Simanti Bandyopadhyay
Mousumi Dutta *Editors*

Opportunities and Challenges in Development

Essays for Sarmila Banerjee

Springer

Opportunities and Challenges in Development

Simanti Bandyopadhyay · Mousumi Dutta
Editors

Opportunities and Challenges in Development

Essays for Sarmila Banerjee

Springer

Editors
Simanti Bandyopadhyay
Department of Economics and Public Policy, School of Management and Entrepreneurship
Shiv Nadar University
Greater Noida, India

Mousumi Dutta
Economics Department
Presidency University
Kolkata, India

ISBN 978-981-13-9980-0 ISBN 978-981-13-9981-7 (eBook)
https://doi.org/10.1007/978-981-13-9981-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Professor Sarmila Banerjee

Foreword

Sarmila Banerjee, affectionately called *Sarmiladi* by her innumerable students she nurtured at the post-graduate department of Economics in Calcutta University and elsewhere, retired as Rajiv Gandhi Chair Professor of Eco-systems and Sustainable Development, after serving for more than 35 years. Her remarkable leadership brought in qualitative changes in the orientation of the *Department* as a centre of economics learning—changing it from a predominantly teaching *Department* to one having a balanced mix of teaching and research. Evolution of her own academic interest was possibly coincident with this regime change of the *Department*. She is the recipient of Panchanan Chakraborty Memorial Award for lifetime achievement in teaching and research in economics from Bengal Economic Association in 2018.

Her doctoral research on trade policies, from the University of Florida, Gainesville, USA, was along the line of theoretical development economics. In course of her doctoral studentship, she received a series of best student awards like Lasky Award in 1985–1986, Lanzilotti Award in 1986–1987, Ann Lynch Award in 1987–1988 and President's Award in 1988–1989. She also received World Bank's Overseas Fellowship for Post-doctoral Research in Environmental Economics in 2001–2002 and studied environmental economics and sustainable development in her post-doctoral research, a pursuit that always remained close to her heart later, along with other areas of development of contemporary significance. It will not be an exaggeration to say that her post-graduate training abroad, in addition to stimulating her interest in developmental policy, issues of environmental economics and sustainable development, inculcated in her a strong appreciation and competence in using econometric techniques, which she later used in most of her quantitative analytic research. As her own research and also research that she supervised reflect, her research interest has covered three broad areas, viz. Environmental Economics and Sustainable Development, Empowerment of Women and Gender Discrimination and Social Sector and Development. The quantum of her research involvement is indeed remarkable, which covers a large number of Ph.D. and M.Phil. research supervisions also.

It is a matter of great pleasure and appreciation that as a tribute to Sarmila Banerjee, Simanti Bandyopadhyay and Mousumi Dutta, on behalf of students who have learnt from her how to think logically and do impactful research, have chosen to edit this volume in honour of Professor Banerjee. The volume contains 25 original articles by eminent colleagues, contemporaries and students of Professor Banerjee. As the title of this volume may suggest, all the chapters of the volume are connected to different areas of development. The editors have chosen to group the articles into following five sections with five chapters in each: Section 1: Macroeconomic Issues, Section 2: Institutions and Development, Section 3: Public Finance and International Economics, Section 4: Labour and Human Capital, Section 5: Inequality and Well-Being.

All the five chapters of Section 1 are woven in the framework of macroeconomics and present interesting analyses relating to different macroeconomic issues. Pradip Maiti demonstrates logically how the use of different forms of expectation formation rules, particularly those of rational expectations, may result in qualitatively different conclusions about the issue of unemployment—inflation trade off—i.e. the Phillips curve.

Soumyen Sikdar revisits a prediction made by J. M. Keynes in 1931 that substantial living standards improvement would take place globally over the next hundred years. He explains why this prediction has gone wrong and attempts to explore the implications of this *Keynesian Failure* for *our grandchildren*.

Ananya Ghosh Dastidar and Rashmi Ahuja analyse the recent investment slowdown in India using a model of aggregate private investment behaviour based on both demand and supply-side factors. Their empirical analysis based on the specified model attempts to identify short- and long-run drivers of private investments in India.

Agnirup Sarkar and Abhirup Sarkar attempt to examine the linkage, if any, between the stock market boom and economic growth. For this, they construct a model of stochastic growth of the economy with stock market, which predicts that whereas a stock market shock leads to a rise in growth, a productivity shock that increases the growth rate has an adverse impact on the market capitalization ratio of the stock market. The main findings of a follow-up empirical study based on a panel data set covering 35 countries suggest that causality from stock market shock to growth is stronger than the one from growth shock to the stock market capitalization.

Ayanangshu Sarkar and Malabika Roy deal with the nature of volatility of stock indices at the sectoral level in banking, oil and gas, health, FMCG and IT. Indian stock markets are analysed to understand the characteristics of the volatility of sectoral returns as well as their relationships with market returns, focussing on its time-varying nature and the presence of characteristics like volatility clustering. Existence of volatility clustering and 'spillover effect' between the domestic stock indices of different sectors are modelled using econometric techniques. The paper also attempts to investigate how far the different segments of stock markets are integrated with the international stock markets.

The second section titled Institutions and Development starts with the analysis of Abhishek Kabiraj and Tarun Kabiraj who argue that uncertainty of R&D outcome, together with associated problems of free-riding and imitation, may lead to under-investment in R&D and hence hamper long-run growth of economy—a problem which may get mitigated if firms cooperate in their R&D effort.

Abhik Mukherjee, Pratip Kumar Datta and Saumya Chakrabarti, in the next chapter of this section, consider informality of production organization as an institutional factor of interest and seek to examine to what extent the inter-firm heterogeneity in the degree of informality of manufacturing units may be responsible for the observed non-inclusive growth performance of the informal manufacturing sector. Using primary data, they design a study of the observed behaviour of informal manufacturing units using the *structure-conduct-performance* paradigm of industrial organization to link up observed behaviour of informal manufacturing units with relevant firm characteristics, market linkages and socio-economic-cultural factors of their locations.

For Sukanta Bhattacharya and Shirsendu Mukherjee, a micro-credit market is the institutional phenomenon of interest and choice of loan contract is the relevant institutional issue. They examine analytically the effect of different types of loan contract on a borrower's incentive for investment in information. Briefly, alternative trade-offs a borrower faces by collecting information on the potential of her intended projects under individual and joint liability loan contracts are found important for the choice of optimal contract type. Among other results, the authors observe that in specific situations, individual liability lending may emerge as the dominant lending arrangement.

Tanika Chakraborty, Reema Kumar, Anirban Mukherjee and Sarani Saha take the court or judiciary as a form of institution and examine the efficacy of court efficiency for development. Their empirical analysis indicates that effective contract enforcement leads to expansion of business enterprises.

Prasenjit Sarkhel and Subhalakshmi Paul address the issue of a trade-off between the quality of drinking water used and the resource cost of procuring that water. Using a discrete choice model on the IHDS data sets, they identify the factors that affect households' choice of the drinking water source and water collection activities.

The third section on Public Finance and International Economics starts with Hiranya Mukhopadhya's article discussing fiscal reform of Indian states from the political economic perspective by bringing in fiscal populism, political alliance and other factors with the purpose of settling empirically the ambiguities in the relationship between fiscal reforms and their political economic determinants.

Madhurima Koley and Kumarjit Mandal are concerned with the issue of vertical fiscal imbalances in the Indian states. Vertical fiscal imbalance is recognized as obstacle to subnational accountability and good fiscal performance. Analysing a data set covering 25 states for a fairly long time period, the authors estimate that the primary deficit of a state measured as a percentage of NSDP declines by 18 per cent per 1 per cent decline in the vertical fiscal imbalance indicator. This result can prove to be of utmost importance as far as policy implications for a country like India is concerned.

Pradyut Kumar Pyne and Saikat Sinha Roy examine the extent of exchange rate pass-through to sectoral import prices in an emerging market economy like India. For their empirical analysis, a simultaneous equation model incorporating demand and supply factors is specified and the corresponding set of reduced-form equations for import prices are estimated. The main empirical result, claimed to be important for trade and exchange rate policy-making for emerging market economies like India, shows that there is an incomplete exchange rate pass-through to import prices with the degree of pass-through varying across import product groups.

Amrita Ganguly and Ranajoy Bhattacharyya in their interesting work deal with the complexities in the pricing of petrol, diesel, liquefied petroleum gas (LPG) and other energy products in India. Set in the background of international pricing mechanisms and recommendations of International Institute of Sustainable Development and the International Monetary Fund, the paper discusses the break-up of the components of pricing, which involves taxes and subsidies by central and state governments. The discussion centres around the point that these products being necessities with very low elasticity of demand, incidence of taxes and/ or subsidies on them are almost 100%, as a result of which the international cartels (seeking profits) and local governments (seeking revenues) use them as a tool to maximize their pecuniary benefits from the manipulation of such prices.

In the last chapter of this section, Tarun Kabiraj and Arijit Mukherjee review the literature on the role of International Joint Ventures in developing countries with special reference to implication of factors like government policy, information asymmetry, transfer pricing and tariff jumping.

The fourth section on Labor and Human Capital opens up debatable issues and offers interesting analyses. Bibhas Saha uses an extended Basu-Van model of child labour that allows a household to split time between outside labour market activities and self-employment-based home production. The main findings indicate that if a minimal proportion of households is empowered with self-employment opportunities, child labour will not arise in equilibrium—a result of significant implications in the literature on Development.

Rajesh Bhattacharya and Sarmistha Sen study labour mobility in the handloom sector of West Bengal. Their study, based on a qualitative field survey in three selected districts of West Bengal, shows that at present, handloom sector of West Bengal is characterized not just by exit of weavers, there is also valuable inter- and intra-sectoral labour mobility. This observation may be important for development strategy formulation for the handloom sector of West Bengal.

Using DEA technique, Arpita Ghosh in the next chapter estimates total factor productivity (TFP) for elementary education in India and identifies the factors accounting for estimated overall TFP and its components due to technical change, efficiency change and scale efficiency change, separately, for categories of elementary education.

Kajal Lahiri and Lin Hua analyse the so-called Gulf War syndrome, viz. effects on the health of the Gulf War veterans. Using the 2001 National Survey of Veterans conducted ten years after the end of Gulf War and appropriate econometric methodology, they brought out evidence on the complexity in identifying the genuine health problems faced by varieties of war veterans.

The last chapter in this section is contributed by Zakir Hussain, Diganta Mukherjee, Mousumi Dutta and Susmita Mukhopadhyay which compares the foetal origin hypothesis and the predictive adaptive response theory (which hypothesizes that nutrition supply at foetal stage leads foetus to adapt to expected future environment). The main findings of this empirical analysis suggest that the predictive adaptive response theory can give comparatively better prediction of the probability of onset of non-communicable disease in mid-life due to starvation at foetal stage.

The fifth, and the last, section is on Inequality and Well-being. Anup Sinha in his interesting article analyses how intra-generational inequality can have a constrained impact on sustainable development. Sustainable development is generally defined in terms of inter-generational equality in the access of resources in an economy. He argues that reduction of intra-generational inequality is also an important factor to ensure sustainable development.

Lopamudra Banerjee and Snehashish Bhattacharya discuss how the impact of a natural disaster on people's lives may get affected by their conditions of informality, class they belong to and location along the cultural axes of victims. They also provide illustrations in support of their conclusion based on panel surveys in countries.

Hee Cheol Chung, Gauri S. Dutta and Jerry Maples build up an interesting application of small area estimation methodology for simultaneous estimation of median incomes of the American states.

Manisha Chakrabarty and Jayanta Mandi measure household-level diversity in consumer spending using an entropy-based diversity measure and then try to explain the inter-household variation of measured diversity in terms of the level of well-being and other relevant household characteristics. Their results indicate that the diversity follows a random two-way error component model as commodity group and monthly per capita consumer expenditure decile groups' random error turn out to be significant.

The last contribution in this volume is by Asis Kumar Banerjee. This article attempts an inter-temporal comparison of inter-state disparities in well-being of selected Indian states from a methodological point of view. Using asset ownership rather than per capita consumer expenditure as the indicator of well-being and the methodology of Lorenz dominance, instead of a specific inequality index in his exercise, he shows that more robust results of well-being comparisons are obtained compared to those available in the literature.

This collection of essays presented to Sarmila Banerjee by her colleagues, contemporaries and students is rich in content and quality and aptly reflects the affection and respect they have for her. Let me close by expressing my grateful thanks to her for the interactions that she could find time to have with me on many issues of mutual interest. I wish her a post-retirement life, equally meaningful and interesting, if not more, than working life as a teacher and researcher of economics that she has lived. May she continue stimulating others in doing meaningful research and leading a fulfilling life.

Dipankor Coondoo
Indian Statistical Institute
Kolkata, India

Preface

Professor Sarmila Banerjee has been an inspiration for many. She taught us how to build up quality and confidence in one's life as well as the lives surrounding him or her in every little way, adding up big. Apart from her significant achievements as an academician, we would like to remember her as a great human being, an efficient administrator, a complete homemaker and a wonderful time manager who can infuse positivity in others through her actions. We have seen her conceptualizing and implementing academic projects, structuring and teaching innovative courses, with the same levels of commitment and love with which she cooks meals and serves to her people or nurses the ailing members of her huge extended family comprising relations, students, friends or anyone in need. There are very few students who pursued their career in economics during her tenure and did not have an occasion to have got some feedback or a brilliant piece of advice on their work from her. Be it professional or personal life, she is ready to help anybody at any point of time. Her decision to step out of academics after retirement and get involved actively with different social issues which have bothered her throughout her life has taught us again that it is never too late.

The idea of this volume came from the deep respect we all have for her. Two of us took the responsibility for paper and tried our best to do justice to this opportunity. Whatever she has given us and to the society in general cannot be acknowledged through editing an academic volume like this. This endeavour may rather be interpreted as an exercise for us to know and understand contemporary research on different areas of development in our own way, involving people whom she knew, we knew and had worked in close association as a colleague, teacher, adviser or friend.

This journey has been very enriching from selecting a theme, to approaching potential contributors, to going through each piece of contribution for the volume. We are fortunate to have 25 quality research articles from scholars, eminent in their fields, from all over the world. Given the diversity of area of work Prof. Sarmila Banerjee was involved in, it was not an easy task to come up with a theme for the volume. We tried to keep it broad so that we can welcome contributions from diverse fields of development and do not restrict the authors in any narrow domain.

But there was a trade-off. Defining a structure to divide the book into parts was a real challenge. As economists, who are trained to understand and model trade-offs efficiently, we took up the challenge and plunged into it. We cannot say we are fully satisfied with the structure to the extent that each contribution is the most appropriate fit for the part it belongs to. But we would justify by saying that we have done constrained optimization in the process of clubbing the articles under each theme and we respected the constraints and the objective function equally!

