23
Williams O, Williams S (1998) Foreign direct investment flows to the eastern Caribbean central bank region. Central Bank of Barbados Working Papers, pp. 179–194
World Trade Organization (2011) International trade statistics. Geneva
Yalta AY (2011) Uncovering the channels through which FDI affects current account: the case of Turkey. Working Paper No. 8. Department of Economics, TOBB University of Economics and Technology, Ankara
Yamawaki H (1991) Exports and foreign distributional activities: evidence on Japanese firms in the United States. Rev Econ Statist 73:294–300
Zhang KH (1999) How does FDI affect a host country's export performance? The case of China. Econ Syst 21(4):291–303
Zivot E, Andrews D (1992) Further evidence on the great crash, the oil price shock and the unit root hypothesis. J Bus Econ Statist 10:251–270

Chapter 30
Contagious Financial Crises in the Recent Past and Their Implications for India

Asim K. Karmakar

There have been big changes in the contours of the global economic map. But the fact remains that the actual extent of global shifts in economic activity is extremely uneven. Only a small number of developing countries have experienced substantial economic growth; a good many are in deep financial difficulty whilst others are at, or even beyond, the margins of survival... although we can indeed think in terms of a new international division of labours, its extent is far more limited than is something claimed.

P. Dicken in Global Shift

1 Introduction

The phenomenon of financial market crises—whether on issued external debt, in the banking sector, on the foreign exchange market augmenting currency crises or in all of these at more or less the same time—which have their origin in one particular country and tend to spread to other countries to be explained by interdependence in trade or exposure to common macroeconomic factors, is not a new one and is called in modern terminology 'contagion' that has been recoined from the earlier term 'international propagation' as coined by Charles Kindleberger in his classic of 1978, *Manias, Panics, and Crashes*. The more the interdependence among the countries in trade or exposure to common macroeconomic factors, the greater will be the contagion, while the main sources of it can be traced in common financial linkages, pathologies in the diffusion of information among the agents and, finally, financial fragility that underlies international contagion: rapid inflows of capital, macroeconomic shocks that occur too rapidly for gradual portfolio rebalancing and a leveraged common creditor.

The present chapter in this context endeavours to have a look back and forward into the sad story of so many financial market crises.

Capital flows into emerging economies have grown twice as fast as those into developed economies since the 1990s. Over time, there have been notable changes in the form and nature of international capital flows. The first important trend is that

A. K. Karmakar (✉)
Department of Economics, Jadavpur University, Kolkata 700 032, West Bengal, India
e-mail: asimkkarmakar@yahoo.co.in, iasimkkarmakar@gmail.com

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6_30,

the vast majority of these flows are driven by portfolio investment. A second trend is the rise in private capital flows. Private capital flows represented more than 80 % of all flows in 2011. A third trend is the increasing integration of developing states into global financial markets. These states have become an important destination for global capital.

Larger capital flows meant larger current account deficits, given the difficulty of sterilizing these inflows, and real exchange rate appreciation. Both the deficits and the large real appreciation are sources of vulnerability when financial market conditions are disturbed.

In the first place in this chapter, we recall early financial market crises in economic history which helps us to understand more about the contagion processes observed in more recent years. In the following section, our aim is to picture the important events that occurred in Chile in 1982, crisis of the exchange rate mechanism (ERM) during the early 1990s, the Mexican crisis of 1994–1995. Last of all, we provide a chronicle of the more recent crisis of 1997–1998 in Thailand and the subsequent turmoil in many Asian economies as well as global economic meltdown and the Eurozone crisis of recent years and their implications for India. The chapter ends with a conclusion.

2 The Background

Financial crisis refers to a rapid financial disintermediation due to financial panic. All the booms and busts, from the 'Tulip Fever' in Amsterdam (1634–1637), the French 'Mississippi Bubble' (1716–1720), 'South Sea Bubble' (1717–1720) in London, through the first Latin American debt crisis (1820s), through all the manias of the nineteenth century, the canal mania, the mining mania, British commercial crisis of 1847–1848 (resulted from railway overinvestment 'boom', business failures 'distress' and a monetary 'panic') on to the crash of 1929 and beyond, many memorable financial crises in the late twentieth century and the first decade of the twenty-first century blotted the landscape of capitalist progress .

The ultimate manifestation of financial crisis includes bank failures, stock market crashes like the 1987 New York Stock market crisis and currency crises, occasionally leading to deep recessions. During the last quarter of the twentieth century, observers focused attention on the growing role of international trigger for financial crises—an outcome of the collapse of the Bretton Woods (BW) system, the rapid rise in the importance of emerging markets in the global economy and the growing financial integration of countries with the global financial system. It is noteworthy that after the end of World War II, when the USA and its allies emerged victorious, the colonial system dismantled, while on the eve of the end of war, the international conference at BW in 1944 had established a monetary order to fix the exchange rate. It is a well-known story. The American, Harry Dexter White's view that the dollar would be used—at fixed parity with gold—and the International Monetary Fund (IMF) as a stabilization fund held sway, while the British John Maynard Keynes' view that the world wanted a world currency, the Bancor, against which all currencies would be exchangeable at fixed rates, and an International Clearing Union, a kind of world central bank, were ultimately rejected. For a while, under the aegis of this imperial system, international trade and investments proceeded, indeed expanded, relatively smoothly. But over time, this historical settlement was brought down by its own contradictions. The de facto 'dollar standard' led, over time, to an over-valuation of the dollar and a massive outflow of American capital. The consequent prolonged US balance of payments (BoP) deficits, thus, became a principal source of international liquidity that gave rise to 'offshore' or 'eurodollar' markets. The practical impossibility of reversing these offshore dollars into 'naïve' dollars, due to the restrictions of BW agreement, led to speculative movements against the dollar in these markets. By 1971, these speculative movements reached to such proportions that in August 1971 the US government, under Nixon, declared the dollar no longer convertible into gold, thereby effectively ending the BW exchange system. A 10 % devaluation of the dollar soon followed. The consequence of this was twofold: on the one hand, a tremendous increase in volatility in the international currency markets, and, on the other, more generally, a huge increase in international credit risk. Companies or banks investing abroad would have to run the whole gauntlet of volatile currencies as well as commercial risk of dealing with counter parties. This increase in risk, over time, led to the invention of a whole class of new financial instruments, to spread risk, to reduce exposure and to ensure against risk: The most important of these derivatives, besides futures and options, were currency swaps, interest rate swaps, credit default swaps (CDS) and the credit debt obligations (CDOs). 'What began as earnest "hedging against risk" to cope with uncertainty and financial volatility after the end of Bretton Woods and the 1973 oil crisis soon developed into endless speculation and gambling' (Chancellor 2000).

However, the 1980s opened with a new financial phase. Here, it was not structural adjustment programmes but financial adjustment crises. Since the 1980s, there have been several financial crises, some famous and some obscure as the individual-country financial crises that happened in more than 80 countries in the 1980s and 1990s as they deregulated their financial systems, mostly under the pressure from global regulators aiming at facilitating the globalizing of financial markets. Glick and Hutchison (1997) investigated a sample of 90 countries during 1975–1997, covering 90 banking crises, 202 currency crises and 37 twin crises. They found that banking and twin crises had occurred mainly in developing countries, and their number increased in the 1990s. Twin crises are mainly concentrated in financially liberalized emerging-market economies. The costs of these crises are substantial—currency and banking crises are very costly, reducing output by 8–10 % over a 2–4-year period (Hutchison and Noy 2005).

The resumption of capital flows to developing countries in the early 1990s led to the waves of 'sudden stops' (the abrupt cessation of foreign capital inflows and the reversals of capital flows), starting with the Mexican crisis of 1994–1995, continuing with the Russian and East Asian crises in the second half of the 1990s and culminating with the Argentinean meltdown in the early 2000s (Calvo 1998; Edwards 2004) .

Most of the financial crises in the 1990s and the early 2000s affected the developing and emerging markets, leading to a heated debate regarding their causes and the needed remedies. There is solid evidence that financial opening (i.e. the dismantling of capital controls) increases the chance of financial crises. In particular, while financial opening increases a country's overall welfare, financial opening may be welfare reducing in the presence of other distortions. An example of such a distortion is moral hazard, which frequently acts as an implicit subsidy to borrowing and investment, ultimately leading to over-borrowing and crisis (McKinnon and Pill 1999). Moral hazards arise when investors believe that they will be bailed out of their bad investments by the taxpayer and, therefore, have little incentive to undertake proper monitoring of their investments.

3 Theoretical Explanation of Financial Crises

In general, the underlying causes of a currency crisis are inconsistencies between a country's exchange rate policy and its domestic economic policies so that the currency is either over- or undervalued relative to the currencies of other countries. The numerous financial crises that have ravaged the world markets as well as mature economies have fuelled a continuous interest in developing models for explanation of financial crises. These theoretical models, sometimes called models on BoP crises, are even catalogued into three different generations. One set of explanations of currency crises, termed first-generation models (e.g. Krugman 1979), directs attention to inconsistencies between government policy commitments and domestic economic fundamentals. In other words, these models explain the inconsistency between the variables determining the exchange rate and the value fixed by the country's monetary authorities.

3.1 First-Generation Models: Fundamental Disequilibrium

The *first generation* of speculative attack models were mostly developed to explain the crisis in Latin America in the 1960s and 1970s. These models consider cases where the government is either unable or unwilling to correct inconsistencies between its exchange rate and other domestic policy goals. As these become more serious, a crisis eventually becomes inevitable. In these models, focus is on the fiscal and monetary causes of crises. And the governments are assumed to pursue fiscal and monetary policies that are inconsistent with maintaining their fixed or slowly adjusting pegged exchange rate regimes. Unsustainable money-financed fiscal deficits lead to BoP deficits and to a persistent loss of international reserves. And when the levels of reserves fall to a certain threshold, there is a sudden BoP crisis and ultimately a currency crisis (Krugman 1979). Walters (1986) gives another variant of a first-generation model, where the source of the speculative attack is a vulnerable anti-inflation policy that undermines the fixed exchange rate. Following Krugman, the basic mechanics can be shown using the open economy models. Incorporating a government sector alongside a private sector, the current account may be expressed as: $X - M = (S - T) + (T - G)$ where X is exports, M is imports, S is domestic saving, I is investment, T is tax revenue and G is government expenditure.

From this expression, it may be seen that if G increases relative to T, and there is no change in $(S - I)$, then $(X - M)$ will fall. If $S - I = 0$, then with $G > T$, it follows that $M > X$. With a pegged exchange rate, international reserves will decline in order to finance net imports. Fiscal deficits and their impact on reserves lie at the heart of the first-generation currency crisis model. In what follows, the first-generation model emphasizes domestic economic mismanagement in the form of fiscal deficits, monetary expansion and pegged exchange rates as the ultimate source of currency crisis. The Mexican crisis in 1976 is an example of first-generation crisis.