Our sincere thanks to all the contributors for their valued submissions and also their kind cooperation at each stage of the publication. We thank Prof. Dipankar Coondoo (Professor retd.), Economics Research Unit, Indian Statistical Institute, Kolkata, for writing an interesting Foreword for this book. Thanks are due to Prof. Anjan Chakrabarti for writing an excellent piece on Prof. Sarmila Banerjee as a colleague in the Department of Economics, University of Calcutta. We thank immensely all our reviewers who reviewed the submissions and had given their precious time for this volume in giving comments in multiple rounds as each article went through a double-blind review process before publication. We thank Aishna Sharma (Shiv Nadar University) and Aditi Jamalpuria (Independent Researcher) for a careful reading of the entire manuscript and their observations on the draft. Our sincere thanks to Debashis Biswas, Inspector of Colleges, University of Calcutta, for taking the photograph of Prof. Sarmila Banerjee for this book. We would like to thank our respective family members for their cooperation and support in this journey, which has played a pivotal role in the timely completion of this publication. We are grateful to our respective institutions, Shiv Nadar University and Presidency University, for the support extended from time to time. Finally, sincere thanks are due to Nupoor Singh and her team of Springer International for their support without which the book would not have been possible. Their untiring efforts to match our expectations are really praiseworthy.

Thanking you.

Greater Noida, India	Simanti Bandyopadhyay
Kolkata, India	Mousumi Dutta
September 2019	

Contents

Public Finance and International Economics

Labour and Human Capital

Inequality and Well-Being

Editors, Authors and Reviewers

About the Editors

Simanti Bandyopadhyay is currently Associate Professor, Economics and Public Policy Area, School of Management and Entrepreneurship, Shiv Nadar University, India. She has nineteen years of research, consultancy and teaching experience, and currently focuses on applied development economics, mainly related to public finance and public policy. She completed her Ph.D. in Economics on a UGC Fellowship from the Centre for Economic Studies and Planning, Jawaharlal Nehru University, New Delhi, and has been an International Centre for Tax and Development Postdoctoral Fellow at the Andrew Young School of Policy Studies, USA, and Institute of Development Studies, UK. She has undertaken research, teaching and research guidance at various respected universities and research institutions in India and abroad and has also been a recipient of MacArthur Foundation Award. She regularly contributes research articles on public policy, economics and operations research, and serves as an external reviewer for a number of international journals. She has led important development projects and has worked closely with organizations like the World Bank, ADB, UNDP, IFMR, ICRIER, Planning Commission and different ministries and government bodies. She has also been the Chief Researcher for the background study of the Fourth State Finance Commission of the state of Tamil Nadu in India.

Mousumi Dutta is currently a Professor and the Head of the Economics Department, Presidency University. She is a Gold Medalist in M.Sc. (Economics) from Calcutta University and completed her doctoral degree on "Economics of conservation of built heritage: the case of Kolkata" from the same University. She has been the Principal or co-Investigator of several projects, funded by The International Growth Center, Rosa Luxemburg Society, Berlin, ICSSR and UGC. Prof. Dutta has published extensively on the built environment, health and gender issues in journals like Tourism Management, Journal of Cultural Heritage, Social Indicators Research and Journal of International Development. Prof. Dutta has also published four books on different socio-economic issues including gender and reproductive health. She has been invited and presented her research papers at Stockholm University, Corvinus University, Gutenberg University, Liverpool University, Leeds University, Winchester University and Shanghai University and at international conferences organized by the European Population Association, British Society for Population Studies, IUSSP and Asian Population Association.

Authors

Rashmi Ahuja Lal Bahadur Shastri Institute of Management, New Delhi, India

Asis Kumar Banerjee Calcutta University, Kolkata, India

Lopamudra Banerjee Bennington College, Vermont, USA

Rajesh Bhattacharya Indian Institute of Management Calcutta, Kolkata, India

Snehashish Bhattacharya South Asian University, New Delhi, India

Sukanta Bhattacharya University of Calcutta, Kolkata, India

Ranajoy Bhattacharyya Indian Institute of Foreign Trade, Kolkata, India

Saumya Chakrabarti Visva-Bharati University, Santiniketan, India

Manisha Chakrabarty Indian Institute of Management Calcutta, Kolkata, India

Tanika Chakraborty Indian Institute of Management Calcutta, Kolkata, India

Hee Cheol Chung University of Georgia, Athens, GA, USA

Ananya Ghosh Dastidar University of Delhi South Campus, New Delhi, India

Gauri Sankar Datta University of Georgia, Athens, GA, USA

Pratip Kumar Datta Rajatpur Indranarayan Vidyapith, Rajatpur, Birbhum, West Bengal, India

Mousumi Dutta Presidency University, Kolkata, India

Amrita Ganguly KPMG Advisory Services Private Limited, Kolkata, India

Arpita Ghose Jadavpur University, Kolkata, India

Zakir Husain Presidency University, Kolkata, India

Abhishek Kabiraj HSBC Pvt. Ltd., Kolkata, India

Tarun Kabiraj Indian Statistical Institute, Kolkata, India

Madhurima Koley K. K. Das College, Kolkata, India

Reema Kumari Gartner India Research and Advisory Services Pvt. Ltd, Bengaluru, India

Kajal Lahiri University at Albany—SUNY, Albany, NY, USA

Hua Lin University of Texas Southwestern Medical Center, Dallas, TX, USA

Pradip Maiti Indian statistical Institute, Kolkata, India

Kumarjit Mandal University of Calcutta, Kolkata, India

Jayanta Mandi Data Science, Vrije University Brussels, Brussels, Belgium

Jerry Maples US Census Bureau, Washington, DC, USA

Abhik Mukherjee Visva-Bharati University, Santiniketan, India

Anirban Mukherjee University of Calcutta, Kolkata, India

Arijit Mukherjee Nottingham University Business School, Nottingham, UK

Diganta Mukherjee Indian Statistical Institute, Kolkata, India

Shirsendu Mukherjee St. Paul's C. M. College, Kolkata, India

Hiranya Mukhopadhyay Asian Development Bank, Mandaluyong, Philippines

Susmita Mukhopadhyay Indian Statistical Institute, Kolkata, India

Subhalakshmi Paul Anandamohan College, Kolkata, India

Pradyut Kumar Pyne Ministry of Finance, New Delhi, India

Malabika Roy Jadavpur University, Kolkata, India

Bibhas Saha Durham University Business School, Durham, UK

Nirupam Saha Visva-Bharati University, Santiniketan, India

Sarani Saha Indian Institute of Technology Kanpur, Kanpur, India

Abhirup Sarkar Indian Statistical Institute Kolkata, Kolkata, India

Agnirup Sarkar Indian Institute of Technology Guwahati, Guwahati, India

Ayanangshu Sarkar Pailan College of Management & Technology, Kolkata, India

Prasenjit Sarkhel University of Kalyani, Kalyani, India

Sarmishtha Sen Syamsundar College, Shyamsundar, West Bengal, India

Soumyen Sikdar Indian Institute of Management Calcutta, Kolkata, India

Anup Kumar Sinha Heritage Business School, Kolkata, India;
Indian Institute of Management Calcutta, Kolkata, India

Saikat Sinha Roy Jadavpur University, Kolkata, India

Reviewers

Rudrani Bhattacharya National Institute of Public Finance and Policy, New Delhi, India

Shampa Bhattacharya Shiv Nadar University, Greater Noida, India

Anindita Chakrabarti University of Leeds, Leeds, UK

Arijit Chakrabarty Indian Statistical Institute, Kolkata, India

Gagari Chakrabarty Presidency University, Kolkata, India

Sujoy Chakrabarty Jawaharlal Nehru University, New Delhi, India

Saumen Chattopadhyay Jawaharlal Nehru University, New Delhi, India

Anirban Dasgupta South Asian University, New Delhi, India

Arghya Ghosh University of New South Wales, Sydney, Australia

Krishnendu Ghosh Dastidar Jawaharlal Nehru University, New Delhi, India

Sabyasachi Kar Institute of Economic Growth, New Delhi, India

Lokendra Kimawat Ramjas College, New Delhi, India

Amita Majumdar Indian Statistical Institute, Kolkata, India

Rajarshi Majumdar Bardhaman University, Bardhaman, India

Subrata Majumdar Sundarban Mahavidyalaya, Kakdwip, India

Sabuj Mandal Indian Institute of Technology Madras, Chennai, India

Arnab Mukherjee Indian Institute of Management Bangalore, India

Partha Ray Indian Institute of Management Calcutta, Kolkata, India

Indrani Roy Chowdhury Jawaharlal Nehru University, New Delhi, India

Prabal Roy Chowdhury Indian Statistical Institute, New Delhi, India

Sahana Roy Chowdhury International Management Institute, Kolkata, India

Debdatta Saha South Asian University, New Delhi, India

Satwik Santra Centre for Studies in Social Sciences, Kolkata, India

Macroeconomic Issues

Theories of Expectations Including Rational Expectations and Their Uses in Different Versions of the Philips Curve

Pradip Maiti

Abstract Expectations are used in almost every branch of economics. And alternative theories of expectations have been developed, including the theory of rational expectations (RE) introduced by Robert Lucas (Jr.). The use of RE in macroeconomic models has changed radically many well-developed earlier results and policy prescriptions. The purpose of the present paper is twofold—first, to discuss the various theories of expectations, and secondly, to show how the Phillips curve in macroeconomics dealing with the celebrated trade-off between inflation and unemployment—yield different results, if alternative expectations mechanisms are used. We thus discuss different versions of the Phillips curve—first, its initial version due to Phillips, showing the existence of a permanent trade-off between inflation and unemployment, next the Friedman–Phelps version using static expectations and showing the existence of only temporary, but no permanent trade-off, and finally, the new classical version with rational expectations showing the absence of any trade-off even temporarily. We add that no new result has been proved here; rather, our objective is mainly to help the students to understand these topics clearly—students who might have felt that the discussion of these topics in the standard textbooks is not always up to expectation.

Keywords Expectations—Static · Adaptive (AE) and Rational Expectations (RE) · Cobweb model · Systematic error · Serially correlated · Mathematical expectation · Mark-up pricing · Trade-off · Phillips curve (PC) · Natural level of output · Expectations-augmented Phillips Curve · White noise · New classical · New Keynesian Phillips curve · Lucas critique · NAIRU

1 Introduction

Expectations are used in almost every branch of economics. But what do we mean by the term "expectations" and why are these so important? A very good introduction to this topic is given in the book by Carter and Maddock (1984). Let us quote from

---

P. Maiti (✉)
(Former Professor) Indian statistical Institute, Kolkata, India
e-mail: pmaiti.econ@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_1

it. "Expectations (in economics) are essentially forecasts of the future values of economic variables which are relevant to current decisions. For example, firms have to forecast the future prices for their products… in order to decide how much to produce in the current period… Union negotiators have to predict the future rate of inflation in their wage bargaining" (p. 12). In fact, for many a crop (say, rice, wheat, etc.), sowing takes place a few months before harvesting and hence, farmers need some forecast of its price in the next period—the period when its harvesting takes place, in order to decide on the size of the area on which the crop is to be cultivated during the current period.

Carter and Maddock (1984, Chap. 2, p. 13) remark, "one of the first economists to give expectations a paramount role in his analysis was John Maynard Keynes …. While expectations play a role in Keynes' earlier writings, they take centre stage in his *General Theory* (1936). His analysis of the level of employment, the demand for money, the level of investment and the trade cycle all depend crucially on expectations". However, Keynes did not discuss how such expectations were formed and/or did not pursue the implications of having alternative theories of expectations in a model.

The purpose of the present paper is *twofold—first*, to discuss the various theories of expectations developed/used in economics, and *secondly,* to show how the Phillips curve—an important topic in macroeconomics dealing with the celebrated trade-off between inflation and unemployment—yield different results, if one uses different expectations mechanisms to discuss this model. We would like to add that we are not going to prove any new result hitherto unknown; rather, the objective is to demonstrate the known results clearly and elaborately. In fact, our targeted readers are mainly students who might have felt that the treatment of these topics in the standard textbooks is not always up to expectation; our objective is mainly to help the students to understand these topics clearly.[1]

The plan of the paper is as follows. Section 2 introduces various models of expectations, i.e. static expectations, adaptive expectations and finally, rational expectations, each illustrated in a suitable model. Section 3 discusses the various versions of the Phillips curve—first, its initial version as presented by Phillips, showing the existence of a *permanent* trade-off between inflation and unemployment, next the Friedman–Phelps version, showing that such a trade-off holds only temporarily, but not permanently, and finally, the new classical version with rational expectations showing that no such trade-off exists even temporarily. Section 4 concludes our discussion.

[1] I guess, this objective suits the basic objective of this volume—the volume which is being brought out in honour of Professor Sarmila Banerjee who was a very popular and respected teacher in the Department.

2 Alternative Theories/Models of Expectations

2.1 Theory of Naïve, i.e. Static Expectations and Cobweb Model

The theory or model of **static** expectations (also called **naïve** expectations) is one of the earliest models of expectations developed in the literature. This model takes the immediate past value of a variable as its expected value for the next period. A well-known application of this model is in the context of the cobweb model which is discussed in detail below.

In case of many agricultural crops, there is a time gap between sowing and harvesting seasons, and hence, farmers have to take a prior decision as to how much land to cultivate so as to produce the desired amount of output.[2] And this decision is influenced by the price of the crop they **expect to prevail** when the crop will be harvested and marketed. Suppose their expectation is the *static* one, namely that p_t^e, the price of the crop *expected* by the farmers at the time of sowing [i.e. in period $(t-1)$] to prevail in the period after harvest (i.e. in period t) is the same as the price *observed now,* i.e. in the period of sowing: $p_t^e = p_{t-1}$. Thus, the quantity of output supplied in period t (q_t^s) depends on the amount of land cultivated in period $(t-1)$, and the latter depends on p_t^e. However, (q_t^d), the demand for output in any period t depends on the actual price which prevails in that period (p_t). And this price p_t is a market-clearing price in the sense that it equates demand with whatever amount is supplied. This model, which is written below in detail, is known as the famous *cobweb* model.

$$\text{(demand):} \quad q_t^d = \alpha - \beta p_t \quad \left(q_t^d = \text{quantity of output demanded at time } t\right); \tag{1}$$

$$\text{(supply):} \quad q_t^s = \gamma + \delta p_t^e \quad \left(q_t^s = \text{quantity of output supplied at time } t\right); \tag{2}$$

$$\text{(market clearing condition):} \quad q_t^d = q_t^s = q_t, \tag{3}$$

where q_t = actual quantity marketed at t. Solving (1)–(3), we get p_t as

$$p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p_t^e \left(\text{i.e. } p_t = a - bp_t^e, \text{ where } \quad a \equiv \frac{\alpha - \gamma}{\beta} \text{ and } b \equiv \frac{\delta}{\beta}\right). \tag{4}$$

The parameters α, β, γ are all positive; δ is generally taken to be positive, but it may be negative also. Now, static expectations give the expected price as

$$p_t^e = p_{t-1} \quad \text{(naïve or static expectations).} \tag{5}$$

[2] A good discussion on the cobweb model is given in Chakraborty (2002, Appendix).

Substituting then p_{t-1} for p_t^e in (4), we may rewrite the latter as follows:

$$p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p_{t-1} \quad \left[= a - bp_{t-1}, \text{ in terms of } a \text{ and } b \text{ defined in (4) above}\right]. \tag{6}$$

Equation (6) is a *first-order linear difference equation* which can be solved provided one *initial condition* is given. Suppose price at time 0 is given at p_0. Then the solution of (6), in terms of p_0 and the parameters, is as follows:

$$p_t = (-\delta/\beta)^t (p_0 - \bar{p}) + \bar{p}, \tag{7}$$

where $\bar{p}$ may be called the *equilibrium price* in this model,[3] i.e. that price which will satisfy both the demand and supply curves [i.e. both sides of (6)] and hence will be repeated period after period, if no other disturbances occur:

$$\text{(equilibrium price): } \bar{p} = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta}\bar{p} = \frac{\alpha - \gamma}{\beta + \delta} \quad \left(= \frac{a}{1+b} \text{ in terms of } a \text{ and } b\right). \tag{8}$$

Now, as Eq. (7) shows, the market-clearing price p_t is expressed in terms of deviation from the equilibrium price, $\bar{p}$, and will converge to the latter, i.e. the model is *stable*, if $|\delta/(-\beta)| < 1$ [i.e. if, with respect to the price axis, the demand curve has a larger (absolute) value of the slope $(-\beta)$ than the value of the slope of the supply curve (δ)]. Figure 1a shows a *convergent cobweb* model, and Fig. 1b, with a downward-sloping supply curve (i.e. with $\delta < 0$), shows a *convergent non-cobweb* model as the market price converges *steadily* to the equilibrium price.

Remark A: (i) Variables have oscillatory characters (from which the model derives its name, *cobweb*)—a period of low supply and high price is immediately followed by a period of high supply and low price.

(ii) The cobweb model may/may not converge. The stability condition is $|\delta/\beta| < 1$. However, if $\delta = \beta$, the model has a *neutral stability*.

Criticisms of the cobweb model

(1) Farmers ignore the impact of similar actions by the other farmers. For example, suppose in a period supplies are high and consequently prices are low. This induces each farmer to plant a smaller crop for the next period with the expectation that low price will be maintained, but since all the farmers will follow

[3]The derivation of the equilibrium price $\bar{p}$ and the solution of Eq. (7) are shown below. Given p_0, we have $p_1 = bp_0 + a$ and hence, $p_2 = bp_1 + a = b\{bp_0 + a\} + a = b^2 p_0 + a(1+b)$. By making repeated substitutions in this way, one gets $p_3 = bp_2 + a = b\{b^2 p_0 + a(1+b)\} + a = b^3 p_0 + a\{1 + b + b^2\} = b^3 p_0 + a(1 - b^3)/(1-b)$. Hence, $p_t = b^t p_0 + a(1 - b^t)/(1-b) = b^t[p_0 - a/(1-b)] + a/(1-b)$, which may be rewritten, by inserting the expressions for b and $\bar{p}$ as: $p_t = (-\delta/\beta)^t (p_0 - \bar{p}) + \bar{p}$

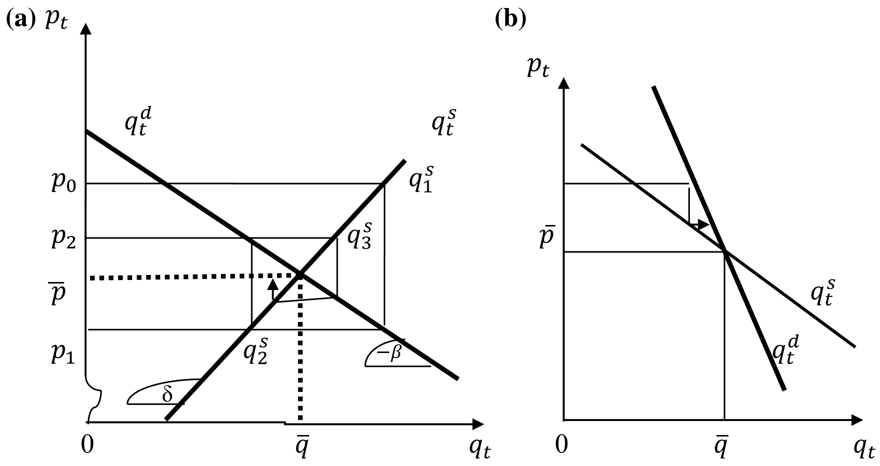


Fig. 1 **a** A convergent cobweb model (to explain **a**, given p_0, supply in period 1 is q_1^s which is demanded at p_1 which is then the market price in period 1; this price then induces farmers to supply q_2^s in period 2, and the market price in period 2 then settles at p_2 at which demand = this amount of supply. The story goes on in this way and ends when the market settles at the *equilibrium price*, $\bar{p}$—the price at which both the quantity demanded and supplied are the same). **b** A convergent non-cobweb model

suite, total supply will be much less and prices higher in the next period. Surely over time, the farmers are expected to appreciate the effects of their combined actions. But the model does not take care of this point.

(2) Even if the farmers are not sufficiently smart to consider the impact of their joint actions, one expects them to learn from their mistakes and adjust their actions accordingly. However, the assumptions of the cobweb model do not allow for any learning on the part of the farmers. Hence, a model of static expectations used to describe the cobweb model is very unsatisfactory.

2.2 *Theory of Adaptive Expectations (AE)*

An expectations mechanism—usually attributed to Cagan (1956) who used it in his *model of hyperinflation*—postulates that agents revise their expected value of a variable each period so as to correct the entire/a part of the error made in their previous expectations. This mechanism is thus called *adaptive expectations* (AE) and is stated below in detail:

$$p_t^e = p_{t-1}^e + \theta\left(p_{t-1} - p_{t-1}^e\right), \quad (0 < \theta \le 1), \tag{9}$$

where θ, called the *coefficient of adaptation*, determines the speed at which expectations adjust to correct past errors.[4]

Remark B: Some features of adaptive expectations theory are noted below.

(B.1) Note that Eq. (9) may be rewritten as[5]:

$$p_t^e = \theta p_{t-1} + (1-\theta) p_{t-1}^e, \quad (0 < \theta \leq 1), \tag{10}$$

which means that the expected price for a period is a weighted average of the actual and the expected price in the preceding period with θ being the weight attached to the former. This is another interpretation of the AE mechanism.

(B.2) By continuously substituting for p_{t-1}^e, p_{t-2}^e, p_{t-3}^e, etc., one obtains

$$\begin{aligned} p_t^e &= \theta p_{t-1} + (1-\theta) p_{t-1}^e = \theta p_{t-1} + (1-\theta)\left\{\theta p_{t-2} + (1-\theta) p_{t-2}^e\right\} \\ &= \theta p_{t-1} + (1-\theta)\left[\theta p_{t-2} + (1-\theta)\left\{\theta p_{t-3} + (1-\theta) p_{t-3}^e\right\}\right] \\ &= \theta\left[p_{t-1} + (1-\theta) p_{t-2} + (1-\theta)^2 p_{t-3}\right] + (1-\theta)^3 p_{t-3}^e \\ &= \cdots = \theta \sum_{k=1}^{\infty} (1-\theta)^{k-1} p_{t-k}. \end{aligned} \tag{11}$$

This is known as a *distributed lag model*, since the weight is distributed over all the lagged (i.e. past) values of the variable, with the *weights* attached to the preceding values *decline* at a geometric rate (since $0 < \theta \leq 1$), as we go further and further into the past. Thus, the **entire past history** of the variable is used to form its expected value for the immediate future (in contrast, the static/extrapolative expectations use one or two past values only).

Until the advent of the idea of rational expectations, adaptive expectations was the most common formalisation of expectations used in economics, partly because of its conceptual simplicity and easy applicability and partly owing to the fact that statistical estimates for the coefficient θ can be easily obtained. Moreover, adaptive behaviour in the face of uncertain environment appears intuitively plausible and appealing. Further, models with adaptive expectations appeared to work well in an environment in which changes were gradual—a characteristic of the 1950s and 1960s (Carter and Maddock 1984, pp. 23–24). Still, there are *problems* associated with the use of an adaptive expectation mechanism. We list some of these problems below.

Criticisms of the Adaptive Expectations Theory

First, there is not much empirical support for the assumption that weights decline geometrically. *Secondly*, sticking to adaptive expectations will, in many cases, lead to

---

[4] Metzler (1941) introduced *another* expectations mechanism, namely *extrapolative expectations*, where the expected value depends on the past value of the variable as well as on its direction of change in the past: $p_t^e = p_{t-1} + \in (p_{t-1} - p_{t-2})$, where $\in$ is called the *coefficient of expectations*. However, this model has not been a very popular one. For further discussion on it, see Carter and Maddock (1984, pp. 17–18).

[5] Note that static expectation is a special case of this mechanism with $\theta = 1$.

increasing forecasting errors (one such case is discussed later in Example I). However, the *most important criticism* is that adaptive (or for that matter, "static") expectations incorporate *only* past values of *the variable* being forecast. However, using *also* past values of *other relevant variables* and/or available relevant contemporaneous information may enable the decision-maker to make better forecasts. But adaptive expectations hypothesis just ignores this. Thus, for example, the prior knowledge of (i) the nature of agricultural production (bumper crop or drought), (ii) the announced changes in (*a*) government's monetary and fiscal policies and/or (*b*) current wage structure may be relevant—more relevant—than the knowledge of the past inflation rates for predicting the future inflation rates.

Let us consider a hypothetical example which will not only illustrate the above points, but also serve to introduce the idea of rational expectations.

Example I (**A Hypothetical Economy**)
(Ia) Suppose an *economy* in period t is described by the following upward-rising *aggregate supply* curve $(\mathrm{AS})_t$ and a quantity theory-type *aggregate demand* curve $(\mathrm{AD})_t$ in which the income velocity of money is taken to be a constant over time and for simplification, its value is also normalised at unity:

$$(\mathbf{AS})_t\text{:}\quad Y_t = Y^* P_t^{\beta}(\beta > 0) \quad \text{and} \quad (\mathbf{AD})_t\text{: } M_t = P_t Y_t, \tag{12}$$

where Y^* denotes the *natural* level of output[6] (assumed to be a constant across periods) and Y_t, P_t and M_t denote, respectively, the level of each of output, price and money supply in period t. Equating demand with supply, we get: $M_t = Y^* P_t^{1+\beta}$; this equation may be rewritten in logarithms, and then denoting the logarithm of a variable by the corresponding lower-case letter (e.g. $p_t \equiv \ln P_t$), we get the equilibrium price (in logarithm) for period t as

$$(\text{equilibrium price for period } t)\text{: } p_t = \sigma\left(m_t - y^*\right) \quad \left[\text{where, } \sigma \equiv \frac{1}{1+\beta}\right]. \tag{13}$$

Finally, writing Eq. (13) in first differences, we get

$$p_t - p_{t-1} = \sigma\left[(m_t - m_{t-1})\right] \quad \left[\text{where } 0 < \sigma \equiv \frac{1}{1+\beta} \leq 1\right]. \tag{14}$$

We make a few additional assumptions about the economy's functioning:
(Ib) Money supply is a *policy* variable. We first assume that the authority follows a *fixed* rule say, it raises M at a *constant* rate g every period:

$$(\text{policy})\text{: } \Delta m_t \equiv m_t - m_{t-1} = g \text{ for all } t \ (\text{where } m_t \equiv \ln M_t), \tag{15}$$

and that this rule is *known* to, and *believed* by, all agents in the economy.

---

[6]The concept of the *natural* level of output is explained later in Sect. 3.1 when we start discussing the different versions of the Phillips curve.

(Ic) Expectations regarding the future price level are formed adaptively:

$$p_t^e = p_{t-1}^e + \theta\left(p_{t-1} - p_{t-1}^e\right) \quad \left[\text{where } p_t^e \equiv \ln\ P_t^e\right]. \tag{16}$$

(Id) Expectations are fulfilled in the initial period, i.e. $p_0^e = p_0$.

The crucial *question* is: given this set-up, *will the expectations be fulfilled in the future periods as well*? Note that the error in expectations in period t, e_t $(=p_t - p_t^e)$, may be rewritten as follows:

$$\begin{aligned} e_t = p_t - p_t^e &= p_t - p_{t-1}^e - \theta\left(p_{t-1} - p_{t-1}^e\right) \left[\text{substituting (16) for } p_t^e\right] \\ &= p_t - p_{t-1} + p_{t-1} - p_{t-1}^e - \theta\left(p_{t-1} - p_{t-1}^e\right) \text{ (by subtracting, adding } p_{t-1}) \\ &= \sigma(m_t - m_{t-1}) + (1-\theta)e_{t-1} \left[\text{using (14) and the definition of } e_{t-1}\right] \\ &= \sigma g + (1-\theta)e_{t-1} \left[\text{using (15)}\right]. \end{aligned} \tag{17}$$

Suppose, initially there is no error in expectations, i.e. $e_0 = 0$. However, given that the authority is raising money supply at a constant rate g every period, Eq. (17) shows that there will be a positive error in expectation in period 1 ($e_1 = \sigma g > 0$) and that this error will go on increasing every period, ultimately settling at the value, $\bar{e}$ $(=\sigma g/\theta) > 0$, in the long-run.[7]

Remark C: We thus observe that, for the system (Ia)–(Id), the adaptive expectation mechanism performs poorly. Even if there is no forecasting error to start with (i.e. $e_0 = 0$), there is a (positive) error (σg) in period 1. And rather than converging to zero, the error in expectations increases from period to period, ultimately stabilising at a *positive value*, $\bar{e}$ $(=\sigma g/\theta)$. So people will be making forecasting error period after period and will be making this error *systematically*.

In summary, forecasts based on adaptive expectations are effective for reasonably stable variables, but are of little use in forecasting trends. This feature presumably accounts for the popularity of this mechanism in the 1950s and 1960s when inflation rates were low/relatively stable (Carter and Maddock 1984; Chap. 2, pp. 23–24). However, when inflation rate started to rise in the 1970s, adaptive expectations turned out to be a poor performer.

Given the limitations of adaptive expectations, a much superior forecasting, i.e. expectations-generating method, suggests itself. Since the working of the economy is described by (Ia) and since the authority follows a fixed rule for money supply, i.e. (Ib) which is known, the proportionality relation (14) between the rate of growth of money $(m_t - m_{t-1})$ and the rate of inflation $(p_t - p_{t-1})$ will be revealed to the rational economic agents, if the situation continues for some time. They are very likely to use this information to find that the optimal forecasting rule is: $p_t^e - p_{t-1} = \sigma(m_t - m_{t-1})$ whence $e_t = 0$ every period. In fact, even if the authority does **not** follow any **fixed** rule in determining the money supply, but **announces publicly** in the current period

[7]The first-order linear difference Eq. (17), if solved, will have a stable equilibrium at a positive value, $\bar{e} = \sigma g/\theta$.

the quantity of money supply in the next period (say, $\hat{M}_t$), the rational economic agents can again make perfect forecasts of future prices, if the economy is still described by (Ia), simply by utilising the publicly announced money supply figures, $\hat{m}_t$ (i.e. ln $\hat{M}_t$), in Formula (14) to get, $p_t^e - p_{t-1} = \sigma\left(\hat{m}_t - m_{t-1}\right)$. And, the error in the expectation of inflation rate would once again be zero every period.