3.2 Second-Generation Crisis Models: The Role of Expectations

Second-generation crisis models, like those by Obstfeld (1986, 1994), tell a rather different story. It analyses cases where the above inconsistencies in the first-generation model place the economy in a 'zone of vulnerability' making a crisis possible but not inevitable. These models focus on investors' expectations and governments' conflicting policy objectives and predict that speculative attacks could occur when a country's fundamentals are nearly in a vulnerable zone. So these models demonstrate that exchange rate crises cannot be identified or predicted only with macroeconomic indicators. A speculative attack here leads to a change in economic policy that justifies a new, higher equilibrium value of the exchange rate. But then, in such circumstances, the speculative attack is self-fulfilling: The successful attack yields its own justification. In first-generation models, the success of the attack is a certainty; in the second-generation model, success only comes if the attack is self-fulfilling—through bringing about the policy change after the attack that the speculators expect. An example of a second-generation crisis is the 1992–1993 crisis in the European Monetary System (EMS).

Thus, the role of expectation is central to the second-generation model. Speculators do not cause second-generation crises. Rather, it is the inconsistency between internal and external targets at the pegged exchange rate—that is, the problem of fundamental disequilibrium—that is, the root cause of the crisis. And yet, given this fundamental disequilibrium, the action of speculators determines whether it will result in a crisis. The 'end game' can be different according to how speculators respond. A key element of the second-generation model is, therefore, the existence of a multiple equilibrium.

Finally, newer crisis models have focused on financial sector weaknesses and the role of policies in the emergence of currency crises. These explanations focus on fundamental economic disequilibrium such as large government budget deficits.

3.3 *Third-Generation Crises: Financial Sector Weaknesses*

Traditional macroeconomic models paid little attention to the financial sector and the crises of the 1990s demonstrated that this was a major mistake. The discovery of serious weaknesses in financial sectors can generate major changes in international capital flows, especially when countries have weak international liquidity position (i.e. high short-term foreign debt relative to international reserves). Such a situation can quickly turn into a run on the currency without requiring any outright speculation. Indeed, many of the flows of fund during the Mexican crisis of 1994 and Asian crisis of 1997–1998 were of this risk-covering nature.

This new wave of currency crises, the Mexican crisis in 1994 and the Asian crisis in 1997, fuelled a new variety of models—known as third-generation models (Krugman 1998) which are at best very much a capital account crisis model as well as moral hazard and imperfect information models. When it identifies fully with a capital account crisis model, fundamental to the model is an understanding of the factors that influence capital mobility, although it points to the importance of differences in interest rates (that is, return) and to risk, in the form of both default risk and exchange rate risk. For that reason, these models are a part of a framework of financial bubbles, moral hazard and lax fiscal policies that implicates both irresponsible fiscal authorities and irresponsible local banks. Speculative bubbles and financial crises are caused by the poor quality of the loans: The consequence is a run on the local currency.

In the case of the third-generation model, fiscal deficits and current account deficits may exist but, in fact, they are not of central importance. The model, therefore, differs from the first-generation model. In common with the second-generation model, there is no significant inconsistency between internal and external policy objectives. Rather, it is a narrowing in interest rate differentials, or a changed perception of default risk, that converts the boom of capital inflows into the bust of capital outflows, and it is this reversal that is at the heart of the crisis. In sum, this third-generation models resort to financial excess argument to explain financial crises. The process begins with large capital inflows increasing the loan capacity of the domestic banks, which, for their part, adopt non-recommended risk management practices that ultimately lead to a banking crisis.

After achieving perhaps the most rapid period of economic growth in human history, described by economists, the World Bank and commentators as representing a ‘growth miracle’, a group of countries in the East Asian regions experienced one of the history’s most precipitous declines in economic fortune.

A crisis in the Thai exchange rate market had been at least partially anticipated. But the contagion was almost entirely unanticipated. Most—if not all—of the affected countries like Indonesia, Malaysia and South Korea had enjoyed low inflation and robust economic growth for quite some time (Table 30.1), where none of the indices would indicate a crisis in the making and appeared to have strong enough economic fundamentals during the first half of the 1990s to preclude the types of crises predicted by first- and second-generation crisis models. In order to explain these events, economists developed ‘third generation’ of crisis models, which placed

Table 30.1 Growth and decline in Asia, 1996–1999. (Source: quoted in Godement 2002 and various other sources)

Crisis countries	GDP Growth				Unemployment (% of labour force)				
	1996	1997	1998	1999	1990–1996 (average)	1996	1997	1998	1999
Indonesia	8.0	5.0	– 13.2	0.2	2.2	–	4.7	15–20	28.5
Malaysia	8.6	8.0	– 7.5	5.4	–	2.6	2.6	3.9	3.0
The Philippines	5.5	5.1	– 0.5	3.2	–	7.6	–	8.0	–
South Korea	7.1	5.5	– 6.7	10.7	2.1	2.3	2.5	9.7	7.4
Thailand	5.5	– 0.4	– 10.4	4.2	2.5	1.5	1.0	6.1	5.9

a greater emphasis on the micro-foundations of currency crises. Among these refinements are the considerations of financial sector weakness, moral hazard, contagion and stock as well as flow disequilibrium.

When emphasis in these models are on moral hazard and imperfect information, they then focus on 'excessive' booms and busts in international lending and asset price bubbles. These models also link currency and banking crises, sometimes known as 'twin crises' (Kaminsky and Reinhart 1999). It is difficult to distinguish whether a financial crisis originates in a run on domestic banks or on the domestic currency. As a result, currency and banking crises can appear to occur simultaneously.

Problems in the banking sector typically precede a currency crisis; the currency crisis deepens the banking crisis, activating a vicious spiral; and the financial liberalization often precedes banking crises. For example, McKinnon and Pill (1995) examine the role of capital flows in an economy with an unregulated banking sector with deposit insurance and moral hazard problems of the banks. Capital inflows in such an environment can lead to over-lending cycles with consumption booms, real exchange rate appreciations, exaggerated current account deficits and booms (and later busts) in stocks and property markets. Importantly, the excess lending during the boom makes banks more prone to crisis when a recession unfolds.

Following the crisis in Argentina in 2001, the links between debt sustainability, sovereign defaults and currency crisis again attracted our attention.

Finally, currency crises have also been linked to the erratic behaviour of the international capital markets. For example, Calvo (1998) has brought to attention the possibility of liquidity crises in emerging markets due to sudden reversals in capital flows, in large part triggered by developments in the world financial centres.

To summarize, all models suggest that a crisis erupts in fragile economies. Importantly, the three generation models conclude that vulnerabilities come in different varieties.

In the empirical literature, Kaminsky (2006) by regression tree analysis catalogues crises into six classes:

1. Crisis with current account problems: This is associated with loss of competitiveness, i.e. real exchange rate appreciates.
2. Crisis with financial excesses: associated with booms in financial markets; types of crises that are preceded by the acceleration in the growth rate of domestic credit and other monetary aggregates; also associated with banking crises—twin crises

episodes—are the costliest. Not only does the domestic currency depreciate the most, but also output losses are higher and the reversal of current account deficit is attained via a dramatic fall in imports. In the aftermath of these crises, exports fail to grow. This evidence suggests that countries are unable even to attract trade credits to finance exports when its economy is mired in financial problems.
3. Crisis of sovereign debt problems: characterized by fragilities with 'unsustainable' foreign debt.
4. Crisis with fiscal deficits: This variety is related to experiencing fiscal policy.
5. Sudden-stop crisis: This type is only associated with reversals in capital flows triggered by sharp hikes in world interest rates, with no domestic vulnerabilities and adverse effects on the economy.
6. Self-fulfilling crisis: This class of crises is not associated with any evidence of vulnerability, domestic or external.

These estimates allow us to answer four important questions about crises:

1. Do crises occur in countries with sound fundamentals?
2. All crises are preceded by domestic or external vulnerabilities.
3. How important are sudden reversals in capital flows in triggering a crisis?
4. While many have stressed that the erratic behaviour of international capital markets is the main culprit in emerging markets' currency crises, only 2 % of the crises in developing countries are just triggered by sudden-stop problems. While sudden-stop problem does occur, reversals in capital flows mostly occur in the midst of multiple domestic vulnerabilities (Calvo et al. 2004).
5. Are crises different in emerging economies?
6. Crises in emerging markets are preceded by far more domestic vulnerabilities than those in industrial countries. Overall, 86 % of the crises in emerging economies are crises with multiple domestic vulnerabilities, while economic fragility characterizes only 50 % of the crises in mature markets.
7. Are some crises more costly than others?
8. It is a well-established fact that financial crises impose substantial costs on society. Many economists have emphasized output losses associated with crises. But these are not the only costs of crises. In the aftermath of crises, most countries lose access to international capital markets. In most cases, countries have to run on capital account surpluses to pay back their debt. Finally, the magnitude of the speculative attack is itself important. For example, large depreciations may cause adverse balance sheet effects on firms and governments when their liabilities are denominated in foreign currencies.

4 Financial Crises, Spillovers and Contagion

As noted in the previous section, a crisis in one country may act as the trigger for a crisis in another by widespread spillover and contagion effects—a situation in which a currency crisis in one country increases the probability of a financial crisis in another country.

One may identify three causal factors that contribute to episodes of contagion: 'sudden stops' in capital inflows (capital flight), sometimes interact with banking sectors to cause banking crises, 'herding' or 'informational cascades' (shift in expectations and confidence) and 'financial linkage' (asset prices). Fundamental-based theories of contagion focus on the role of *real shocks* that are correlated across countries. These shocks felt by a large number of countries could be in the form of changes in the terms of trade—the sudden rise or fall in a key export or import price, liquidity shocks or macroeconomic shocks, apart from real shocks that can also be spread through *trade linkages* and *financial linkages.*

On the other hand, there are three alternative theories of contagion that are belief based. The *first* is that contagion is *self-fulfilling.* A crisis in one country leads investors to believe that investors are going to flee to other countries, regardless of their economic fundamentals. If this happens, investors across a broad range of countries begin to withdraw their funds and capital flight spreads. The *second* belief-based channel of contagion is referred to as *herding*. When information is costly to investors, small investors have an incentive to follow the leaders. If the small investors see even one large investor with good reputation pulling back based on a crisis in a single country, regardless of their reasons for doing so, enough investors follow to trigger a speculative attack. The *third* channel is referred to as a *wake-up-call hypothesis,* a term coined by Morris Goldstein (1998, pp. 18–20). According to this, the other neighbouring countries share a similar fate with the crisis country: weak financial sectors, large external deficits, etc. The crisis then spreads because a collapse in one country alerts investors to the existence of deeper problems. Another channel of contagion is the *competitive dynamics of devaluation* operating through bilateral and third-country trade linkage, from which a country becomes susceptible to speculative attacks. For example, Thailand may have exported little to Indonesia and Malaysia, but these countries all sold into the same markets in other parts of the world. Thailand's devaluation, therefore, worsened the BoP of all its neighbours and competitors. Moreover, the lack of transparency of bank balance sheets heightened the difficulty of distinguishing good credit risks from bad ones. In this environment, bank runs can lead to systematic banking crises and spill contagiously across countries.

Let as now turn to real-life experiences of these contagious diseases.

In the aftermath of the December 1994 Mexican peso crisis, the large Latin American countries experienced varying degrees of volatility in their foreign exchange markets and decline in their equity markets. There are a variety of explanations for this 'contagion' phenomenon. First, contagion can be the result of common external shocks, with simultaneous crises triggered by a change in the external environment, such as in international interest rate. Second, portfolio rebalancing by investors or common bank lenders in response to developments in one country may affect other countries' access to funds. Third, crises may spread for reasons that cannot be accounted for by fundamentals. Such 'pure' contagion is usually attributed to 'herd behaviour' by investors. If investors lack complete information about the economic environment in which they invest, including the way borrowers use their funds and what their financial situations actually are, they may pay attention to other investors.

Such herding behaviour is most likely to occur if the behaviour of individual investors is viewed as revealing important information about borrowers' creditworthiness. Finally, crises can spread though financial and trade linkages. A concern in a world where economies are interconnected via trade and finance crises will not be contained in the countries in which they originate. For example, the global financial crisis of 2008–2009 unleashed in the US sub-prime mortgage market, after crossing the Atlantic and the Pacific Oceans hard hit the European ocean and lastly the Indian Ocean. In the case of a crisis, aggregate demand will fall in the crisis country. As a consequence, national income falls. Since imports depend on income ($M = \mathrm{mY}$), imports will also fall. This will be one way in which the crisis country deals with the crisis and reduces its current account deficit. However, the crisis country's imports will be other countries exports. Their $(X-M)$ schedules will shift down, causing both their current account to weaken and their national income to fall. The weakening in the current account may then make these countries more prone to crisis themselves. Country A is the country in which the crisis originates and Country B is the one that is 'infected' by the crisis through its trade with Country A. The potency of the contagion will depend on the size of the fall in national income in Country A's marginal propensity to import and the pattern of its trade; what proportion of Country A's imports from Country B?

In addition to the trade route described above, contagion may occur in other ways. Devaluation in the initial crisis country, which is designed to strengthen its competitiveness vis-à-vis other countries will, if successful, clearly weaken the competitiveness of these other countries. This is another reason why the $(X-M)$ schedule in infected countries will shift down, creating current account deficit and economic recession. But contagion may also occur via the capital account. The response to the crisis in the initiating country is likely to raise domestic interest rates. If this policy is effective in attracting foreign capital, the capital will, by definition, be leaving other economies. The capital inflow to the original crisis country will be a capital outflow from elsewhere. The countries experiencing the capital outflow may now encounter their own capital account crises.

5 Global Financial Crises

5.1 The Latin American Debt Crisis

The decade of the 1980s in Latin America began with a crisis (Table 30.2) that included worsening terms of trade of debtor nations, a rapid appreciation of the US dollar (in such a way that bankers and borrowers had not anticipated this surge in the dollar), an unprecedented increase in real interest rates (after a switch to anti-inflationary monetary policy in October 1979, with the prime rate topping 20 %, thus making the payment of interest difficult for many borrowers, and the repayment of principal just about impossible), a high 27 % decline in non-oil commodity prices and the use of debt for subsidizing consumption rather than investment and

Table 30.2 Financial crises in the 1980s. (Source: compiled by the author)

Financial crises	Consequences	Spillovers and contagion effects
The Latin American debt crisis, 1982–1989	It reached a head in 1982, when several Latin American borrowers were unable to meet scheduled payments. The much-feared run on banks and financial panic as well as runaway inflation did not occur in the 1980s because international institutions and private banks cooperated to provide credits. The 1980s ended without widespread bank failures and financial chaos. But Latin American countries, such as Mexico, Argentina, Peru, when they liberalized in the 1980s, allowed their exchange rate to appreciate which damaged the trade balance and impacted negativity on growth	Nations have come to recognize that failure of other countries' banks will spill over to their own banks and economic setbacks among their customers will hurt their own firms that supply these customers
Mexican debt crisis following default and devaluation in August 1982	When the crisis was over, most Latin American countries had devalued their currencies and defaulted on the foreign debts. The debt crisis was followed by a decade of negative growth and isolation of international capital markets. The output costs of this crisis were so large that the 1980s became known as the 'lost decade' for Latin America	This Mexican crisis spread rapidly to all Latin American countries

removing these subsidies became difficult due to political pressure, and so borrowing continued. These increased borrowing costs and reduced export earnings for many developing countries, including Mexico. On 12 August 1982, in the face of capital flight, the Mexican finance minister Jesus Silva Herzog, telephoned the US secretary to the treasury Donald Regan, the US federal reserve chairman Paul Volcker and the managing director of the IMF Jacques de Larosiere to let them know that Mexico had almost run out of foreign currency reserves and could no longer make payments on around US$100 billion external debt it had taken from US and European banks, not even its interest payments on this massive debt obligations. By November, Brazil, similarly, had given up on repaying its own US$90 billion of debt, and other debtor states, including Argentina, the third biggest Latin American borrower, and the Philippines soon followed suit. This was a global banking crisis. Many large Western banks were threatened with insolvency if their loans to emerging markets were written off.

External debt in Latin America had quadrupled from a level of US$75 billion in 1975 to more than US$315 billion by 1983, which was about 50 % of gross domestic product (GDP) of that region and the ratio of debt to exports from around 200–300 % during the period from 1973 to 1983. Debt service (interest payment and the repayment of principal) grew even faster, rising to US$66 billion by 1982 from US$12 billion in 1975 (Grupen 1986).

The excessive private lending to Latin America has its costs in the oil price dynamics of the early 1970s. How did this crisis arise? Just as today's banking and credit crises has its roots in global current account balances, the problem of Latin American debt was rooted in the 'recycling' of the surpluses of oil-exporting countries of the Middle East, such as Saudi Arabia. In 1973, following restrictions on output agreed by the Organization of the Petroleum Exporting Countries (OPEC), the world price of oil jumped. This phenomenon, combined with the fact that oil bills were paid almost exclusively in dollars, resulted in large inflows of dollars into oil-producing countries. The only choice for OPEC countries was to deposit dollars mostly in US and Japanese banks. These massive deposits became known as petrodollars. The stage was set for a massive lending spree from private banks to Latin American countries. Most scholars agree that there was significant over-lending during the period from 1973 to 1982. The important question is why.

On the supply side of the loanable funds market, the answer is: Eurodollar deposits dramatically increased banks' desire to lend. On the demand side of the loanable funds market, Latin American countries exhibited considerable eagerness to borrow. Even the oil-exporting developing countries in Latin America, Mexico and Venezuela in particular saw the availability of private bank credit as an opportunity for industrialization backed by state leadership, after avoiding the conditionality of IMF for BoP financing.

The financial condition of Latin America was so bad that they had to borrow. Inflation had skyrocketed after the 1973 oil price hikes. There was negative real rate of interest.

The stage had been set for a major crisis when another oil crisis hit in 1979—though the intensity of oil-price rise was much less as compared to 1973. The effects of 1979 had been ‘hyper’ stagflation (up to five-digit inflation together with double-figure GDP declines): severe recession coupled with accelerating inflation. Meanwhile, lenders became increasingly hesitant to give new loans. The 1980s ended with only four countries (Chile, Columbia, Costa Rica and Uruguay) having recovered macro-balance. In the other parts, despite many years of costly adjustment, hyper-stagflation and persistence imbalances were the rule.

Despite the efforts of the IMF to effectively address these Latin American crises, international commercial banks began to withdraw credit from many of the developing countries of the world, and the debt crisis became global. Within a few years of the outbreak of these crises, the phenomenon of net capital outflows appeared in which the capital account payments of debtor countries exceeded their capital account receipts. Poverty increased substantially, and much of the developing world, particularly Latin America, entered what came to be known as the *lost decade*.

The debt crisis led to an abandonment of import substitution industrialization policies and a move towards trade and financial liberalization policies. Proponents of neo-liberalization view debt crisis as a necessary lesson for policymakers about the inefficiency of the state intervention policies.

Opponents of neo-liberalism view the debt crisis as a tragedy. Regardless of which view one takes, there is little dispute that the Latin American debt crisis triggered an era of neo-liberalism reform.

5.2 Italian Crisis of 1992

As a member of the EMS, Italy had pegged its exchange rate to the deutschmark. Although its inflation rate was far below triple digits, it was still above the inflation rates of most of the partners in the EMS, and in particular higher than Germany. The resulting disparity eventually led to a currency crisis in 1992.

The collapse of the BW system in the 1970s provides an example of an exchange rate system failing due to budget deficits in the USA. For domestic political reasons, the US government delayed tax increases, required to pay for the large increases in expenditures associated with the Vietnam War. Seeking to prevent interest rates from escalating, the Federal Reserve Board financed a sizable portion of its resulting budget deficits with monetary accommodation. The consequent overheating of the US economy led to increasing BoP deficits and, ultimately, to the currency crisis that was the final straw leading to the widespread abandonment of the BW regime of pegged exchange rates.

5.3 1992 EMS Crisis

A currency can only come under attack if the exchange rate regimes in question is neither fully flexible nor really fixed, meaning, thereby, that all types of exchange regimes are in between, such as crawling pegs, managed floating, target zone around a central parity and so on are prone to speculative attacks.

In the late summer of 1992, several currencies of EMS member countries came under speculative pressure: The Italian lira had to be devalued by 7 % against its EMS partner currencies on 13 September. On 17 September, the Italians and the British suspended their membership in the ERM.

The two proximate causes of this remarkable crisis were the reunification of Germany, which led to a large increase in German government expenditures, as well as an accelerating German domestic demand. As German policymakers were unwilling to offset these higher spending with higher taxes, a large budget deficit emerged. In response, the Bundesbank, sticking to its monetary orthodoxy, substantially increased interest rates to 8.75 % on 16 July 1992, which contributed to the strengthening of the German mark against the remaining EMS currencies. Since Germany was the EMS centre country, the logic of the currency system required all other EMS members to tighten monetary policy as well. The British government, which was at that time fighting a strong recession at home, was unwilling to follow the restrictive monetary and interest rate policy of the Deutsche Bundesbank as a consequence of its own exchange rate target in the ERM. More powerful is the argument that not only the pound and the lira but also the Spanish peseta and the Portuguese escudo were considerably ‘overvalued’. This generated crises that led to a breakdown of the EMS system.

The ERM crisis of September 1992 did spread to other European currencies. It not only involved countries from the centre or the south of Europe but also Nordic countries. The Finnish markka was attacked by speculators. The attack triggered speculation against the Swedish krona and then it spread to the Norwegian and the Danish krone.