Let us summarise our observations. The inflation rate is determined as a fixed proportion of the growth rate of money supply [Eq. (14)]. And, if money supply is determined by a *fixed* rule (Ib), i.e. Eq. (15) as well, a rational economic agent, in all likelihood, will be able to figure out both the features and make perfect predictions of future price level or inflation rate. But even when the authority does not follow any fixed rule for money supply growth, but announces each period the money supply figure for the following period, rational agents will again utilise this announcement and predict the next period's price level/inflation rate perfectly [via (14)].

We have thus come to the essence of the idea of rational expectations. Over time, economic agents accumulate relevant information about an economic system, e.g. relationships governing economic variables [such as (Ia) or its reduced form, the relation (14)], the behaviour of other agents, in particular, the government [such as its policy rule (Ib)]. They would behave rationally and use this information in forming expectations about the future values of various variables. This is the basic presumption of the theory of rational expectations.

2.3 Theory of Rational Expectations

The basic idea of rational expectations comes from a path-breaking paper by Muth (1961) who observes that traditional expectations formulae (e.g. static, extrapolative and adaptive) which have been used in the analysis of dynamic models have little resemblance to the way the economy works. He argues that the character of the dynamic processes is typically very sensitive to the way expectations are influenced by the actual course of events, and it is often necessary to make sensible predictions about the way expectations would change, when either the available information or the structure of the system is changed and in this respect, the traditional models of expectations (discussed earlier) perform very poorly, as the latter do not permit any such changes when the set of available information or the structure changes. Muth argues that the behaviour of any economic system depends on its structure, and the rational economic agents would use their knowledge of this structure in forming their expectations and "expectations, since they are informed predictions of future events, are essentially the same as predictions of the relevant economy theory" (Muth 1961, p. 316).

Let us try to expound the above statements. One serious limitation of earlier expectations theories, e.g. the adaptive expectations mechanism, is that these are based on **ad hoc rules** which allow individuals to make **systematic forecasting errors**, period after period, without requiring any amendment to the basis of the given forecasting rule followed.[8] This does not, however, imply that individuals forecast accurately in a world in which some random movements are inevitable; rather, the assertion is that guesses about the future must be correct *on average*, if individuals were to remain satisfied with the mechanism of expectations formations followed by them.

Now what degree of information should the individuals be then assumed to possess? A simple assumption could be that an agent knows *only* the past values of the particular variable (say, x) whose future value is to be predicted. But this assumption is unduly restrictive, since intelligent agents would come to realise that a number of other variables are likely to affect the value of x (i.e. some kind of a structure of an economic model), and hence, they may be supposed to realise that past values of other variables are likely to convey additional information about future realisations of x. When such information is widely available, it will be more reasonable to suppose that individuals (a) **know** the **structure** of the model, (b) observe the **previous values** of **all** the **relevant variables** and (c) also know—in case of a *stochastic* model (i.e. a model which also covers randomness of the world in which we live)—the **statistical properties** of these random disturbances. They then make **use** of all this information to **form expectations** about future values of these variables, and in such a set-up, **agents' expectations** will be the **same as** the **mathematical expectations** of the variables, conditional on the information set available at the time when expectations are to be formed. Thus, according to rational expectations hypothesis, (subjective) expectations of individuals = (objective) mathematical expectations (i.e. means) of the conditional distributions of the stochastic variables (conditional on the information set available up to date).

Illustration of Rational Expectations in the Context of the Cobweb Model

We now discuss the implications of having rational expectations in the cobweb model introduced at the beginning of this article. It was discussed then as a deterministic model. We shall now consider its ***stochastic version*** by adding a stochastic term u_t to the supply function $-u_t$ which assumes a random value at each time t (e.g. weather could be one such random variable whose variations affect crop *yield*). Thus, q_t^s, the quantity of the crop supplied in any period t, is affected by both the *area* planted (which depends upon the expected price, p_t^e) and the *yields* (represented by the random variable, u_t) We also assume that the *unconditiona*l distribution of u_t has *zero mean,* i.e. $E(u_t) = 0$. So, Eqs. (1) and (3) of the earlier model remain unchanged, but supply Eq. (2) and the model's *reduced form* equation, i.e. Eq. (4), change [with $E(u_t) = 0$] as follows:

$$\text{(supply): } q_t^s = \gamma + \delta p_t^e + u_t; \tag{2s}$$

$$p_t = \frac{\alpha - \gamma}{\beta} - \frac{\delta}{\beta} p_t^e - \frac{1}{\beta} u_t. \tag{4s}$$

[8]Recall that in our Example I, discussed earlier, the "forecasting error", under adaptive expectations, was found to be increasing over time, ultimately settling at a *positive* value.

Now, the actual price depends on both p_t^e and u_t. How p_t^e is determined now? The theory of rational expectations introduced by Muth takes the expected price to be generated endogenously by the model itself. The agents are being assumed (i) to know the structure of the model and (ii) to equate their subjective expectations for p_t with $E_t(p_t)$, since they have rational expectations, where $E_t(p_t)$ $\{\equiv E(p_t|I_t)\}$ is the mathematical expectation of p_t conditional on I_t (the set of information available at the beginning of period t). Thus, by substituting for p_t^e in Eq. (4s) by its rationally expected value, $E_t(p_t)$ and then taking conditional expectation, given I_t of both sides of the resultant equation [with $E_t(u_t)$ denoting $E(u_t|I_t)$] we get

$E_t(p_t) = (\alpha - \gamma)/\beta - (\delta/\beta)E_t(p_t) - E_t(u_t)/\beta$, or, simplifying we get

$$E_t(p_t) = (\alpha - \gamma)/(\beta + \delta) - E_t(u_t)/(\beta + \delta) = \bar{p} - E_t(u_t)/(\beta + \delta), \tag{5s}$$

where $\bar{p}$ is the *equilibrium price* of the deterministic cobweb model:

$$\bar{p}\ (equilibrium\ price \text{ of the cobweb model with } u_t = 0\,\forall\, t) = \frac{\alpha - \gamma}{\delta + \beta}. \tag{6s}$$

But what is $E_t(u_t)$—the mean of the conditional distribution of u_t? Note that $\{u_t\}$ represents a *stochastic process* so that successive terms in this sequence—$u_{t-2}, u_{t-1}, u_t, u_{t+1}, u_{t+2}$, etc.—are each a *random* variable with a given probability distribution. The question is: whether these successive terms are *serially* uncorrelated or correlated. When the $\{u_t\}$ series are ***serially uncorrelated***, the past history of the $\{u_t\}$ series contains no information relevant to predicting its future values, i.e. u_t, u_{t+1} and u_{t+2}. Thus, conditional distribution of u_t is the same as its unconditional distribution, being a*ssumed to have zero mean,* and hence, $E_t(u_t) = 0$. For lack of space, we consider here the cobweb model with rational expectations under the first case only, i.e. when the $\{u_t\}$ series are ***serially uncorrelated***.[9] Then we find from (5s) that the rationally expected price level is just $\bar{p}$, the *equilibrium price* under deterministic cobweb model.[10]

Our main objective is to discuss the much familiar theory of the **Phillips curve** (PC) with its policy implications and, in particular, the chronological development of the literature on PC. Quite interestingly, the development of different models of PC follows almost the same path as that of the development of different models of expectations—*first*, the initial version of the PC introduced by Phillips himself without any expectations; *next*, the version of the PC with static or adaptive expectations mechanism—the so-called Friedman–Phelps version of PC and *finally*, the version of the PC with rational expectations—the neoclassical version of PC introduced by

[9]Note that when u_t's are *serially correlated*, the past realisations, u_{t-2}, u_{t-1}, etc., contain information which will be available at the end of the period in question and hence will be used—according to the rational expectations hypothesis—in forming expectations like $E_t(u_t)$, $E_t(p_t)$, etc., and in this case, $E_t(u_t)$ need not be zero, even if $E(u_t) = 0$ for each t.

[10]As we have seen, the cobweb model with *rational* expectations and the cobweb model with *other* expectations mechanisms (e.g. the static expectations, $p_t^e = p_{t-1}$) yield completely different results.

Lucas. We shall discuss all these versions in brief. But before that we need to discuss the background of the development of the Phillips curve (PC).

3 Phillips Curve—Development of Its Different Versions

One non-rigorous feature of the classical or even the Keynesian macroeconomic model was the "wage–price relation". Both the models assumed that firms operated on their labour demand curve, i.e. *cet. par*. employment could increase only if real wages decreased. As Blanchard (1990, pp. 782–4) writes, Keynes added the assumption that workers focused mostly on nominal wages so that nominal wages were more rigid, "sticky" (the word appears to be Keynes') than prices. In such a model then, an expansionary policy would raise aggregate demand and prices, thereby leading to a reduction in real wages and an increase in employment and output so that real wages were *countercyclical*. However, Dunlop (1938) contradicted this proposition by showing that (i) for the UK, real wages were, if anything, *procyclical* and that (ii) unions cared explicitly about the cost of living, i.e. real wages. As Blanchard writes, "These findings led many economists, including Keynes himself (1939), to conclude that a more drastic departure from classical theory was needed and that price-setting in particular could only be understood by appealing to imperfect competition".

A rigorous analysis of "wage–price" relation was thus called for. However, as Blanchard (1990, pp. 783–784) remarks, this task was not taken up by macroeconomists working within the "neoclassical synthesis"—the consensus view of macroeconomics which emerged in the 1950s and 1960s—due probably to *two* main *reasons*. The first was that it was hard, and the second was that less need was felt for this, owing to the discovery of the "Phillips curve", establishing a reliable empirical relation between the rate of change of nominal wages and rate of unemployment (Phillips 1958). Let us then discuss the Phillips curve in detail.

3.1 Phillips Curve—Its Initial Version (and also Mark-up Pricing)

By the early 1970s, there was a wide consensus as to the main empirical characteristics of the "wage–price mechanism". As summarised by Tobin (1972), this mechanism consisted of *two relations*, a price or "mark-up" equation and the "Phillips curve" explaining the wage change. The ***mark-up pricing*** **rule** for industrial products says that prices are set on the basis of costs per unit of output—but *not* the *actual* unit costs, but the *normal* unit costs, which was interpreted by many as unit costs at *capacity level* of production (these costs include wage costs as well as material and fuel costs). Let P_t, W_t, Y_t and N_t denote, respectively, the price level, money wage rate, level of output and level of employment in period t. Assuming labour to be the only variable input, the mark-up pricing rule is given by

$$P_t = \gamma(W_t/d)\ \big[\text{where}\ \gamma,\ \text{the mark-up factor (a}\ parameter)>1\big], \tag{18}$$

and where d stands for the average product of labour (APL), if the level of production were at full capacity; further, the APL at capacity production, d, is being assumed to be a constant over the short run we are considering. Taking logarithms of (18) and writing p_t for ln P_t and w_t for ln W_t, we have

$$p_t = \ln\gamma\ + w_t - \ln d = \mu + w_t,\ \{\mu \equiv (\ln\gamma - \ln d)\ \text{being a}\ parameter]. \tag{19}$$

The question now is: if output prices are determined by wages [and also by other input costs, not shown in the simple mark-up rule (18)], how is the wage rate to be determined? This brings us to the celebrated **Phillips curve** showing trade-offs between unemployment and wage inflation.

Phillips (1958) made a remarkable contribution in macroeconomics by developing and estimating an empirical equation relating the rate of change of money wages to the rate of unemployment. Phillips (1958) argued that wage rate (the price of labour services) moves just like the price of any other commodity or service; it is expected to rise (fall), if its demand were higher (lower) than its supply and the rate of rise (fall) would be greater, the greater the excess (deficiency of) demand.[11] Thus, one would expect a direct (i.e. positive) relation between the excess demand for labour and the change of money wages. However, the excess demand for labour is not directly observable. Lipsey (1960) provided a solution by developing a theoretical framework where it is argued that an inverse relation exists between the *excess demand* for labour and the *proportion* of the labour force *unemployed*, *u.* However, given that workers change jobs and that a finite time is taken to change, it follows that when excess demand for labour is zero, i.e. "when the wage rate is stable … there will be some quantity of unemployment" (Lipsey, *op.cit*., p. 14). To relate this discussion to the subsequent analysis in the literature, we shall call this quantity of unemployment (as a proportion of the total labour force, *LF*) the "*natural* rate of unemployment" *a la*' Friedman (1968)[12] and denote it by u^*. Lipsey's argument may then be put as follows. Wage rate will rise (fall) if u is less (greater) than u^*, and the extent of the rise (fall) will also depend on the extent of the discrepancy between u and u^*.

---

[11] "When the demand for labour is high and there are very few unemployed we should expect employers to bid wage rates up quite rapidly, each firm and each industry being continually tempted to offer a little above the prevailing rates to attract the most suitable labour from other firms and industries. On the other hand it appears that workers are reluctant to offer their services at less than the prevailing rates when the demand for labour is low and unemployment is high so that wage rates fall only very slowly" (Phillips 1958, p. 283).

[12] "At any moment of time, there is some level of unemployment which has the property that it is consistent with equilibrium in the structure of *real* wage rates" (Friedman 1968, p. 8). Friedman termed this level of unemployment (as a proportion of the total labour force) the *natural rate of unemployment* and argued that an unemployment higher (lower) than the natural rate would indicate that there was an excess demand for (supply of) labour that would produce upward (downward) pressure on real wage rate.