5.4 A New EMS Crisis of the French Frank

A new EMS crisis erupted on a Monday in August 1993, this time involving the French Franc. In concert with the Bundesbank, the Bank of France intervened massively in the foreign exchange market to put an end to the speculative attack. Bundesbank came forward and sold more than US$35 billion worth of marks in support of frank.

In what follows is that while the crisis in the ERM of the EMS dominated the early 1990s, the rest of the decade and the first half of the 2000s were more associated with currency and financial crises in a number of emerging economies. The sequence of crises began in Mexico in 1994–1995, followed by further crises in East Asia in 1997–1998, Russia and Brazil at the end of the 1990s and Argentina and Turkey at the beginning of the 2000s, and they introduced new elements of instability in the global economy, largely the result of many countries liberalizing their BoP capital accounts and thus becoming vulnerable to massive short-run inflows and outflows of short-term capital (the so-called hot money), all of which are discussed in the following.

5.5 *Mexican Tequila Crisis 1994–1995: A Second-Generation Crisis*

The Latin American debt crisis of the early 1980s had been caused by public sector overspending. But in 1994, something new happened. A major financial crisis, caused by the outflow of private capital, of the kind which had brought down the BWs in 1971 and the EMS in 1991, happened in Mexico. The Mexican crisis was different from the Latin American turmoil of the 1980s in that it was set off not just by fundamental weaknesses, such as sustainable fiscal deficits, in conjunction with high domestic consumption, monetary expansion, relatively rapid inflation, an appreciating real exchange rate, current account deficits, a reluctance to devalue and a fated attempt to ride out the storm by running down reserves but also by the currency mistakes on the public sector balance sheet (Calvo and Mendoza 1996). These caused a 'second-generation crisis' in the form of self-fulfilling currency run. This crisis presented new challenges for the IMF.

The country had successfully brought down inflation from triple- to single-digit levels, but the combination of strong capital inflows and a slow rate of depreciation designed to limit domestic wage increases had resulted in a large current account deficit and a substantially appreciated real exchange rate. Thus, although Mexico's domestic fundamentals were strong, the economy was vulnerable to a drop in capital inflows and negative as with the Brazilian crisis to severe shocks to political stability.

A series of crises punctured the generally strong flows of international lending to developing countries since 1990. The first of these struck Mexico in December 1994.

In the period from 1990 to 1994, Mexico received large capital inflows following the pattern of uncovered interest rate parity[1] worth US$95 billion, of which the portfolio investment into high-yielding tesobonos were US$43 billion and foreign direct investment (FDI) and equity investment were only US$24 billion and US$24 billion, respectively, as investors sought and anticipated high returns on their investments in Mexico and were impressed with Mexico's economic reforms, combined with trade and financial liberalization, and its entry into the North American Free Trade Area (NAFTA) and also into the Organisation for Economic Cooperation and Development (OECD). This at best means that the value of peso within the fluctuating band maintained with the dollar was consistently high. But the use of the exchange rate

---

[1] Since 2000, carry trade has become one of the important phenomena. This is a speculation based on borrowing funds in a country with a low interest rate and investing the funds in a country with a higher interest rate. As long as the exchange rate for the two currencies involved in carry trade is not fixed irrevocably, the carry trader bears exchange rate risk fully according to the equation: $i = i^* + (E^e - E)\,/E.100$, where i and i^* represent the domestic and the foreign interest rates (in percentages). E^e is the expected exchange rate, while E is the exchange rate on the spot exchange market in terms of the domestic currency price of the foreign currency. In the event of carry trade, the Left Hand Side (LHS) expresses the borrowing costs, while the Right Hand Side (RHS) shows the return on lending, including the expected exchange rate profit (which can, of course, be negative). One of the most popular channels for carry trade in the past few years has been borrowing in Japan (with quite low interest rate) and lending in Australia (with relative high interest rate). In the summer of 2007, when Japan increased its interest rate, part of the carry trade was stopped.

as an anchor to curb inflation has had an inherent inconsistency. As a result, the country's competitive position deteriorates, causing a continuous worsening of the current account.

Before the exchange rate crisis, Mexico's government debt consisted mainly of short-term, peso-denominated bonds. But strains arose. The real exchange rate value of the peso increased, because the government permitted only slow nominal peso depreciation, while the Mexican inflation rate was higher than that of the USA, its main trading partner. The current account deficit increased to 6 % of Mexico's GDP in 1993, although this was readily financed by the capital inflows. Mexico's banking system was rather weak in this era of capital account liberalization, with inadequate bank supervision and regulation by the government. With the capital inflows adding funds to the Mexican banking system, bank lending grew rapidly, as did defaults on these loans. The year 1994 was an election year with some turmoil, including a guerrilla uprising in the Chiapas region in January and two political assassinations in March 1994. One such event was the assassination of the ruling party's presidential candidate, Colosia. Following a series of political disappointments, the deputy attorney general, Massieu, resigned in November 1994. At the end of the year, anti-government riots broke out again among the Indians in Chiapas. As a result, foreigners began to withdraw invested capital for the first time. The peso came under some downward pressure. The government used sterilized intervention to defend its exchange rate value, so its holdings of official international reserves fell from a level of US$30 billion to a new level of US$18 billion. On 20 December 1994, in spite of the government's decision to devalue a 15 % of the new peso's band ceiling, no tangible result occurred with the pressure of the currency still continuing. For Mexico, 21 December 1994 is a watershed date because its central bank lost US$6 billion on that very day in going for an attempt to stabilize the exchange rate. One day later, the authorities switched to free floating after abolishing the intervention band. Another 15 % devaluation of the new peso instantly followed. By the end of 1994, the new peso had depreciated by no less than 71 % for the year as a whole. By December 1994, there was a massive reversal of capital flows and the peso plummeted. The consequences for Mexico were as follows: Inflation rose from 7 % in 1994 to 357 % in 1995 and GDP fell by 6.2 % in 1995 compared with a growth rate of 4.4 % in the preceding year.

In the middle of 1994, the ratio of M_2 to reserves had a value of seven, and this rose to ten in November. In April, the foreign currency part of M_3 was already larger than the amount of official reserves. When M_3 'headed for Miami'—and so became flight capital—in late 1994, available official reserves proved unable to cover the country's liabilities, and the currency crisis became a reality.

The US government became worried about the political and economic effects of the financial crisis in Mexico, and it arranged a large rescue package that permitted the Mexican government to borrow up to US$50 billion, mostly from the US government and the IMF. The Mexican government did borrow about US$27 billion and used the money to pay off *tesobonos* (short-term dollar-indexed government liabilities), as they matured, and to replenish its official reserve holdings. The currency depreciation and the financial turmoil caused rapid and painful adjustments in Mexico.

As the rescue took hold, the pure contagion that led investors to retreat from nearly all lending to developing countries calmed after the first quarter of 1995. The adverse tequila effect lingered for a smaller number of countries, as investors continued to pull out of Argentina, Brazil and, to lesser extents, Venezuela and the Philippines. Still, much of the Mexican financial crisis of 1994–1995 was resolved quickly.

The Mexican crisis began to raise questions concerning the design and structure of the international monetary system. In fact, a theme throughout the 1990s was that fundamental reform was needed to better deal with excessive international capital mobility.

5.6 Japanese Stagnation and the Asian Financial Crisis 1997–1998

With the bursting of asset bubble in the early 1990s, the Japanese economy entered into a period of prolonged recession—in fact, the longest in its modern history. In the meantime, some East and SouthEast Asian economies (Thailand, Malaysia, Indonesia, Korea and Hong Kong) experienced major financial crises in 1997. The difficulties arose from the following events.

First, the private sector had to rely on borrowing, rather than equity issuance, to raise investment funds. As a result, firms became highly leveraged, but banks continued to lend because they were underpinned by implicit government guarantees. When growth slowed, as it first did in Thailand in 1996, and then in East Asian economies, these banks were exposed to the inability of borrowers to repay loans.

Second, a further difficulty arose, as so many times before, from the existence of fixed exchange rate systems in some East Asian economies, but with a new twist. Banks financed much of their domestic corporate lending by borrowing in foreign exchange from abroad, often at shorter maturities. Very little of this borrowing was hedged, as a result of the implicit guarantee on the exchange rate. The financial sector was already in trouble after the initial slowdown in growth in 1996. Currencies fell in mid-to-late 1997 because of the foreign investors' concerns about these difficulties; as a consequence, widespread bankruptcies and potential bank failures loomed large because of the unhedged foreign currency obligations. Fear grew that fiscal systems would be unable to bear the cost of large-scale bank rescues.

The East Asian debacle marked the advent of 'third-generation' crises in which currency crises and banking crises were intimately intertwined. Unlike the earlier Latin American debt crisis, or even the Mexico crisis in the 1980s and 1990s, fiscal profligacy and macroeconomic crises played no explicit part in the East Asian crisis. In East Asia, most of the important macroeconomic indicators were generally healthy. The fiscal balance was in surplus, inflation was low and domestic savings and investment as a proportion of GDP were among the highest in the world; there were well-developed stock markets, but bond and other security markets were underdeveloped along with the lack of transparency in the banks' operations. There had been for some time, however, a major imbalance in the external accounts of the

Asian countries. No country in the long run can grow faster than the rate compatible with equilibrium on the current account of the BoP unless it can finance ever-growing deficits—which, in general, it cannot. The root cause of the crisis is well documented in Table 30.3.

Early first-generation model predicts that ongoing fiscal deficits, rising debt levels or falling reserves precede the collapse of the fixed exchange rate regime. This prediction is inconsistent with the 1997 Asian currency crisis. This inconsistency led many observers to discuss the fiscal explanations of this crisis. However, some economists show that bad news about prospective deficits can trigger a currency crisis. Under these circumstances, a currency crisis will not be preceded by persistent fiscal deficits, rising debt levels or falling reserves. These models assume that agents receive news that the banking sector is falling and that banks will be bailed out by the government. The government plans to finance the bank bailout by printing money, beginning at some time in future. Burnside et al. (2001) show that a currency crisis will occur before the government actually starts to print money. Therefore, in their model, a currency crisis is not preceded by movement in standard macroeconomic fundamentals, such as fiscal deficits and money growth. They argue that their model accounts for the main characteristics of the Asian currency crisis.

This explanation of the Asian currency crisis stresses the link between future deficits and current movement in exchange rate. Using the fiscal theory of the price level, Corsetti and Mackowiak (2006) show that prices and the exchange rates jump in response to news about future fiscal deficits.

Moral hazard: The combination of moral hazard and weak risk management and regulatory systems had led to money ill-advised loans, so that in a number of countries the financial sectors were suffering from serious solvency and liquidity problems. Such interaction came in huge quantity as a result of capital account liberalization (CAL).

In July 1997, the already overvalued Thai baht came under speculative attack. This crisis served as a wake-up call to international financial markets, alerting them that Thailand's unhedged fungi foreign borrowing was much riskier than many had believed. This, in turn, prompted fungi foreign investors to reassess the vulnerability of Thai investments to financial sector risks.

Among the Southeast Asian countries, the experience of the financial crisis in 1997–1998 in Malaysia as well as in other affected countries (particularly Indonesia) demonstrates the fact that the implementation of neo-liberal version of globalization, particularly financial liberalization, has brought about widespread hardship among the disadvantaged groups in these countries and caused political and social turmoil. Policy prescriptions from the IMF, which involved tight monetary policies, worsened the suffering of those adversely affected by the crisis.

5.7 *Russian Rouble Crisis, 1998: First-Generation Example*

Russia weathered the Asian crisis in 1997 reasonably well, but its underlying fundamental position was remarkably weak. It had a large fiscal budget deficit, and

Table 30.3 Financial crises in the 1990s. (Source: compiled by the author)

Financial crises	Consequences	Spillovers and contagion effects
The 1992–1993 EMS crisis (centred in Germany)	In the summers of 1993, by the end of this crisis, the lira and the sterling had been driven from the ERM. Finland, Norway and Sweden had abandoned their unofficial peg to the European Currency Unit (ECU); the Spanish peseta, the Portuguese escudo and the Irish pound had devalued; and Europe's central bank governors and finance ministers had widened the ERM's intervention margins to + 15 % from + 2.25 %. Only then did the currency market stabilize	In 1992, speculation by currency traders caused a run on the pound and the lira, destroying the fledgling European Monetary System (EMS)
1994–1995 Mexican peso crisis (the so-called Tequila crisis)	Mexico underwent a *second* crisis in late 1994 and early 1995 Peso depreciated by 71 %. The forex reserve fell. Peso and banking system collapsed The authorities had to abandon its pegged exchange rate for the peso at the end of 1994 and switched to free floating from the intervention band	Currency jitters spread around the Latin American regions under the new name Tequila effect. Argentina was not only hard hit by the Mexican turbulence but also thrown into a full-fledged economic crisis
Asian currency and banking crisis, 2 July 1997 (the so-called Asian contagion)	The Central Bank of Thailand, after spending some US$9 billion, gave up its futile attempts to support the baht and protect its peg to the US dollar. The newly 'floating' baht promptly collapsed, losing about a fifth of its value. Following the massive depreciation, there was massive speculative selling of the currency by the investors, anticipating that the pegged rate could not be maintained. Very quickly, the 'contagion' spread to other Asian countries	Within a few days, the crises had spread from Thailand to Indonesia, Malaysia and the Philippines, Singapore and Hong Kong Special Administration Region and finally the Republic of Korea, thus plunging them into financial turmoil
	Poor domestic financial structure as witnessed by the high percentage of non-performing bank loans (15–20 %), regulation and imprudent lending, worsening terms of trade, corporate weakness, cronyism or nepotism as well as unsustainable BoP deficits financed by	These shock waves within a matter of 18 months, as one currency after another came under speculative attack, spread outright from the region. Russia, the Ukraine and the Brazil, Argentina, Mexico all saw their currencies come under

Table 30.3 (continued)

Financial crises	Consequences	Spillovers and contagion effects
	short-term, volatile capital inflows, made possible by financial liberalization, while at the same time East Asian authorities' endeavour to maintain a fixed rate of the currencies were the root causes of the financial crisis. The rapid reversal of these capital inflows (mostly portfolio investment and private lending of various types—rose from US$36 billion in 1994 to US$86 billion in 1996—became one of the driving forces behind the financial crisis. Nay, more importantly, the crisis arose from inherent problems in the global financial system: people taking excessive risks but with inadequate capital. Also, the crisis was driven fully by capital account The grave consequences of the crisis included: a slowdown of international trade, falls in output, accompanied by bankruptcies and rise in unemployment, sharp rise in prices, cuts in public expenditure and erosion of the social fabric	speculative attack. In each of the cases, sharp depreciations of the currencies resulted. All were forced to raise interest rates in an attempt to prevent further depreciation of their exchange rates
The Russian rouble crisis (default and devaluation on 1 August 1998)		The case of 'pure contagion': Turmoil in the forex market following Thai's flue heightened in August 1998 in Russia resulting in default and devaluation. Thereafter, the crisis spread around the world with speculative attacks in economies as far apart as South Africa, Brazil and Hong Kong
Brazilian crisis (January 1999)	In January 1999, there had been devaluation of the Brazilian real	One can note the lack of contagion accompanying 1999 devaluation of the Brazilian real

government borrowing led to rapid increases in government debt to both domestic and foreign lenders. In mid-1998, lenders balked at buying still more Russian government debt. In July 1998, the IMF organized a lending package under which the Russian government could borrow up to US$23 billion, and the IMF made the first loan of US$5 billion. However, the Russian government failed to enact policy changes included as conditions for the loan. With substantial debt service due on government debt during the second half of 1998, investor confidence declined, with selling pressure driving down Russian stock and bond prices.

In August 1998, the Russian government announced drastic measures. The government unilaterally 'restructured' its rouble-denominated debt, effectively wiping out most of the creditors' value. It placed a 90-day moratorium on payments of many foreign currency obligations of banks and other private firms, a move designed to protect Russian banks. And it allowed the rouble to depreciate by shifting to a floating exchange rate. Russia requested the next instalment of its loan from the IMF, but the IMF refused, because the government had not met the conditions for fiscal reforms.

Foreign lenders were in shock. They had expected that the IMF rescue package would provide Russia with the funds to repay them. They reassessed the risk of investments in all emerging markets and rapidly sought to reduce their investments. The reversal of international bank lending and stock and bond investing in 1998 led to the first decline in net long-term financial flows to developing countries since the mid-1980s.

The fall of the Berlin Wall in 1989 and the dissolution of the Soviet Union in 1991 enabled the IMF at last to become a universal institution. Although the IMF can claim credit for helping Russia to install some monetary discipline by the mid-1990s, the process took time: FDI remained low, tax collection was poor and the fiscal deficit remained large. Growth in real GDP did re-emerge by 1997. But following the onset of the 'East Asian crisis', the rouble came under speculative attack in November 1997. Pressure on the rouble was compounded by foreign investors' attempt to hedge their rouble holdings as well as by a drop in the price of oil, which accounted for about one third of Russian foreign exchange inflows. Oil prices fell to half due to decreased demand for oil in Asia. The exchange rate value of the rouble came under severe pressure as capital flight by wealthy Russians led to large sales of roubles for foreign currencies.

As a result, the Yeltsin government, dependent as it was on oil revenues to fund its fiscal programmes, found its treasury rapidly being emptied. Russia sought additional IMF financing in early 1998. Agreement on the terms of a new programme could not be reached owing, in part, to a failure by the Russian authorities to secure an increase in fiscal revenue. As a result, foreign investors began to unload Russian assets and about US$4 billion fled the country in the summer of 1998. By the time, additional IMF financing was agreed in July 1998, but fears of devaluation led to such a pronounced sell-off of Russian securities that the authorities were forced to devalue the rouble and halt payments on both domestic and foreign debt.

After the IMF negotiations failed, the Russian legislature, the Duma, removed the most critical components of Boris Yeltsin's proposed fiscal reform plan. However, it became clear that Russia was not going to be able to adjust its policy enough to avoid a fiscal crisis. In August 1998, there was a several speculative attack on the rouble, and the Russian government announced that it would no longer support the crawling peg and would default on its foreign-held debt. The remaining investors immediately and frantically sought to divest themselves of the Russian liabilities and the rouble plummeted. The government played a passive role in the first-generation model .

5.8 Brazil Crisis in 1999

Political considerations can play a role in the outbreak of a crisis—for example, Brazil 1999. Brazil was among the countries hit hard by the fallout from the Russian crisis. In November 1998, the IMF organized a package that allowed the Brazilian government to borrow up to US$41 billion (US$18 billion actually borrowed), in an effort to allow Brazil to fight pressures pushing towards a crisis. Brazil had a large current account deficit, and the government was defending its crawling exchange rate with intervention and high domestic interest rates. However, the government failed to enact the fiscal reforms called for in the IMF loan, and capital outflows increased.

In January 1999, the Brazilian government ended its pegged exchange rate, and the real depreciated. However, this situation did not escalate into a full crisis because the problems did not spread to the Brazilian banking system, which was sound and well regulated. By April 1999, Brazil and other developing countries were able to issue new bonds to foreign investors. More generally, the market prices of emerging market financial assets began to increase, although the net capital flows to developing countries remained lower than they had been in 1997.

5.9 Turkish Crisis, 2001

Turkey's economy and its government policies had been problematic for decades, and it had borrowed from the IMF continually since 1958. In January 2000, Turkey entered into another borrowing programme of US$10 billion from the IMF and committed to reduce its inflation rate (which had been close to 100 % for a number of years), improve its regulation of the banking system and close failing banks, privatize state-owned businesses, end various subsidies and reduce its fiscal deficit. As part of the inflation fight, Turkey adopted a crawling exchange rate (pegged to a basket of the euro and the US dollar).

The announcement of the new programme brought large capital inflows. However, bank regulation and supervision remained weak. Turkish banks took on substantial additional exposure to exchange rate risk, as they borrowed foreign currencies at low interest rates and converted the funds into liras to invest in high-interest Turkish government bonds. The country grew quickly, and inflation was lowered below 50 %, but the fiscal deficit remained high, and the current account deficit widened to about 5 % of GDP. November brought the first signs of new trouble, and foreign lenders began to pull back. The Turkish government used a large amount of its official reserves to defend the pegged exchange rate. December brought new pressures as several prominent bankers were arrested. Overnight, interest rates rose to an annual rate of nearly 2,000 % to stem the capital outflows. A new IMF programme promised additional loans of up to US$7.5 billion during the next year.

After calming for a while, conditions deteriorated again in February 2001, because of legislative delays and political fighting between the President and Prime Minister about reforms. Overnight, interest rates again went into quadruple digits, and the government again used up a large amount of official reserves defending the pegged exchange rate. Then the government gave up, and the lira lost a third of its value in 2 days. Turkey's banks incurred large losses. The Turkish economy experienced a severe recession, with real GDP declining by 9 % during 2001. Turkey entered into yet another IMF programme in May 2001.

5.10 Argentina's Crisis, 2001–2002: A Second-Generation Crisis

In the 1980s, Argentina's economy was in a mess. Argentina suffered from a high rate of inflation that appeared to be a permanent feature of the economy. The government tried several stabilization programmes, but these had all failed, and, at the end of the 1980s, hyperinflation had reached a rate of around 10,000 %.