To facilitate some further manipulations, we shall take the following specific function for the "wage change–unemployment" relation which satisfies the above Phillips–Lipsey postulates, taking u^*, the *natural rate of unemployment*, to be a *constant* over time in our short-run analysis:

$$\frac{W_t}{W_{t-1}} = \left(\frac{1-u_t}{1-u^*}\right)^{\tau} \quad (\tau > 0). \tag{20}$$

It is now easy to verify that if $u_t < (>) u^*$, Eq. (20) ensures that W_t/W_{t-1} would be $> (<) 1$, i.e. wages would rise (fall) and the extent of such rise (fall) depends on the extent of the difference,[13] $|u_t - u^*|$. We shall now use the notations LF_t, N_t, N_t^*, Y_t, Y_t^* and P_t to denote, respectively, the labour force, the level of employment, the *natural level* of employment, the level of output, the *natural* level of output and the price level in period t, and use a lower-case letter to denote the logarithm of the corresponding upper-case letter (e.g. $y_t = \ln Y_t$). The Phillips curve, given in the "wage-adjustment-unemployment space" by Eq. (20), may be redefined in the "wage-adjustment-output space" by considering the following production function and the corresponding natural level of output:

$$Y_t = BN_t^{\theta} \quad \text{and} \quad Y_t^* = B\left(N_t^*\right)^{\theta} \ (B > 0 \,\text{and}, \ 0 < \theta \leq 1). \tag{21}$$

As $(1 - u_t)$ equals N_t/LF_t, using Eqs. (20)–(21), we may then write

$$\begin{aligned} \frac{Y_t}{Y^*} &= \frac{BN_t^{\theta}}{B(N_t^*)^{\theta}} = \left(\frac{N_t/LF_t}{N_t^*/LF_t}\right)^{\theta} = \left(\frac{1-u_t}{1-u^*}\right)^{\theta} \\ &= \left(\frac{W_t}{W_{t-1}}\right)^{\theta/\tau} \left[\text{as } = N_t^*/LF_t = (1-u^*)\right]. \end{aligned} \tag{22}$$

Converting Eq. (22) into logarithms, one can write the Phillips curve as

$$w_t - w_{t-1} = \beta\left(y_t - y^*\right) \quad [\text{where}\, \beta = \tau/\theta > 0]. \tag{23}$$

Expressing Eq. (19) in terms of first differences (with a constant μ), the rate of inflation equals the rate of change in (money) wage rate in a period t:

$$\pi_t = (w_t - w_{t-1}) \left[\text{where}\, \pi_t \equiv (p_t - p_{t-1}), \text{the rate of inflation in}\, t\right]. \tag{24}$$

By using (23)–(24), the Phillips curve [Eq. (23)] may now be expressed in the "inflation-output space"—the space where Phillips located it initially (Lipsey 2000, p. 61) and where it is being used nowadays—as follows:

[13] In fact, Phillips (1958, p. 290) estimated a relation comparable to our Eq. (20) (but without u^*) for the UK economy for the period 1861–1913. His estimated equation is:

(% rate of change of wage rates) + 0.900 = 9.638 (% rate of unemployment)$^{-1.394}$.

$$\text{(Phillips Curve: } \textbf{\textit{Initial Version}}\text{): } \pi_t = \beta\left(y_t - y^*\right) \quad (\beta \equiv \tau/\theta > 0). \tag{25}$$

Remark D: One clarification is needed. Usually, Phillips curve is presented as showing *trade-off between inflation and unemployment*, which means that *unemployment* can be reduced only if the economy bears a positive rate of *inflation*. However, we have expressed our Phillips curve equation [e.g. Eq. (25)] in terms of inflation and output above the natural rate. We shall, however, continue to use the term "*trade-off* between inflation and unemployment". And, in the present context, that is to be interpreted as meaning that *output above the natural level* (i.e. unemployment below the natural rate) can be achieved only at the cost of a *positive inflation*.

The celebrated trade-off between inflation and unemployment is clear from Eq. (25). We state this result formally as Proposition 1 below.

Proposition 1 (Phillips' Trade-off between Inflation and Unemployment) *If the authority wants to raise the level of output above y^* say, by the amount g (i.e. $y_t - y^* = g$), it has to take an expansionary policy (e.g. to raise the level of money supply at a given rate); then aggregate demand and output will increase, but as* Eq. (25) *asserts, that will be achieved only at a positive rate of inflation* (βg). *And this can continue as long as the authority continues to raise the money supply at the given rate every period. Moreover, if output is to be raised* further *above y^*, the economy has to bear a higher* rate of inflation *(higher than βg). Thus, there is a* **permanent trade-off** *between inflation and unemployment (as asserted by Phillips).*

3.2 Friedman (1968)–Phelps(1967) Expectations-Augmented Phillips Curve

The possibility of a long run/permanent trade off, as asserted by Phillips, was questioned by two economists—Phelps (1967) and Friedman (1968). Let us discuss Friedman's arguments in detail. Friedman (1968) argued that it is not true that workers do not care about real wages. When unemployment is sought to be reduced by demand-generating policies, inflation occurs and workers find their real wages falling. They would then try to protect their real wages. Thus, workers would go for wage negotiations, even if the level of employment (or output) was unchanged, but prices had been rising. However, when the workers go for wage negotiations, they cannot observe the future price change. They would, therefore, base their wage demands on their expected price change. Thus, the nominal wage change would be affected by *two* things: (i) expected price change and (ii) the employment/output change (a′ la′ the original Phillips curve). And, since the price change would be guided primarily by the wage change (via the assumed mark-up pricing mechanism), price change, i.e. the rate of inflation in period t (π_t), would be equal to the rate of *inflation expected* to occur in period t $\left(\pi_t^e\right)$ *plus* the amount caused by the *deviation of the actual output from the natural output* in period t (as is postulated in the initial version of the

Phillips curve). Formally, we may write, what is called the expectations-augmented Phillips curve (PC), as

$$\text{(Expectations-augmented PC):}\quad \pi_t = \pi_t^e + \beta\big(y_t - y^*\big) \tag{26}$$

The question is, of course, how expectation of inflation is formed. In the textbook discussion of the Friedman–Phelps model, two alternative expectations mechanisms are normally used, namely the *static expectations* and the *adaptive expectations*. Here, we shall illustrate the model under *static expectations* only. Thus, the Friedman–Phelps (expectations-augmented) PC (26) will read as

$$\text{(Friedman-Phelps-PC):}\quad \pi_t = \pi_{t-1} + \beta\big(y_t - y^*\big) \tag{27}$$

Now, considering Eq. (27) and the initial conditions (28) given below, one can get the results of this model quite easily, as we show now.

$$\text{(Initial conditions in period 0):}\quad \big[y_0 = y^*;\ \pi_0 = \pi_{-1} = 0\big] \tag{28}$$

Suppose the authority wishes to raise the level of output above y^* in period 1, by taking appropriate policy (say, by raising appropriately the amount of money): $y_1 = y^* + g$. However, as Eq. (27) shows, this can be achieved only at an inflation rate of $\pi_1 = \beta g$. Again, if the authority wants to keep the output at this higher level in period 2 as well, $y_2 = y^* + g$, then (27) shows that this can be achieved only at a higher rate of inflation: $\pi_2 = 2\beta g$. Continuing in this way, one finds that, to keep the level of output above y^* by the amount g every period, the rate of inflation will have to accelerate over time: $\pi_t = t\beta g$ (for $t \geq 1$). This is stated formally as follows.

Proposition 2 (Friedman–Phelps Expectations-Augmented Phillips Curve) *Friedman–Phelps* expectations-augmented *Phillips Curve,* Eq. (27)*, shows that to keep the output above the natural level by a given amount each period the economy has to bear an increasing rate of inflation over time.*[14]

Remark D: We thus find that the level of output *cannot permanently* deviate from the natural rate. There may be a trade-off between inflation and unemployment but that is *only in the short run*. And there is *no* such *trade-off in the long run*. Thus, the long-run Phillips curve is vertical at the natural level of output, y^*.

We now discuss the neoclassical version of the Phillips curve. This model was initially built up by Lucas (1973) for an economy with separately located markets. We, however, discuss it in an aggregative framework below. We discuss here its aggregative version, [see also Lucas (1972)].

---

[14]As Friedman (1968, p. 11) writes, "there is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off. The temporary trade-off comes not from inflation per se, but from unanticipated inflation, which generally means, from a rising rate of inflation. The widespread belief that there is a permanent trade-off is a sophisticated version of the confusion between 'high' and 'rising' that we all recognise in simpler forms. A rising rate of inflation may reduce unemployment, a high rate will not".

3.3 New Classical Version of Phillips Curve

As in the Friedman–Phelps model, new classical model also incorporates the expected rate of inflation in the Phillips curve. However, unlike the ad hoc expectation theory usually used in the Friedman–Phelps model, expectation is formed *rationally* in the new classical model. To state the model formally, the Phillips curve or the aggregate supply curve is the same as before, viz. Eq. (26) [which states, $\pi_t = \pi_t^e + \beta(y_t - y^*)$]. However, π_t^e is formed rationally, viz. $\pi_t^e = E_t(\pi_t)$, where $E_t(\pi_t)$ denotes the mathematical expectation of the distribution of π_t *conditional* on I_t (I_t being the set of information available up to, but not including, the current period t). Thus, the Phillips curve (i.e. the aggregate supply curve) in a given period t, $(\mathrm{AS})_t$, of the new classical school may be stated as follows[15]:

$$\text{(New Classical Phillips Curve) } (\mathrm{AS})_t\colon \pi_t = E_t(\pi_t) + \beta\left(y_t - y^*\right) \tag{29}$$

Now, the agents have to form *rational expectation* of π_t, i.e. $E_t(\pi_t)$, and for that they have to have some idea of an aggregative structure of the economy. Phillips curve, as we have stated earlier, is the *Aggregate Supply* curve. But to describe the economy's structure, we need at least an *Aggregate Demand* curve and a *Policy Rule*. For this purpose, we introduce *two equations*—(30) and (31)—described below in detail. Equation (30) represents a simple quantity theory-type *Aggregate Demand* curve for period t, $(\mathrm{AD})_t$ [introduced earlier as Eq. (12) to describe the economy in Example 1]: $M_t = P_t Y_t$. Recall that for this equation, the income velocity of money has been taken to be a constant and normalised at a value of unity. We now write this equation in logarithms so as to get $m_t = p_t + y_t$. Writing it further in first differences, and then subtracting y^* from both sides, we get

$$(\mathrm{AD})_t\colon y_t - y^* = y_{t-1} - y^* + (m_t - m_{t-1}) - \pi_t,\ (\pi_t \equiv p_t - p_{t-1}). \tag{30}$$

To introduce the second equation, note that in the stochastic world that we are considering the policy rule is *not completely* deterministic. Suppose, the authority wants to increase the money supply at a constant rate (say, g) every period from period 1 onwards as a demand-raising policy. However, the authority does not have complete control on money supply, and the actual rate of growth of money may deviate from g by some random element, v_t:

$$(\text{monetary } policy)\colon\quad m_t = m_{t-1} + g + v_t. \tag{31}$$

Here, g is ***policy determined*** but v_t is a ***random variable*** say, a ***white noise***.

However, the Phillips curve now contains rational expectation of π_t: $E_t(\pi_t)$. And, to find such a rational expectation for any period t, the agents have to have idea about

---

[15]By adding p_{t-1} to both sides of Eq. (29), it may be rewritten as $y_t - y^* = (1/\beta)[p_t - E_t(p_t)]$. This version is known as the famous Lucas supply function [see Blanchard and Fisher (1989, p. 358); but y^* is not mentioned there].

the entire structure of the economy, i.e. both supply and demand curves as well as policy—Eqs. (29), (30) and (31)—for that period. Now, to find $E_t(\pi_t)$, we have to first solve the first two equations for π_t. For this purpose, the expression $(y_t - y^*)$ in Eq. (29) is first substituted by the RHS expression of Eq. (30) so that one gets

$$\pi_t = E_t(\pi_t) + \beta\left[\left(y_{t-1} - y^*\right) + (m_t - m_{t-1}) - \pi_t\right] \tag{32}$$

Taking expectations of (32) conditional on I_t, we get
$E_t(\pi_t) = E_t(\pi_t) + \beta\left[(y_{t-1} - y^*) + E_t(m_t) - m_{t-1}\right] - \beta E_t(\pi_t)$ or solving,

$$E_t(\pi_t) = \left(y_{t-1} - y^*\right) + E_t(m_t) - m_{t-1}. \tag{33}$$

And we also write

$$\text{(Initial conditions)}: \left[y_{-1} = y^*, v_0 = 0, m_0 = m_{-1}\right] \tag{34}$$

$$\text{Equations (31)–(34) yield, } \{E_0(\pi_0) = 0, \pi_0 = 0\} \text{ and (29) yields, } y_0 = y^*]. \tag{35}$$

Note that the model is completely specified by Eqs. (29)–(35). To start with, we shall *assume* that the random term $v_t = 0$ for all t. Now, Eqs. (31)–(35) yield that in period 1, $E_1(\pi_1) = (y_0 - y^*) + E_1(m_1) - m_0 = g$.

Also, $(\mathrm{AS})_1$ and $(\mathrm{AD})_1$ curves, using the above results, may be rewritten as:

$$(\mathrm{AS})_1 : \pi_1 = E_1(\pi_1) + \beta\left(y_1 - y^*\right) = g + \beta\left(y_1 - y^*\right). \tag{36}$$

$(\mathrm{AD})_1$: $y_1 = y_0 + (m_1 - m_0) - \pi_1 = y^* + g + v_1 - \pi_1 = y^* + g - \pi_1$.

In Fig. 2, $(\mathrm{AD})_1$ and $(\mathrm{AS})_1$ curves intersect at the point N where $(y_1 = y^*, \pi_1 = g)$. Thus, output remains at the natural level, y^*; but the rate of inflation is now *positive* (i.e. g). Also, for any $t > 1$, one may find out that $(\mathrm{AS})_t$ and $(\mathrm{AD})_t$ curves [(29) and (30)] are the same as those for period 1 with $E_t(\pi_t) = (y_{t-1} - y^*) + E_t(m_t) - m_{t-1} = g$, if $y_{t-1} = y^*$ and $(\mathrm{AS})_t$: $\pi_t = E_t(\pi_t) + \beta(y_t - y^*)$; $(\mathrm{AD})_t$: $y_t = y_{t-1} + (m_t - m_{t-1}) - \pi_t$.

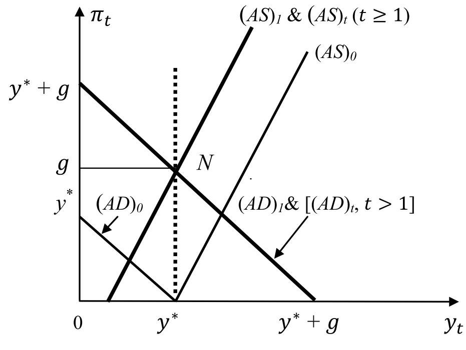


Fig. 2 New classical Phillips curve: **no** trade-off even in the short run

This may be argued easily. Take $t = 2$. Since $y_1 = y^*$, $E_2(\pi_2) = g$. Hence, $(\mathrm{AS})_2$: $\pi_2 = g + \beta(y_2 - y^*)$; $(\mathrm{AD})_2$: $y_2 = y^* + g - \pi_2$. These curves being the same as those for period 1 intersect at the same point ($y_2 = y^*$ and $\pi_2 = g$).

This argument may then be repeated for all subsequent periods. We may now state this result formally as Proposition 3 below.