After a sustained period of hyperinflation in the 1980s, Argentina then decided in April 1991 to peg its currency, the peso, to the US dollar under a quasi currency board regime at a one-to-one party (the so-called convertibility law). Although the IMF cautioned that Argentina had neither the fiscal discipline nor the robust export sector needed to sustain such a system, Argentina supported their macroeconomic programme under a series of lending arrangements.

In the first few years of its operation, the Argentine currency board achieved considerable level of success. And Argentina was widely hailed as a model of successful economic reform as the rate of inflation fell to single digits and growth increased. However, not all of the indicators were favourable. Between 1990 and 1994, the Argentine current account changed from a surplus of 3.5 % of GDP to a deficit of 4 % of GDP.

The year 1995 proved to be a dramatic one for the Argentine economy. The country experienced a substantial contagion from the Mexican crisis at the end of 1994 with capital outflow along with high inflation rate reaching 17 % in 1995 and a high unemployment rate. Another negative development for Argentine economy was that between 1996 and mid-1999 the real exchange rate worsened by 20 %, thus making its goods uncompetitive in regional and international markets. As a result, Argentina chose to borrow substantial amount in US dollars. Brazil's decision to float the real in 1999 in response to pressure from the Russian crisis made it even harder for Argentina to compete under its quasi currency board regime. Meanwhile, the Argentine people began to fear for the continuation of the fixed exchange rate and the soundness of the banking system. In response to depositor runs on banks, the government closed the banks in November. When the banks reopened in December, withdrawals were severely limited. Angry protest spawned looting and rioting, with 23 deaths. The country's President resigned, and Argentina then had four new Presidents in 2 weeks.

In September 2001, the IMF made an unusually large disbursement of US$6 billion to Argentina, but it was to be the last. The IMF refused to make additional loans under the rescue package because the government had not met the conditions set by the Fund for improvements in government policies.

The Argentine authorities allowed the peso to float in January 2002, and it quickly collapsed from parity with the US dollar to an exchange rate of nearly 3.9 to the dollar in June 2002 and the peso lost about 75 % of its value in the first 6 months of the year. The government defaulted on about US$140 billion of its debt, much of it owed to foreigners, the largest default ever. In addition, the peso depreciation caused huge losses in the banks because of some mismatch of dollar liabilities and dollar assets, and especially because of the terms under which the government mandated the conversion of dollar assets and liabilities into pesos. A number of banks closed, and the banking system was nearly non-functional. During 2002, real GDP declined by 11 %, inflation reignited, the government defaulted on its debt and the banking system was largely paralysed. IMF did not play a master role. Ultimately, in the period between the end of 2001 and early 2002, the Argentine currency board, which had remained in place for more than a decade, finally broke down.

At first, it appeared that Argentina's collapse would have few effects on other developing countries, since it had been widely expected. But after a few months, Argentina's problems did spread to its neighbours, especially Uruguay. Uruguay relied on Argentina for tourisms and banking business. The tourism dried up, and Argentine withdrawals from their Uruguayan accounts increased. After its holdings of official reserves plummeted defending Uruguay's crawling pegged exchange rate, the Uruguay government floated its currency in June; within 2 weeks, the currency had fallen by half. In August, Uruguay received an IMF rescue package and used it to stabilize its financial situation. Still, it suffered a severe recession, with real GDP declining by more than 10 % during the year.

5.11 Dotcom Bubble Bursting, 2002

In the late 1990s, the share-price-to-earnings ratio of the US-listed companies, especially IT companies, underwent a surge that was driven by what some observers at the time referred to as 'irrational exuberance'. This trend suddenly reversed in late 2000 as investors realized that their wild expectations were not being fulfilled and prices began to plummet. Overall, price-to-earnings ratios went from an average of 20:1 in 1995 to 44:1 in 2000 and back to 22:1 in 2003–2006—one of the purest manifestations of bubble dynamics that recent history has produced. The fallout caused the US economy to tip into recession in early 2001.

5.12 Global Financial Crisis (2008–2009)

Of the 200 or so financial crises since the late 1970s, the most far-reaching ones had occurred in the past 18 years following the Mexican (1994–1995) crisis and its reparations. The world has been further alarmed by the Asian crisis (1997)—which spread to Russia and Brazil (1998) and the Dot-com bust (2001). The ongoing crisis unleashed with the sub-prime mortgage crises in 2007 and the subsequent failures of large financial institutes in the USA and elsewhere, the 2008–2009 crisis developed rapidly into a global credit crisis, deflation and reductions in international trade. And it is more central and serious than any of the previous ones that the experts do not hesitate to say that. It is the worst economic crisis capitalism had faced since the depression of the 1930s. For the first time everybody, from the richest person in the richest city to the poorest person in the poorest slum, was affected by the crisis and although its roots are global, its impact was local, directly felt on nearly every high street, on nearly every shop floor, around nearly every kitchen table.

The US banks failed as the result of the households defaulting on mortgages that, in an increasingly deregulated financial market, were able to grow unchecked because banks relied on financial innovations that allowed them to repackage the relevant securities in such a manner that the new bundles looked safer than the original loans to their acquirers. When the fraud came to light and the banks failed, the confidence of consumers and businesses, which was already shaken, finally collapsed, and they sought protection by avoiding consumption and investment. Besides, as banks lost confidence, a credit crunch materialized. In consequence, aggregate demand plunged vertically everywhere, and the turmoil, which was at first limited to the banking industry, became a global crisis. In the USA alone, 22 banks collapsed in 2008 and 77 more by August 2009. The crisis has also involved major investment fund failures, sharp declines in stock indexes and large reductions in the market value of the commodities and housing worldwide (Krugman 2008). According to the IMF estimates, from April 2009, total global output in 2009 declined by 1.3 % when measured in terms of purchasing power parity (PPP) and while per capita output declined drastically by 2.5 % in PPP terms and 3.68 % in market rate terms. Moreover, these developments are unequal. Overall, the advanced economies contracted significantly and negative growth was found in middle-income central and eastern European countries, the ones that were more penalized for having not learnt the lessons of the 1990s and incurred high current account deficits, that is, they insisted on adopting the growth with foreign savings policy, while contradictorily the immediate consequence of the crisis in emerging and developing countries was a sharp devaluation of their currencies in relation to the dollar with economic growth increasing by a modest rate of 1.6 %. In some poor regions, this means negative per capita growth. Unemployment rates were showing everywhere. The decoupling thesis popular in financial markets before October 2008 lost credibility as developing countries saw their exchange rates sharply devalue, their commodity prices fall, their local stock market bubbles burst and the first signs of investment plans (Karmakar 2011).

A worldwide contagious collapse happens only under conditions of global networks of intra-dependence (Table 30.4). In 2007–2008, global financial markets were tightly integrated than ever before. Although the financial collapse clearly hit North America and Europe more seriously than other parts of the world, the way financial markets in different parts of the globe react to changes has been synchronized to a remarkable degree.

On the other hand, the 2008–2009 crisis was generated by relatively enduring mechanisms of international political economy and related political choices (Sheng 2009). Global economy consists of states themselves engaging in financial activities and giving or taking loans. Their immediate financial situation depends on the prevailing current and capital account (im)balances. Since 1971–1973, currencies have been floating and the foreign exchange market has grown to be an essential part of the global financial system. The re-emergence of global finance has in effacement the return of some of the key features of the gold standard and haute finance era of 1870–1914, involving unprecedented integration and increased volatility.

With the re-emergence of the Europe and the rise of China, the systemic imbalances have been reflected in changing levels of confidence and exchange rates. As of the early 1990s, an increasing part of the world surplus has been accumulated by China in its foreign reserves. Moreover, since introduction in the 1990s, the euro has been the second most widely held international reserve currency after the US dollar providing an alternative to dollar. The euro has already surpassed the US dollar in terms of the combined value of cash in circulation.

The 2008–2009 global financial crisis has been caused by weak regulation of financial markets and intermediaries and excessive financial innovation and risk taking by finance and banking intermediaries at the micro-level on the one hand, and the loose monetary policies adopted by the reserve currency-issuing advanced economies (particularly the USA) and the mercantilist policies adopted by export-led economies leading to accumulation of large current account surpluses and reserves on the other as well as unlearning of the lessons of the 1930s and 1940s, and by short-sighted attempts by different state actors to use the options provided by the global monetary system for their own narrowly defined short-term benefit. As states had become entangled with free market global finance capital, and there is trade-off between states' relative competitiveness and financial stability, regulators and lawmakers have found it acutely difficult to control the system in concert. There is 'a glaring lack of governance of international monetary and financial relations'.

Policymakers in India were under the impression that global economic crisis in 2008 would in no way affect India, but when the crisis after crossing the Atlantic and Pacific oceans hit the oceans of the emerging markets like India and China through different channels as trade channel, financial channel and confidence channel, they quickly abandoned the 'decoupled' hypothesis.

Following the global financial crisis, net portfolio investment in India turned negative and stood at US$(−) 14 billion. Stock market prices fell sharply. Overall, stock price index continued to become low as compared to pre-crisis level even after 4 years business sentiments fear of contagion.

Table 30.4 Financial crises in the new millennium. (Source: compiled by the author)

Financial crises	Consequences	Spillovers and contagion effects
Turkish crisis, 2001	The Turkish economy experienced a severe recession, with real GDP declining by 9 % during 2001	One can note the *lack of contagion* accompanying the 2001 devaluation of the Turkish lira
Argentine financial crisis 2001–2002	Largest forming debt default in history. Links between debt sustainability, sovereign defaults and currency crisis came to the fore. And with banks closed and bills unpaid, the Argentine economy sank into depression The crises in East Asia, Mexico, Russia, Brazil and Argentina introduced new elements of instability in the global economy, largely the result of many countries liberalizing their BoP capital accounts and thus becoming vulnerable to massive 'hot money'	By early 2001, there was a flood of money leaving the country One can also note the *lack of contagion* accompanying the 2001 demise of the Argentine currency board
2008–2009 Global financial crisis	*But not all financial crises originate in developing countries.* The world economy, including its developing countries, is now trying to manage a crisis with origins in the developed world. The crisis hit hard in the fall of 2008 The causes of the crisis were located in the USA Unprecedented inflated housing prices bubble (nearly doubling in real terms in the decade up to 2007) in the USA suddenly began to drop, concessionary NINJA mortgages proved to be uncollectible, bank failures due to a lack of prudential financial regulation loomed (in the USA alone, 22 banks collapsed in 2008 and 77 more by August 2009) accompanied by huge inflows of capital (financing a current account deficit of more than 6 % of GDP) and major investment fund failures, sharp declines in stock indexes and large reductions in the market value of the commodities and a simple subprime mortgage meltdown in the USA quickly developed into a global financial tsunami and deep recession that was worse than anything else in the world	The unhealthy dimension of global financial liberalization put stress on the present as usually it did in the past The 2008–2009 crisis originated in developed country like the USA rapidly snowballed into a global credit crisis, deflation and reductions in international trade and then it spread to Europe and other emerging markets through contagion channel as trade, financial and investor's confidence channel