Proposition 3 (New Classical Phillips Curve Results) *Consider an economy having the Phillips curve of the neoclassical type* [*i.e.* Eq. (29)] *and an aggregate demand curve, viz.* Eq. (30) *and suppose the economy is initially in equilibrium at the natural level of output y* with zero inflation. If the authority raises money supply at the rate g every period from period 1 onwards* [*with the stochastic term* $v_t = 0$ *for each t in* Eq. (31)], *the economy experiences inflation at the rate g every period but its output remains unchanged at the level* y^*. *Thus, there is* ***no trade-off even in the short run****. In fact, even if the stochastic term* v_t *in the money supply process* [Eq. (31)] *were positive (negative) in any period t, the* $(AD)_t$ *curve for that period would be above (below) the* $(AD)_1$ *curve, intersecting the* $(AS)_t$ *curve to the right (left) of the point N—the level of output would be higher (lower) than the natural level* y^* *at a rate of inflation higher (lower) than g. But* ***that trade-off could not be exploited*** *since it was purely random and the authority did not have any prior knowledge whether* v_t *was going be positive or negative in any period t. Thus, unlike in the case of Phillips Proposition 1 or Friedman–Phelps Proposition 2, new classical Phillips curve shows no trade-off even in the short run and even the* ***short-run Phillips curve*** *is* ***vertical*** *at the natural level of output.*[16]

4 Concluding Remarks

We have discussed different theories of expectations—static expectations, adaptive expectations and finally, rational expectations. We have also analysed the chronological development of the Phillips curve where these expectation theories have been used; we have shown that different results are obtained if one uses different expectations mechanisms. There are (***a***) many other aspects/implications of rational expectations used in a model [see articles in Lucas and Sargent (1981)] as well as (***b***) many other extensions of/additions to the Phillips curve. To consider case (***a***) first, models with rational expectations yield many interesting results, e.g. Lucas critique [Lucas (1976)], rules versus discretion—which policy is better, etc. On the other hand, New Keynesians point out that if there are wage and/or price rigidities, policies would affect output and employment even in models with rational expectations [see, for instance, the articles in Mankiw and Romer (1991)]. To talk about case (***b***) now, while Friedman–Phelps considered the natural rate of unemployment, research

[16] One may also incorporate in ***each*** of our previous two models (viz. original Phillips curve and the Friedman–Phelps model), the $(\mathrm{AD})_t$ curve given by Eq. (30) and draw the two curves in a diagram like Fig. 2 and then derive those results in the way done here. We leave it as an exercise for the students.

on the Phillips curve particularly by the New Keynesians has focused on the concept of the ***non-accelerating inflation rate of unemployment*** (NAIRU). [For instance, comparable to Eq. (26), the Phillips curve, according to this approach, would read as: $\pi_t = \pi_t^e - \alpha(u_t - u^*)$, where u_t denotes the actual rate of unemployment and u^* the NAIRU; see Fuhrer et al. (2009) for further observations on it and references.] Finally, whereas the new classical economists deny, while the New Keynesian economists, by bringing in some kind of rigidity in the model, believe in, positive roles of policies, the latter have developed the so-called **New Keynesian Phillips curve**—popularly known as NKPC—which relates actual rate of inflation in a period to (i) its *expected rate of inflation* in the next period as well as to (ii) the *output gap* (i.e. the *deviation* of actual output from the equilibrium level of output under flexible prices). This model, built also on microfoundation, has a structure different from that of a new classical model (see, for instance, Gali 2008, Chap. 3 and a note by Whelan 2005, for the derivation of this model and further discussion). There are many other extensions as well. All these are very interesting topics. However, shortage of space forces us to conclude our discussion here.

References

Blanchard, O. J. (1990). Why does money affect output? A Survey. In B. M. Friedman & F. H. Hahn (Ed.), *Handbook of monetary economics* (Vol. II, Chap. 15). Elsevier Science Publishers B. V.

Blanchard, O. J., & Fisher, S. (1989). *Lectures on macroeconomics*. Cambridge: Massachusetts Institute of Technology.

Cagan, P. (1956). The monetary dynamics of hyperinflation. In M. Friedman (Ed.), *Studies in the quantitative theory of money*. University of Chicago Press.

Carter, M., & Maddock, R. (1984). *Rational expectations: Macroecnomics for the 1980s?*. London: Macmilan.

Chakraborty, S. R. (2002). *Microeconomics*. New Delhi: Allied Publishers Pvt. Ltd.

Friedman, M. (1968, March). The role of monetary policy. *American Economic Review, 58,* 1–17.

Fuhrer, J., Kodrzycki, Y. K., Little, J. S., & Olivei, G. P. (2009). The Phillips curve in historical context. In J. Fuhrer, Y. K. Kodrzycki, J. S. Little, & G. P. Olivei (Eds.), *Understanding inflation and the implications for monetary policy: A Phillips curve retrospective*. Cambridge: The MIT Press.

Gali, J. (2008). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework*. Princeton and Oxford: Princeton University Press.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. The Macmillan Press Limited.

Keynes, J. M. (1939). Relative movements in real wages and output. *Economic Journal, 49,* 34–51.

Lipsey, R. G. (1960, February). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862–1957: A further analysis. *Economica, New Series 37*(105), 1–31.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory, 4,* 103–124.

Lucas, R. E. (1973). Some international evidence on the output-inflation trade-offs. *American Economic Review, 63,* 326–334.

Lucas, R. E. (1976). Econometric policy analysis: A critique. *Carnegie-Rochester Conference on Public Policy, 1,* 19–46. 19. (Available in the book by Lucas, R. E. (1981). *Studies in business cycles theory*. Cambridge, MA: MIT Press).

Lucas, R. E., & Sargent, T. J. (Eds.). (1981). *Rational expectations and econometric practice*. London: George Allen and Unwin.

Mankiw, N. G., & Romer, D. (Eds.). (1991). *New Keynesian economics.* In *Imperfect competition and sticky prices* (Vol. 1). Cambridge: The MIT Press.

Metzler, L. (1941, August). The nature and stability of inventory cycle. *Review of Economics and Statistics, 23,* 113–129.

Muth, J. F. (1961). Rational expectations and theory of price movement. *Econometrica, 29,* 315–335.

Phelps, E. S. (1967). Phillips Curves, expectations of inflation and optimal unemployment over time. *Economica, 34,* 254–281.

Phillips, A. W. (1958, November). The Relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica, 25,* 283–299.

Tobin, J. (1972). The wage price mechanism: Overview of the conference. In O. Eckstein (Ed.), *The econometrics of price determination* (pp. 5–15). Board of Governors of the Federal Reserve System.

Whelan, K. (2005). *Topic 7: The New-Keynesian Phillips curve*. https://www.tcd.ie/Economics/Staff/whelanka/topic.pdf.

Keynes, Working Hours, Lifestyles and the Environment: A Note

Soumyen Sikdar

Abstract In 1931, in the midst of serious economic downturn, John Maynard Keynes published a short paper, 'Economic Possibilities for our Grandchildren'. On the basis of an insightful discussion of the determinants of future economic growth, he predicted substantial improvement in living standards over the next hundred years. Those improvements, he claimed, would herald the end of the economic problem. One component of the higher quality of life would be a drastic reduction in working hours for everybody. He put it at 15 hours a week. Freed from the crippling drudgery of work, the typical person would have ample time to utilise her leisure virtuously and well. In reality, although the pace of economic growth has far outstripped Keynes' predictions, the hope for increase in leisure has not happened. The economic problem has not been conquered, and man is as hard pressed for time as in 1931, if not more. Working hours have declined significantly only in a very small number of even the more advanced countries. Higher leisure consumption would have shifted demand away from material production, and this would have been a boon from the point of view of preserving the world's resources and environment. This paper explores some aspects of the 'Keynesian failure' and the consequences for our grandchildren.

Keywords Income growth · Leisure · Environment

In 'A Tract on Monetary Reform', John Maynard Keynes famously stated that in the long run, we are all dead. But, assuming rightly that grandchildren will be alive, early in 1928 at the age of forty-five, he wrote a short essay, titled Economic Possibilities for Our Grandchildren, that was read to a student society of the elite public school in Winchester. He was still at work revising and refining the essay when Wall Street crashed horrendously in the fall of 1929. Though he quickly recognised that the slump 'will take its place in history amongst the most acute ever experienced', it did not distract him away from extrapolating the trajectory of Western-style capitalism and contemplating a future where mankind will have all the material goods it could possibly want. Capitalism, driven by relentless accumulation and the lust for profit, was a means—a rather distasteful means, according to him—to this end.

S. Sikdar (✉)
Indian Institute of Management Calcutta, Kolkata, India
e-mail: ssikdar@iimcal.ac.in

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_2

The essay was finally published in 1930. Biographer Robert Skidelsky noted that between February 1928 and June 1930, Keynes presented the paper in front of five different audiences and put it in print twice. It was also included in Essays in Persuasion of 1931. Evidently, it was taken seriously by the author.

'Let us, for the sake of argument', he wrote, 'suppose that a hundred years hence we are all of us, on the average, eight times better off in the economic sense than we are today'. In such a world of abundance, the perennial economic problem of scarcity, he believed, will wither away. All the 'absolute needs' will be satiated, and 'relative needs'—keeping up with the Joneses—will become of a second order of importance. Highly efficient labour-saving technology will drastically reduce the need for human labour, and this will be consistent with the preference pattern of affluent individuals who will substitute leisure for irksome labour. The normal working week will not exceed fifteen hours.

But then the real problem will begin. To quote from the essay: 'Thus for the first time since his creation man will be faced with his real, his permanent problem—how to use his freedom from pressing economic cares, how to occupy the leisure which science and compound interest will have won for him, to live wisely and agreeably and well. The strenuous purposeful money-makers may carry all of us along with them into the lap of economic abundance. But it will be those peoples, who can keep alive, and cultivate into a fuller perfection, the art of life itself and do not sell themselves for the means of life, who will be able to enjoy the abundance when it comes'.

Coming from one of the most brilliant minds of the modern age, the essay has engaged the attention of economists and other social scientists. Most readers will be surprised at both how accurate Keynes was in his forecast of the growth of income levels over the next century and how wildly wrong he was in speculating about working hours. The reactions of some major economists of today are on record in the excellent volume, Revisiting Keynes, edited by Pecchi and Piga (2010).

One major flaw in Keynes' vision is the total neglect of distributional issues and their consequences for economic well-being. As Benjamin Friedman writes about the USA in the Pecchi–Piga volume: 'more unequal distribution has prevented the great majority of the nation's families from seeing any increase in real terms'. Stiglitz also points out that the demand for leisure may not have increased in America the way Keynes anticipated because real wages, for most workers, have not in fact increased. Another flaw is the failure to recognise that both 'absolute needs' and preferences in general may be endogenous, susceptible to manipulation by commercial sales promotion. As 'needs' kept growing in an unbounded manner, so did the need to earn more and hence working hours could not decline. This is aptly expressed by Becker and Rayo, in the same book: 'After an initial period of excitement, the average consumer grows accustomed to what he has purchased and rapidly aspires to own the next product in line'. This seems to be hardwired into us. Humans evolved 'so that they have reference points that adjust upwards as their circumstances improve'.

Keynes has been faulted for his failure to recognise that demand for quality is universal and inexhaustible and will forever continue to provide stimulus for more earning. But here the tragedy is that 'quality' has come to be accepted in a very narrow materialistic sense. Better quality basically means additional features of existing products. And usually, it comes with a price tag. Consuming more (and more and more, driven both by the principle of keeping up with the Joneses and the force of effective sales promotion) products (and services) which are of better quality in this narrow sense is indeed a far cry from 'cultivating to a fuller perfection the art of life itself'. To quote Zilibotti (Pecchi–Piga, Chap. 2): 'Keynes anticipated that the progressive satiation of material needs would naturally generate a shift of preferences, whereby people would become better at appreciating arts and beauty. Moral values would also change, and the obsession for money-making would be replaced by a new humanism'. Plainly, this anticipation has woefully failed to pass the test of time. Emphatically emblematic of the emphasis on quantity at the cost of quality is the alarming problem of obesity in the developed world, in the USA in particular.

Technological progress at a rapid rate on a very wide spectrum of activities has increased hourly productivity quite dramatically. Productivity improvement can be useful in two ways. One can produce and consume more working the same number of hours or produce and consume the same amount while working fewer hours and enjoying more free time. Overall life satisfaction depends on both consumption and free time. Many economists such as Blanchard think that USA and some European countries have chosen different paths to happiness. The difference in output per capita may be partly due to the fact that relative to the USA, France or the Netherlands, for example, have used the growth in productivity to enjoy more free time rather than consumption. In 2014, the average employed American worked roughly a hundred and forty hours more per year than the average Englishman and three hundred hours more than the average Frenchman. An excellent analysis of the secular decline in leisure for Americans caught in the endless 'work-and-spend cycle' is contained in the book by Juliet Schor (1992).

Table 1 gives an idea about working hours in a number of developed countries.

Table 1 Average annual working hours per worker

Country	2000	2005	2010	2017
Australia	1780	1732	1700	1676
Austria	1654	1612	1532	1487
Denmark	1466	1451	1422	1408
France	1550	1527	1528	1514
Japan	1821	1775	1733	1710
Netherlands	1462	1434	1421	1423
Norway	1455	1423	1415	1419
Sweden	1483	1449	1476	1453
Switzerland	1713	1690	1624	1570
UK	1539	1515	1476	1514
USA	1832	1794	1773	1780

Source OECD.stat

The competitive pressures unleashed by globalisation are also responsible in a big way for inducing longer hours of work in the developed countries. In their battle for survival against their lower-wage counterparts in countries such as China, India and Mexico, workers in these countries had to accept wage cuts as well as shorter free time. In 2006, Germany's IG Metall union, an active defender of workers' rights in the German automobile industry, agreed to extend the 29-h workweek to 33–34 h at Volkswagen plants without any pay rise. Without the agreement, the company would have laid workers off and shifted operation to lower-cost European or Asian countries.

In the rest of this note, we shall focus on one component of 'life satisfaction' that was ignored by Keynes, but which has come to assume greater and greater importance in recent decades. This is the link between lifestyles and environmental quality. The issue is of moving away from the traditional mode of exploiting nature ruthlessly for material consumption towards a lifestyle which involves greater care of nature and natural resources. In their book 'How Much is Enough', Robert and Edward Skidelsky (2012) list seven elements of the good life: health, security, respect, personality, harmony with nature, friendship and leisure. Leisure in an affluent society could have been enjoyed in peaceful and contemplative harmony with nature in the company of family and friends. Instead, the vacation and entertainment industries have moved in on a grand scale to ensure that even our 'free time' is packed with the consumption of goods and services.

Since clean environment is an income-elastic superior good, richer countries, compared to poorer ones, have put in better mechanisms to manage environmental resources such as soil, forest, water and biodiversity. However, they emit more CO_2 per capita. Scatter plot of CO_2 emissions per capita against real GDP per capita of countries has been studied for many time periods. Invariably, the line of best fit has an upward slope. More affluent countries are also the more polluting ones. Interestingly, the better performers (lying below the best fitting line—with per capita income measured on the horizontal axis) are also by and large the countries that have chosen to take the 'more free time' route to happiness. These include France, Germany, the Netherlands, Norway, Sweden and Switzerland.

This rising 'social preference' for leisure has strong positive implications for the environment because free time is generally less consumption-intensive. Let us consider a dramatic example. Workers in the Netherlands worked fewer than half as many hours in 2000 as they did in 1900 and enjoyed lot more free time. Had they continued to work more than three thousand hours a year as in 1900, their total earning and consumption expenditure and the resultant adverse impact on the environment would have been much higher. The re-orientation of Dutch preferences has certainly helped in a big way in their commitment to conserve nature.

Encouraging innovations in green technology is part of the environmental policy of many nations today cutting across levels of affluence. This supply-side attack is something very positive and should by all means be encouraged. If it can work hand in hand with a shift towards 'greener' lifestyles, particularly in the affluent consumerist societies, chances for reversing (or at least arresting) the ongoing trend in environmental destruction will improve enormously. And we see no reason to disbelieve that the author of Economic Possibilities, despite all his upper-class British prejudices of the day and fashionable fondness for 'elite communism', would have heartily applauded such a change.

References

Pecchi, L., & Piga, G. (2010). *Revisiting Keynes: Economic possibilities for our grandchildren*. Cambridge, MA: MIT Press.
Schor, J. (1992). *The overworked American: The unexpected decline in leisure*. New York, NY: Basic Books.
Skidelsky, E., & Skidelsky, R. (2012). *How much is enough? Money and the good life*. Baltimore, MD: Penguin Books.

A Perspective on the Slowdown in Private Corporate Investments in India

Ananya Ghosh Dastidar and Rashmi Ahuja

Abstract Investment is universally acknowledged as one of the chief drivers of economic growth; private corporate investment is a key component of aggregate demand as well as an important driver of productive capacity, on the supply side. In India, aggregate investments touched a peak of nearly 36% of GDP in 2007–2008, declining to 34% in 2012 and further to around 28% of GDP in 2016. Even at the current conjuncture, lacklustre recovery in Indian business investments remains puzzling, especially with recovery well under way in the global economy. This paper attempts to analyse the slowdown in private corporate investments in India in the post-2008 crisis phase, taking into account both demand as well as supply-side factors. The first part of the paper attempts to identify the main determinants of business investments by drawing on economic theory and relevant literature in the Indian context. Thereafter, an expository data analysis is carried out on key correlates of corporate investments in India with a view to identifying emerging trends especially since the middle of the 2000s decade. A simple empirical model is also estimated, using annual aggregate data for the period 1995–96 to 2016–17, in an attempt to confirm some of the insights emerging from the literature. Results indicate that uncertainty in the overall macroeconomic and business environment, demand-side factors, especially external demand and real interest rates, and the pace of public investments plays an important role in affecting private business investments in India.

We are deeply honoured to be able to contribute to this volume in honour of Professor Sarmila Banerjee, who is a very fine economist, a wonderful human being and a superb teacher. She has been an inspiration to generations of students. We would like to thank the reviewers for their valuable comments and suggestions and the editors for providing this opportunity. The usual disclaimer applies.

A. G. Dastidar (✉)
Department of Business Economics, University of Delhi South Campus,
1st Floor, Arts Faculty Building Benito Juarez Road, New Delhi 110021, India
e-mail: agdastidar@gmail.com

R. Ahuja
Department of Economics, Lal Bahadur Shastri Institute of Management, Dwarka, Sector 11, New Delhi, India
e-mail: rahuja2012@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_3

Keywords Private corporate investment · India · Real interest rate · Expectations

1 Introduction

This paper sets out to examine the key determinants of private investments in India and seeks to explore to what extent these can explain the investment slowdown that became evident in the aftermath of the global financial crisis. The world economy has largely recovered since the crisis, led by fall in unemployment and strong growth recovery in the USA, but the lacklustre performance of private corporate investments in India against these positive developments in the global scenario remains a puzzle. The Indian government undertook major policy initiatives like 'Make in India' to boost manufacturing growth, kept a strict vigil on fiscal deficit and inflation targets and strived to keep the overall macroeconomic environment conducive for investment growth. Yet these measures failed to revive and spur growth in private investments which is essential for growth and job creation. Gross fixed capital formation (GFCF) in the private sector in India fell from 26.9% of GDP in 2011 to 21.7% in 2015–16 and robust signs of recovery remain elusive till date.[1] It has been argued that policy shocks such as demonetisation in November 2016 and major structural reforms such as the introduction of the goods and services tax (GST) in July 2017 slowed down recovery in private investments. Yet the duration of the slowdown, which set in well before these policy shocks came and have continued even as these seem to have been absorbed,[2] suggests that other factors have also been at play.

The paper is structured as follows. Analysis of the chief determinants of private business investments outlined in the theoretical literature is followed by a discussion of main findings in the context of India. The literature on determinants of private investments in India as well as specific explanations for the current slowdown is explored in an effort to identify policy takeaways to address this problem. Based on the review of literature, expository data analysis is carried out on key correlates with a view to identifying the factors affecting private corporate investments in India. A simple empirical model is also estimated in an attempt to confirm the key insights emerging from the literature. The main findings and policy conclusions therefrom are presented in the concluding section.

---

[1] A recent study by the RBI showed a steady decline in phased capital expenditure plans of the private corporate sector in absolute terms from 2010–11 down to 2017–18 (reported in The Indian Express, dated 13 March 2019).

[2] Analysis of firm-level data does indicate signs of a turnaround in net sales of listed firms since the last quarter of 2017 after a long period poor performance (quarter on quarter growth in net sales of listed firms were either negative or below 1%, from 2011 up to the third quarter of 2017) (Patnaik 2019). However, it is too early to say to what extent this would usher in a pickup in business investments.

2 Determinants of Business Investments: Lessons from Theory and Evidence from India

2.1 Theoretical Underpinnings

Standard macroeconomic theory suggests that the opportunity cost of investible funds is one of the most important determinants of aggregate private investments in an economy.[3] An increase in the rate of interest (the lending rate), ceteris paribus, lowers the net present value of future income flows from investment projects, thereby reducing planned investment by firms. In a micro-theoretic framework with profit maximising firms and perfectly competitive markets, the optimal level of investments is determined by the equality between user cost of capital and value of the marginal product of capital. The inverse relation between the real interest rate, price of capital goods and business investments emerges clearly in this framework. Also based on a micro-theoretic framework is Tobin's q theory of investments which suggests that firms' market value or stock market valuations of their shares and net debt are also likely to affect investment decisions. That is, when a firm's q ratio (ratio of stock price to the replacement cost of capital) exceeds one it creates an incentive for increasing capital formation. In this case, higher returns on capital (proxied by the stock value of the firm) exceed the cost of acquiring the same, creating an incentive to undertake further capital formation.

With imperfect competition and price-setting firms, the level of demand also affects investment decisions as prices are set in line with expectations regarding sales. As per the simple accelerator theory, firms' desired capital stock is proportional to their expected change in output levels. This clearly suggests that change in GDP is an important determinant of aggregate investment which is an addition to the existing stock of capital. However, the process of adjustment of actual capital stock to the desired level is not costless as explained by the flexible accelerator theory. In the presence of capital adjustment costs, investment is affected directly by the existing stock of capital as well as the level of output, while policy variables such as investment tax credits and interest rates also have an indirect impact as they affect adjustment costs.

Apart from firms' stock values, the volume of their internal cash flows and retained earnings are also important drivers of their investment decisions. In particular, under asymmetric information and credit market imperfections, firms face borrowing constraints. This is quite common in developing economies, where high costs of selection and monitoring tend to limit the extent of bank credit, especially to small and medium enterprises. In such cases, firms' internal funds assume immense importance in financing capital formation. Further, internal funds are also cheaper compared to the costs associated with leverage.

The larger macroeconomic scenario and institutional factors also affect private investment decisions. For one, in developing economies public investments in infras-

[3]The following discussion is based on Carlin and Soskice (2006) and Sikdar (2006).

tructure may be instrumental in alleviating supply bottlenecks and encouraging private investments. That is, there can be a case for 'crowding in' rather than 'crowding out' of private investments due to the rise in government spending. Overall, macroeconomic stability can also matter a great deal by shaping business confidence and influencing 'animal spirits', originally highlighted by Keynes as an important driver of investment decisions by private firms. In particular, firms' expectations regarding the future is extremely important as the decision to invest today is essentially shaped by an expected future stream of returns. Therefore, uncertainties associated with political instability, lack of clarity about future policy directions can end up having disproportionately large impacts on private investments.

In the context of developing and emerging countries, certain structural characteristics shape economic outcomes and should be taken into account in economic analyses. For instance, certain factors can hinder the inter-sectoral flow of resources in such countries. Broadly, these can be captured under the category of institutional and regulatory factors that create hidden costs and entry barriers (e.g. bureaucratic red-tapism, complicated regulatory compliance requirements, etc.). Effectively, these forces prevent equalisation of economy-wide factor returns, and owing to these rigidities, some sectors could be facing demand constraints and others supply constraints within the same economy. Typically, market-based economic reform programmes attempt to reduce and ultimately eliminate such bottlenecks; however, their impact can be time-consuming and the induced behavioural changes may take effect only in the very long term. In the Indian context, the biggest wave of reforms was introduced in 1991. These initially targeted the external and production sectors, through policies of trade liberalisation, privatisation and deregulation. Alongside, in gradual steps, financial sector reforms were put in place with a focus on banking, deregulating capital markets and easing access to global finance. By and large, there is a consensus that the reforms have been instrumental in unleashing the growth potential of private businesses in India. Currently, the need for a second wave of reforms on simplifying regulatory clearances and procedures, easing access to credit and labour reforms is also being felt and widely articulated across public platforms and by media and industry lobby groups.

2.2 *Empirical Studies on Determination of Private Corporate Investments in India: A Review*

In what follows, we examine the literature in the Indian context to see what the main drivers of private investment behaviour are and whether these have changed over time.

Findings from empirical studies in the Indian context indicate the importance of conventional theories in explaining the behaviour of private corporate investment.[4]

[4]See Goldar (1997) for a survey of early studies on investment demand function estimation for India.

The user cost of capital has an important influence on private investments, evident in both early studies (e.g. Sarma 1988), as well as later ones covering the post-liberalisation period (Athukorala and Sen 2003). An analysis (Athukorala and Sen 2003) of the surge in private corporate investments during 1991–1996 suggests that the main driver was the sharp fall in the cost of capital (driven by a reduction in import tariffs on capital goods) between 1985–90 and 1990–96. Indeed, this effect was strong enough to overcome the negative impact of the slowdown in public investments and a rise in bank lending rates witnessed after the external crisis in 1991. This underscores the importance of user cost of capital as a key driver of private investments in India.

Changes in aggregate income also affect investments, underscoring the importance of demand-side factors in the Indian context (Krishnamurty and Sastry 1975; Athukorala and Sen 2003; Rajakumar 2005), and there is evidence pointing to the importance of public investments in affecting capital formation by private businesses. For instance, Athukorala and Sen (2003) find that private investment increases: (i) the lower is the user cost of capital, (ii) higher the increase in GDP, (iii) higher is a public investment and (iv) lower the existing stock of capital. In terms of magnitude, the user cost of capital is found to have the largest impact followed by income change and public investment in this study. Studies also find the presence of significant adjustment lags and delivery lags in the process of capital formation (Sarma 1988; Athukorala and Sen 2003) so that present investment is affected significantly by changes in the levels of output in the past.

Regarding the financing aspect, the importance of both external finance as well as retained earnings has been established for different industries in an early study using industry-wise data, which also provided evidence of competition for funds between fixed and inventory investment (Krishnamurty and Sastry 1975). Findings from studies using firm-level data indicate the importance of internal funds depends on firm size, with larger firms relying more on this source compared to smaller ones. Also, heterogeneity in terms of access to finance within a size class of firms indicates the presence of asymmetric information that creates borrowing constraints for 'lesser-known' firms (Athey and Reeser 2000). Firm-level studies also find that investment behaviour differs by type of financing availed. Equity-financed firms face capital market imperfections and rely relatively more on internal funds compared to ones that are debt-financed (Rajakumar 2005). Therefore, the literature indicates the presence of financing constraints both for debt and equity-financed firms in India especially owing to various forms of market inefficiencies.

2.3 Studies on the Current Investment Slowdown in India: A Review

A number of explanations have emerged in the context of the slowdown in private investments that set in since 2011–12 (Rakshit 2016). Many of these are embedded in the nature of the boom in private investments that was witnessed between 2003 and

2007. The low cost of capital has been emphasised as one of the factors underlying the phenomenal growth in private corporate investments in this phase. On the supply side, a plentiful supply of domestic savings and a surge in capital inflows ensured easy availability of credit to the industry at relatively low rates and contributed to the boom (Nagaraj 2013). On the demand side, this was the period of robust growth in world trade, and the stellar performance of Indian exports led to this phase being termed as 'export-led' growth.[5] However, economic growth faltered with the onset of the subprime crisis in 2007–08, which led to a withdrawal of the external demand stimulus, especially from the developed country markets. A coordinated fiscal expansion by developing country governments triggered a recovery, and once this was underway, compulsions of prudent macro-management policy soon restored priority on fiscal consolidation.

Indeed, the tight monetary stance adopted by the Reserve Bank of India since early 2010[6] is considered as one of the triggers of the investment slowdown. However, this is by no means the only factor; the importance of demand side factors and the role of several structural problems have also been emphasized in this context. The unprecedented private investment boom which took place earlier since 2003–04 resulted in large scale capacity creation.[7] Utilisation of capacity requires fast-paced demand expansion, a process that was hindered at least in part due to policy focus on fiscal consolidation and monetary tightening. In the absence of policy-induced positive demand shocks, the external sector, characterised by lacklustre export growth and stagnant world demand, failed to bridge the shortfall in domestic demand. Further, lack of investments in agriculture constrained demand growth of a large segment of the population drawing their livelihoods from the rural economy. The neglect of agriculture also allowed bouts of food inflation to dampen demand for non-agricultural commodities. Certain structural characteristics of the economy such as widespread inequality and existence of a large informal sector, that have persisted even after two decades of economic reforms, also weakened the domestic demand response. Investment itself being a component of aggregate demand, the slowdown compounded the problem—while positive gross investments meant capacity expansion continued, aggregate demand would fall by a multiple of the reduction in gross investments, reinforcing capacity underutilisation.

The sectoral pattern of the investment boom also contributed to the subsequent slowdown. Between 2000 and 2013 growth of capital formation in infrastructure outpaced that in the other sectors of the economy. Since capacity utilisation in infrastructure is wholly dependent on growth in demand in other sectors of the economy and since these cannot be exported, any shortfall in domestic demand can lead to prolonged slump in new investments in this sector.[8] In turn, the slowdown in infras-

[5]See Dastidar (2015) for the argument that this episode is better described as 'export induced' growth since it resulted from favourable world market conditions, rather than from directed state policies as in case of the East Asian episode of 'export-led' growth.

[6]The repo rate touched a high of 8.5% in the third quarter of 2011–12 (Rakshit 2016).

[7]The following discussion draws heavily on Rakshit (2016).

[8]Till 2014, 60% of stalled projects were in infrastructure (Rakshit 2016).

tructure investment transmits a negative demand shock to the rest of the economy, contributing to the investment slowdown in the other sectors, which would in turn affect growth and profitability in infrastructure itself. In this way, a downward spiral could feed into itself and persist over a long duration.