Table 30.4 (continued)

Financial crises	Consequences	Spillovers and contagion effects
	The World Bank estimates that 94 out of 116 developing countries have experienced a slowdown in economic growth due to which 53 million people, in addition to 130–155 million people, were pushed into poverty. Hunger is 'back on the march', with 115 million people added to the ranks of the hungry, driven by the economic crises (fuel, food and finance). Those countries mostly integrated into the global economy through trade, finance, FDI and remittances suffered most	
Eurozone crisis 2010	Until 2011, almost equivalent *global financial imbalances* caused essentially by increasing globalization of trade and finance of China's large current account surplus with all other countries in the world vis-à-vis total deficits of the USA with all countries have been a major contributor to the financial crisis. And they are likely to become larger with the increase in international capital flows (comprising 19 % of global GDP in 2012) along with larger global current account imbalances (*Bank of England, December 2011*). Likewise, the Eurozone problem arising out of deficits and surpluses of Eurozone countries is one of a failure of internal financing. The real problem which became manifest in 2010 is political as well as financial	The crisis started in Greece but spread rapidly to Ireland, Portugal and Spain and, subsequently, Italy and major economies there with sovereign debt level started to mount in the aftermath of the global financial crisis in 2008. The financial markets quickly transmitted the shocks which led to a sharp rise in the credit default swaps (CDS) spread and later impacted capital flows elsewhere
2010–2011 Sovereign debt crisis in EU countries	Austerity policies, reflecting harsh neoliberal financial discipline always favoured by TNCs and financial markets, have generated domestic resentment and instability, and regional financial imbalances in the Eurozone which have led to a sovereign debt crisis for some single currency zone members like Iceland, Ireland, Greece, Portugal, Spain and Italy. From the latter half of 2011, the crisis in the Eurozone peripheral countries started to adversely affect the major economies there. This crisis has also impaired the confidence of investors through contagion channels and is contributing to the uncertainty in global economic recovery. As per OECD Employment Outlook, unemployment remained high, with Spain exhibiting the highest unemployment rate (21.7 %) among the European countries	The Eurozone debt crisis and the rise in oil prices affected the emerging markets including India. The onset of the crisis in Euro region again put pressure on India on the exchange rate with the rupee depreciating to above Rs. 55 to a dollar in mid-2012. Real effective exchange rate (36-currency index) computed by the RBI also shows similar trend

Export growth fell to 13 % in 2008–2009 prior to the crisis level (20–25 %). Services exports in particular dropped to 13 % in 2009–2010. Contraction of export demand affected aggregate demand and reduced GDP growth in the economy. Nominal exchange rate depreciated sharply from Rs. 40.3 per dollar in 2007–2008 to Rs. 46 in 2008–2009 and to Rs. 47.4 in 2009–2010, but appreciated to Rs. 45.6 in 2010–2011.

East Asia, like other regions, was seriously affected by the global economic crisis that began in late 2007. China's current account surplus alone contracted from over US$191 billion in January–June 2008 to US$130 billion in the same period in 2009. The scale and speed of that downturn is breathtaking and broader in scope than in the financial crisis of 1997–1998. China's GDP, which expanded by 13 % in 2007, scarcely grew at all in the last quarter of 2008 on a seasonally adjusted basis. In the same quarter, Japan's GDP is estimated to have fallen at an annualized rate of 10 %, Singapore's at 17 % and South Korea's at 21 % (*The Economist*, 31 January 2009, p. 13). Besides, the Asian financial crisis gave the region's hybrid globalizers their first sharp setback, forcing them to reconsider many of the practices they had used to achieve rapid economic development. The event was a test for the new global order and brought home a morass of self-doubt, criticism and denigration of Asia's past achievement (*Godement*, 2002, p. 17, 125).

As the global economic system has struggled to recover from the global financial crisis, the regional differences are evident, with GDP growth of emerging market economies in 2010 and 2011, especially China (10, 9.7 %), India (7.7 %) and the Association of Southeast Asian Nations (ASEAN; 4.7 %) projected to be considerably higher than advanced economies such as the USA (2.7, 2.4 %), Japan (9, 1.7 %) or the Euro area (1.0, 1.6 %) (IMF 2010).

The Food and Agricultural Organisation (FAO) estimates that an additional 100 million people fell into hunger over 2008 and 2009—so the number of hungry people will rise to 1.02 billion in 2010, well up from the 825 million in 1995 (Craig and Porter 2005). The UN's *Millennium Development Goals Report* 2009 records that an estimated 55–90 million people will fall into US$1.25-a-day poverty in 2009 and the achievement of many other poverty goals will slow down.

Asia, especially China, India and the members of ASEAN was hit hard by the global financial crisis (Scott-Quin 2012). It suffered a dramatic reduction in exports to the West and weakening of domestic demand but, in general, fared better than other regions.

The fissiparous effects of financial turmoil in the Eurozone have produced Franco-German efforts to thicken economic and financial integration within the region. The global financial crisis has sorely tested European unity. The Eurozone protected its members from currency speculation, but the impact of the economic downturn has been very uneven. Several countries amassed deficits that were so large as to threaten sovereign default.

5.13 *The Eurozone Crisis 2010–2011*

After passing through good times in terms of both a decline in long-term interest rates from 2002 to 2006 and an increase in the degree of convergence in the interest rates of member countries, the Eurozone (a currency union of 17 European countries) faced a major crisis when, in early 2010, cross-border holdings of sovereign debt and exposure of banks came to light. Meanwhile, the Eurozone witnessed a decline in share of world GDP from 22.3 % in 2005 to 19.3 % in 2010 at current prices. The crisis started in Greece but spread rapidly to Ireland, Portugal and Spain and, subsequently, Italy with sovereign debt level starting to mount in the aftermath of the global financial crisis in 2008. These economies had witnessed downgrades in the ratings of their sovereign debts due to fears of default and a rise in borrowing costs which ultimately led to a spiral of rising bond yields and further downgrade of government debt of other peripheral Eurozone economies as well. Re-financing government debt for some of the countries became very difficult for this sovereign debt crisis. The banking and insurance sector with large sovereign debt exposure stood adversely affected. The financial markets quickly transmitted the shocks which led to a sharp rise in the CDS spread and later impacted capital flows elsewhere. Resolving the crisis becomes very much difficult as the Eurozone lacks a full-fledged central bank, a single fiscal authority capable of strict enforcement and it cannot adjust through a depreciation of currency. Though several packages of measures were taken by the European finance ministers and the European Central Bank during 2011, the overall uncertainty about the effectiveness of all these measures still remains. And it still faces problems such as the continuing recession; the existence of a monetary union without fiscal union; the slow progress of the proposed European banking union, the continuing need for austerity etc.

6 Financial Crises and Their Implications for India

Financial crises are devastating to poor people. They are vulnerable to crises because they do not have the savings or safety nets to protect themselves from the income losses. In the case of the Asian crisis, World Bank estimated that it involved 20 million persons falling back into poverty and 1 million children being withdrawn from school. In Indonesia alone, 35 million persons were pushed into absolute poverty. During the Argentine crises in 2001, close to one fourth of the population became extremely poor, while one half of the population fell below the poverty line. The global nature of the 2007–2009 and the Euro crisis of 2010–2011 continuing up until now may mean its costs to India and other developing countries are even higher than for the Asian crisis: The World Bank estimated that low- and middle-income countries lost 3–8 % of potential output compared with the pre-crisis path with 64 million more people in absolute poverty than if the crisis had not occurred. So far as the Indian economy is concerned, it after reporting fairly robust growth of more than 9 % during 2005–2008, moderated to a growth of 6.7 % because of the global financial crisis and to a growth of 5.5 % following the sovereign debt problem in the Euro zone. The recent global slowdown has thrown up new challenges for India with its export growth being continuously negative since May 2012 compared to very high growth rates of even above 50 % in some months of the previous year. The FAO estimated that the crisis increased the number of undernourished people by tens of millions. Given these poverty effects, caution is warranted, and there is evidence that the sequence and timing of financial sector reforms can mitigate financial turmoil and, thereby, prevent negative effects on poor people. But financial liberalization without the proper surveillance capability against the systematic risk inherent in global capital flows may destabilize local financial sectors, real economics and domestic political environments (Fergusan and Mansbach 2012). However, implementing the prudential regulation to shelter developing countries from the ups and downs of global finance capital is not easy. As is now evident and clear in the current crises, even the developed countries with their proclaimed advanced financial systems have not been able to effectively take on this important task (Gills 2011).

On the macroeconomic front, financial crises manifest capability failure on the part of monetary authority (to maintain exchange rate stability, besides stabilizing rate of interest or protecting the foreign exchange reserve) as well as the commercial banking system (in maintaining a balance among liquidity, profitability and solvency) along with non-functioning of the free market system. The sovereign debt crisis in Eurozone in recent years has not only aggravated the macroeconomic conditions of the countries of the Eurozone but also, in turn, deeply affected the balance sheets of global banks having exposures to these countries. The economic groups even in the BRICS (Brazil, Russia, India, China and South Africa) nations which serve as an engine of growth in the developing world have slowed down considerably. The European debt crisis and the global slowdown are creating serious headwinds for the Indian recovery and posing major challenges for the economy. On the domestic front, the large twin deficits pose significant risks to macroeconomic stability and growth sustainability. Global slowdown deters capital inflows and this makes difficult to finance the increasing current account deficit. Thus, it appears that financial crises developed in developed economies are bound to have considerable direct and indirect influence on our domestic economy and the financial system.

The Asian crisis and two others recent new-style financial collapses offer India to learn some hard economic policy lessons: applying the optimum order of liberalization, applying temporary restrictions on capital inflows and applying a temporary exchange rate anchor in that unrestricted movements of capital, for example, are dangerous; that there is no simple risk-free, fast track to sustained growth by opening up too quickly to capital flows and to allowing exchange rate to appreciate; and that, where possible (Karmakar and Mukhopadhyay 2011). India must finance growth through its savings. They also clearly demonstrate the need for strengthening domestic banking for its stability, and to achieve sustainable economic growth, strong financial system is urgently called for in order to discriminate against the inflow of hot money, to create financial safety nets and the necessary institutional framework to resolve the problems of poor policy response, moral hazards and information asymmetry.

To protect the Indian market from the short-term investment boom and undesirable shifts of capital flows, policymakers in India need to devise an appropriately prudent regulatory and supervisory framework that covers different aspects of governance with regard to capital account management reducing, thereby, its dependence on volatile portfolio flows, even if we ignore official flows, FDI and commercial lending by banks and other institutions (Karmakar 2010). Economic Survey, 2012–2013, itself voices forth that India's current account deficit, widened to 4.2 % of GDP, has been financed largely by capital inflows, altogether indicating that the dependence on private capital flows to finance the same has widened in recent years. This has increased BoP vulnerability to 'sudden stops' and reversal of capital, especially when sizeable flows comprise of debt, and volatile portfolio investment. It shows the danger (and this is a warning to other countries) of the rapid liberalization of international capital flows before the domestic banking system has developed sufficient control, and ever-growing BoP deficits relative to GDP by increasing short-term capital flows (Chatterjee and Karmakar 2011). The liberalization of all international private capital flows creates a vulnerability to financial crisis and introduces several risks: *currency risk* (culminating through the sudden inflows of capital to put pressure on the domestic currency to appreciate, and a large appreciation of the domestic currency is problematic because it undermines net export performance), *capital flight risk* (this induces a vicious cycle of additional flight and currency depreciation, debt service difficulties and reductions in stock or other asset values, thus making the investors panicky for which they sell their assets *en masse* to avoid the new capital losses being brought about by anticipated future depreciations of currency or asset values and when government fails to restrict the kinds of capital flows, viz. portfolio investment, short-term foreign loans and liquid form of FDI, this risk is severe), *fragility risk* (essentially referring to the vulnerability of an economy's internal and external borrowers to internal and external shocks that jeopardize their ability to meet current obligations, causing maturity mismatch or 'Ponzi' financing as coined by Minsky when borrowers finances long-term obligations with short-term credit, for example), *sovereignty risk* (risk in which a government will face constraints on its ability to pursue independent socioeconomic policies) and *contagion risk* (this refers to the danger of a country falling victim to financial and macroeconomic instability that originates elsewhere). Among them, severity of contagion risk obviously depends on the extent of currency, flight and fragility risks, while financial integration is the carrier of contagion risk. Countries can reduce their contagion risk by maintaining their degree of financial integration and by reducing their vulnerability to currency, flight and fragility risks through a variety of financial controls (Grabel 2003). From the above analysis, we may say that following Tarapore II recommendations, India has prematurely liberalized its capital accounts, on which most of the capital flows take place, though it resolved, to a great extent, the trilemma of the famed 'Impossible Trinity', which disallows the simultaneous achievement of exchange rate stability, monetary independence and capital market integration through capital account convertibility (Fig. 30.1). Any two of the goals may be attained (at the vertices of the triangle) but never all three. So far, India has successfully enjoyed substantial monetary independence and a fair degree of exchange rate flexibility by

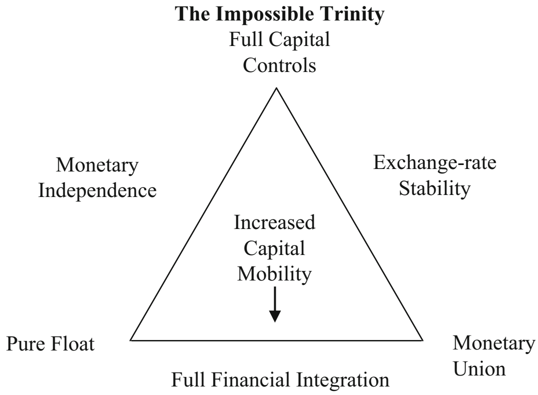


Fig. 30.1 The impossible trinity. (Source: Frankel 1999)

the good combination of managed flexibility and partial capital controls. It is by now known that the burden of adjustment in the current international monetary system falls predominantly on non-reserve-issuing current account deficit countries (like India) (karmakar 2012).

Also in view of the arguments advanced above and empirical evidence supporting that liberalization is strongly associated with banking, currency and generalized financial crises of different types, we may argue that controls on capital flows should not be hastily dismantled and there is a strong case for controlling international private capital flows in emerging markets like India on the following logic: Capital controls (that refer to the measures that manage the volume, composition or allocation of capital flows and/or maintenance of restrictions on investor exit or entrance opportunities) can promote financial stability and, thereby, prevent the economic and social devastation associated with financial crises. Second, capital control can promote desirable types of investment and financing arrangements and discourage less desirable types of investment/financing strategies. Third, capital controls can enhance democracy and national policy autonomy by reducing various external actors to exercise undue influence over domestic decision making.

And India will inevitably need to play an active role at global level in influencing the rules for the global economy on overarching macroeconomic issues such as trade, capital flows, financial regulation, climate change and governance of global financial institutions.

7 Conclusions

International financial crises have become more common, first after World War II with the proliferation of currency crises under the BW system in response to unsustainable macroeconomic policies, and, more recently, as banking crises have become more frequent after the financial liberalization of the 1980s and the 1990s. One disastrous result of this has been more twin crises, in which weak banks quicken the pace of capital flight and the size of currency crises, while more capital flight and large devaluations further weaken fragile banking systems in a vicious cycle. In the case of the East Asian crisis, for example, the result was an international economic contraction unmatched by any economic crisis except the Great Depression.

Fundamentals definitely play an important role in creating the conditions under which international financial crises occur. Macroeconomic imbalances such as high inflation, budget deficits and current account deficits are clearly associated with currency crises and capital flight. Trade and financial linkages have been important in the spread of contagion from one country to another. In recent crises, financial linkages between countries are increasingly creating contagion by spreading capital flight through things such as the carry trade (see footnote 1). Most importantly, indiscriminate financial deregulation, moral hazard lending and currency mismatches between domestic-denominated assets and foreign-denominated debt lead to fragile banking systems that cannot withstand even in a modest shock to the financial system.

But poor fundamentals alone are only necessary for a crisis, they are not sufficient to guarantee a crisis. Crises are too idiosyncratic to be explained by fundamentals alone; countries with poor macroeconomic and financial fundamentals do not necessarily experience a crisis, and often the timing and the events that trigger a crisis or the spread of contagion are not easily identifiable. Beliefs and expectations also appear to be important, particularly in explaining the spread of capital flight and contagion once a single crisis has taken place, either through the wake-up-call effect, through the herding behaviour of investors or because investors' returns are so intertwined that crises are self-fulfilling.

References

Bank of England (2011) The future of international capital flows. Bank of England Financial Stability Paper, December 2011

Burnside C, Eichenbaum M, Rebelo S (2001) Prospective deficits and the Asian currency crisis. J Polit Econ 109:1155–1198

Calvo G (1998) Capital flows and capital market crises: the simple economics of sudden stops. J Appl Econ 1(1):35–54

Calvo G, Izquierdo A, Mejia L (2004) On the empirics of sudden stops: the relevance of balance-sheet effects. Working PaperNo.10520: Cambridge, MA: NBER

Calvo G, Mendoza E (1996) Mexico's balance-of-payments crisis: a chronicle of a death foretold. J Int Econ 41:235–264

Chancellor E (2000) Devil take the hindmost: a history of financial speculation. Penguin, New York

Chatterjee B, Karmakar AK (eds) (2011) Capital account convertibility in India. Deep and Deep Publications, New Delhi

Corsetti G, Mackowiak B (2006) Fiscal imbalances and the dynamics of currency crises. Eur Econ Rev 50:1317–1335

Craig D, Porter DR (2005) Development beyond neoliberalism? The poverty reduction paradigm. Roultledge, London

Dicken P (2011) Global shift: mapping the changing contours of the world economy, 6th edition. Sage Publications Ltd, London

Edwards S (2004) Financial openness, sudden stops and current account reversals. Am Econ Rev 94(2):59–64

Fergusan YH, Mansbach RW (2012) Globalization: the return of borders to a borderless world? Roultledge, New York
Frankel J (1999) No single currency regime is right for all countries at all times, Princeton Essays in International Finance No. 215 (August)
Gills BK (ed) (2011) Globalization in crisis. Roultledge, London
Glick R, Hutchison M (2001) Banking & currency crises: how common are twins? In: Glick R, Moreno R, Spiegel M (eds) Financial crises in emerging markets. Cambridge University Press, Cambridge, pp 35–69
Glick R, Moreno R, Spiegel MM (eds) (2001) Financial crises in emerging markets. Cambridge University Press, Cambridge
Godement F (2002) The downsizing of Asia. Routledge, London
Goldstein M (1998) The Asian financial crisis: causes, cures, and systemic implications. Institute for International Economics, Washington, D.C.
Grabel I (2003) International private capital flows and developing countries. In: Chang H-J (ed) Rethinking development economics. Wimbledon Publishing Company, London
Grupen N (1986) The debt crisis in Latin America. Institute of Latin American Studies, Stockholm
Hutchison M, Noy I (2005) How band are twins? Output costs of currency and banking crisis. JMCB 37(4):725–52
International Monetary Fund (2010) World economic outlook update: a policy-driven multi-speed recovery. January 26, 2010
Kaminsky GL, Reinhart CM (1999) The twin crises: the causes of banking and balance of payments problems. Am Econ Rev 89(3):473–500
Karmakar AK (2010) Balance of payments theory and policy: the Indian experience. Deep and Deep Publications, New Delhi
Karmakar AK, Mukhopadhyay D (2011) Macroeconomics of exchange rate management in developing countries with special reference to India. Artha Beekshan 20(2–3):42–63 (September and October)
Karmakar AK (2011) Global financial meltdown and the Indian economy. In: Sen RK (ed) Modern Indian economy. Deep and Deep Publications, New Delhi
Karmakar AK (2012) Capital flows and financial crises. Artha Beekshan 20(4):77–93 (March)
Kindleberger CP, Aliber RJ (2011) Manias, panics and crashes: a history of financial crises, 6th edn. Palgrave McMillan, Basingstoke
Krugman P (1979) A model of balance of payments crises. JMCB 11:311–325
Krugman P (1998) What happened to Asia? http://web.mit.edu/www/DISINTER.html (January, 1998)
Krugman P (2008) The return of depression economics and the crisis of 2008. Penguin Books, Harmondsworth
McKinnon R, Pill H (1995) Credible liberalizations and international flows: the overborrowing syndrome. In: Ito T, Krueger A (eds) Financial deregulation and integration in East Asia. University of Chicago Press, Chicago
McKinnon R, Pill H (1999) Exchange-rate regimes for emerging markets: moral hazard & international overborrowing. OXREP 15(3):19–38
Obstfeld M (1986) Rational and self-fulfilling balance of payments crises. Am Econ Rev 76:72–81
Obstfeld M (1994) The logic of currency crises. Cashiers Economiques et Monetaries 43:189–213
Scott-Quin B (2012) Commercial and investment banking and international credit and capital markets. Palgrave Macmillan, London
Sheng A (2009) From Asian to global financial crisis: an Asian regulator's view of unfettered finance in the 1990s and 2000s. Cambridge University Press, Cambridge
Walters A (1986) Britain's economic renaissance: Margaret Thatcher's reforms, 1979–1984. Oxford University Press, Oxford

Index

A. N. Ghosh, A. K. Karmakar (eds.), *Analytical Issues in Trade, Development and Finance,* India Studies in Business and Economics,
DOI 10.1007/978-81-322-1650-6, © Springer India 2014