Another structural problem relates to the nature of the financial and real sector inter-linkages that developed during the investment boom and in its aftermath. It is linked to the pattern of financing, which led to a substantial build-up of corporate debt during the investment boom and also under the policy stimulus as capacity creation was incentivised via access to cheap credit from domestic and external sources (Nagaraj 2013; Rakshit 2016). Indeed, analysis of firm-level data suggests that rise in the burden of debt servicing, owing to high levels of leverage, is one reason Indian corporates are devoting a smaller share of profits for physical investments[9] (Sen and Dasgupta 2018).

With the onset of the downturn, the cost of debt servicing ate into firms' profitability and eroded their ability to service debt, thereby transmitting the problem to the financial sector in the form of debt defaults and rising burden of non-performing assets (NPAs). This 'twin balance sheet problem' plaguing the balance sheets of private business and banks accentuated the operation of a borrowing constraint and contributed to the persistence of the slowdown in new investments. This is identified as a key bottleneck that needs to be addressed via institutional reforms to revive and strengthen credit flow for private business investments in the Indian context (Ahmad et al. 2018).

Yet another perspective on the slowdown draws on developments in the financial sector of the Indian economy which experienced a rapid pace of growth, the emergence of a plethora of financial instruments and a much higher degree of integration with global markets. It has been argued that 'financialisation' of the Indian economy has tilted domestic firms' incentive structures in favour of investment in short-term, financial assets with risky but high returns and away from investment in physical assets (Sen and Dasgupta 2018). High returns on mutual funds and trade in equity (in the secondary market) provided managers with the opportunity to enhance shareholders' profits in the short term by investing in financial assets at the cost of physical capital formation that yields returns only in the long run. The link between managers' compensation and firms' stock values would create a natural incentive for such a trend. Overall, both aggregate- and firm-level data show a fall in the share of physical assets and a rise in the share of short-term assets in total assets of Indian firms over the 2000s decade.

Policy uncertainty has also been seen as an important factor affecting the investment slowdown (Anand and Tulin 2014; Ahmad et al. 2018). In particular, uncertainty in the policy environment can have a negative impact on capital formation by the private sector through adverse effects on business expectations. Sen et al.

[9] Sen and Dasgupta's (2018) analysis of firm-level data shows the declining importance of equity finance, stagnant reserves (especially since 2008) and increasing importance of borrowings, with a rising share of overseas credit among Indian private (non-financial) firms over the 2000s decade. There is even evidence of Indian firms taking on more debt to meet interest liabilities on a significant scale.

(2014) add to this literature as they delineate the role of deeper institutional factors at play in explaining episodes of a growth slowdown. They argue that the processes of institutional change and spurts in economic growth are nonlinear with complex feedbacks into each other. In developing countries with relatively weak formal institutions, the process of growth may not always lead to an improvement in institutional quality owing to problems such as rent-seeking and corruption with associated consequences involving prolonged litigation and policy reversals. The resultant uncertainties inevitably have negative repercussions for growth. Indeed, under such circumstances, the slowdown is likely to persist, given the deep-rooted institutional correctives needed for recovery of the growth momentum.

3 Evidence on the Current Slowdown

3.1 Potential Determinants of Private Corporate Investments in India: Recent Trends

What is the recent body of evidence on the core determinants of private business investments in India? Analysis of recent trends reveals an interesting picture.[10]

The slowdown in private corporate investments is clearly evident from the sharp dip in the rate of growth of corporate investments after 2007–2008, following the global financial meltdown. This was followed by recovery, influenced by factors such as expansionary fiscal measures adopted by the government in the wake of the financial crisis. However, post-crisis growth rates have remained subdued, way below the pre-crisis levels (Fig. 1). Further, the policy of demonetisation and the implementation of GST soon afterwards brought about a dip in growth rates[11]; however, the effects of these shocks may have been transient as a pickup in investment rate is evident from the second quarter of 2018 (Ahmad et al. 2018).

Quarterly data is available from 2008 on capacity utilisation, an important outcome variable associated with the pace of investment growth. This clearly shows a downward trend, with a pickup visible only since 2018 (Fig. 2). These trends are in sync with Rakshit's (2016) argument regarding the role of demand-side factors. Commonly, a prolonged slowdown tends to be preceded by an economic boom, during which sustained demand growth creates positive expectations and fuels capacity

[10]Despite comparability issues of the new GDP series (2011–12 base year) with previous 2004–05 data, the trends based on the spliced data reveal a few interesting insights. Without splicing the two series, it is not possible to generate a continuous time series data including the period post-2012–13. While the exact values in the spliced series differ from the 2004–05 series, the basic pattern regarding year-on-year growth is the same. Also refer to Footnote 17 in this regard.

[11]From the third quarter of 2016–17 to the first quarter of 2017–18, quarterly GDP growth slipped below 7%, growing at 6.8, 6.1 and 5.6% respectively over these three quarters that witnessed major policy shocks by way of demonetisation (in November, 2016) and introduction of the GST regime (in July, 2017) (Ahmad et al. 2018).

Fig. 1 Real private corporate GFCF: year-on-year growth (2011–12 base). *Source* Authors calculations based on CEIC India Database

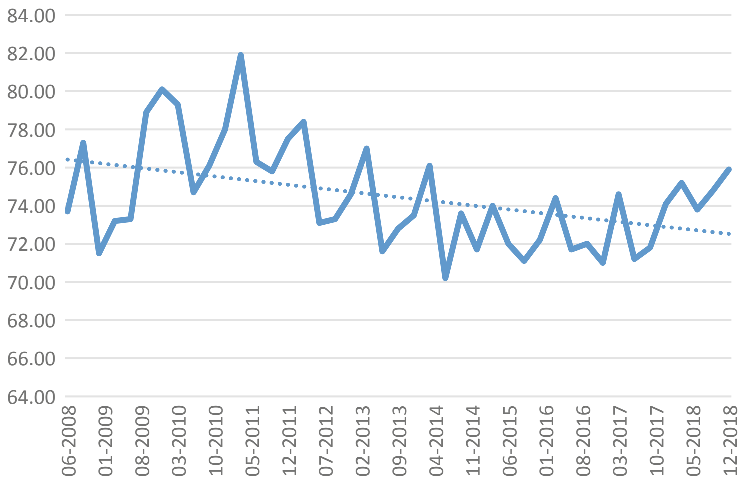


Fig. 2 Capacity utilisation. *Source* RBI OBICUS, CEIC India Database

creation. A slowdown sets in with the onset of negative demand shocks as was experienced during the financial crisis and capacity utilisation slows down. Accumulation of excess capacity serves to depress investment demand, thereby further intensifying the slowdown.

The literature in the Indian context indicates that an important factor affecting private corporate investments relates to the opportunity cost of undertaking expansion in productive capacity. The literature above indicates that both the lending rate as well as the rate of returns on financial assets affect the opportunity cost of investible

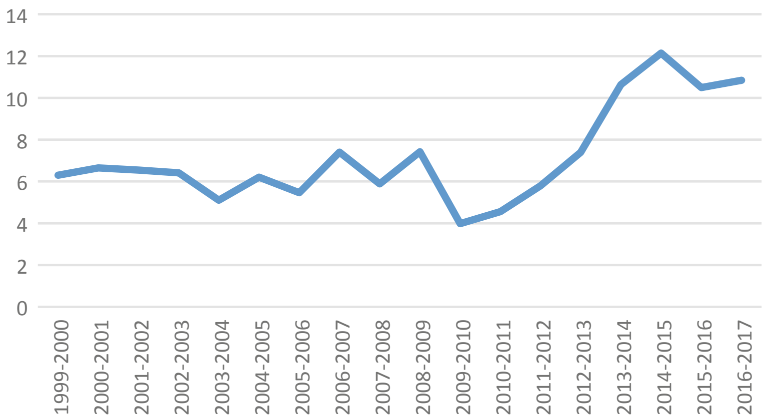


Fig. 3 User cost of capital. *Source* Authors calculation based on CEIC India and WDI Databases

funds. The user cost of capital broadly captures the interest lost from investing in capital goods rather than investing in a financial asset. There was a sharp rise (Fig. 3) in the user cost of capital from 2010 onwards, reflecting the hawkish monetary policy stance which lasted till early 2015 (Rakshit 2016).

Several firm-level characteristics also affect investment decisions. The literature above indicates that state of firms' balance sheets, especially with respect to their stock of debt and the cost of debt servicing, would have an important bearing on their decision to invest. High debt burdens impair the ability to undertake capacity expansion, whereas with relatively lower debt stocks, healthy growth in internal cash flows and retained earnings enable firms to create capacity, even when a borrowing constraint is an operative. In this context, indicators such as firms' interest coverage ratios (ICR), calculated as the ratio of EBIT (earnings before interest and tax) to interest payment can be useful.

Quarterly survey data from RBI on companies' ICR shows this has been on a downward trend, with its average value from 2009 to 2017 being distinctly lower than for the 2004–2007 period (Fig. 4a). Quarterly survey data also from RBI on profit growth, available from 2011, shows large year-to-year fluctuations (Fig. 4b). Together these trends of a growing burden of interest in the face of fluctuations in profit growth demonstrate the inherent instability associated with retained earnings as a source of financing fixed capital formation. Simultaneously if lenders also face balance sheet woes with rising NPAs and the credit constraint becomes binding, financing of fresh capital formation may indeed be quite problematic in such a situation.

Given that the decision to invest is based on a future income stream, the importance of firms' expectations can hardly be overemphasised. Uncertainty with regard to the policy framework, the state of overall macroeconomic stability and prospects for economic growth are important factors influencing such expectations. The overall macroeconomic developments and growth in aggregate demand also affect private investments by affecting business' expectations about the future. The CEIC Leading

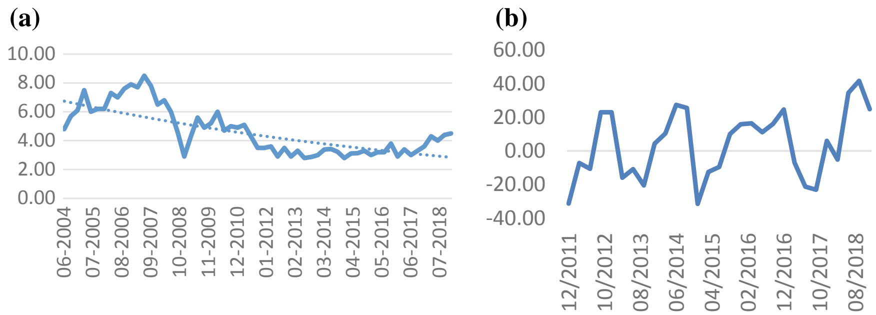


Fig. 4 **a** Interest coverage ratio. **b** Net profits: year-on-year growth. *Source* RBI Enterprises Survey, CEIC India Database

Indicator, available at monthly frequency since November 1995, is an overall index encompassing a broad range of macroeconomic and business-related variables that can be used as an index of expectations regarding future economic trends. The *standard deviation* in this indicator over a fiscal year is used as a measure of volatility that would affect expectation formation by private businesses (Fig. 5). During 2004 to 2007, the standard deviation in this index was relatively low and stable, before it increased sharply and peaked during the financial crisis. Even though the peak declined since then, fluctuations in the series indicate the presence of uncertainties about the future that may affect capital formation in the private sector.

The literature above also indicates that public investments can have a positive impact on private investments by boosting aggregate demand and capacity utilisation. It can ‘crowd in’ private investments by alleviating supply bottlenecks, especially in the context of developing countries. Trends in the growth rate of public investment (Fig. 6) reveal a couple of interesting points. High growth in private investments in the mid-2000s decade saw relatively high growth in public investments, and this was a period when user cost of capital was relatively moderate for the period under

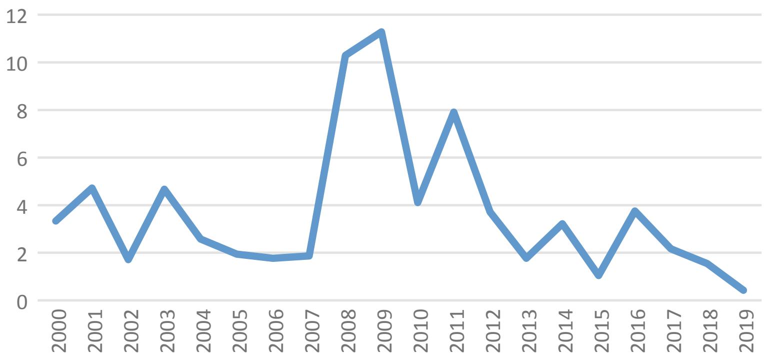


Fig. 5 Volatility in CEIC Leading Indicator. *Data* Authors calculations based on CEIC India Database

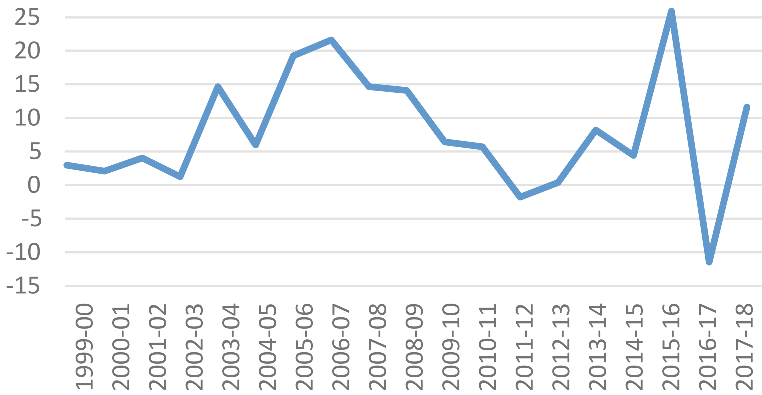


Fig. 6 Real public GFCF: year-on-year growth (2011–12 base). *Source* Authors calculations based on CEIC India Database

consideration. However, the fiscal push in the post-crisis period, which has seen lacklustre recovery in private business investments, has not taken public investment growth to its previous high; average year-on-year growth has been much lower and more unstable without any sustained increase. The spike in public investment growth in 2015–16 occurs at a time when user cost of capital is very high, indicating a possible reason behind its low impact on private investments.

Finally, on the demand side, the importance of external demand has been emphasised in the literature especially as an important factor behind India's stellar growth performance between 2003 and 2008. Export growth collapsed during 2009–10 at the peak of the financial crisis in the OECD, recovered briefly thereafter in 2010–11 and has been on a downward trend especially since 2011–12 with little sign of recovery, especially to the rates witnessed in the pre-crisis period (Fig. 7). The overlap in the period of the slowdown in exports and private investments indicates that external demand may have been an important factor in affecting capital formation by domestic firms. Not only does this impact on capital formation by exporter firms, but the weakening of the export multiplier can impact firms in general via negative demand-side effects.

In what follows, an attempt is made to identify whether opportunity cost or demand-side factors or expectations have played the most important role in affecting private corporate investments in the Indian context by estimating a simple regression model.

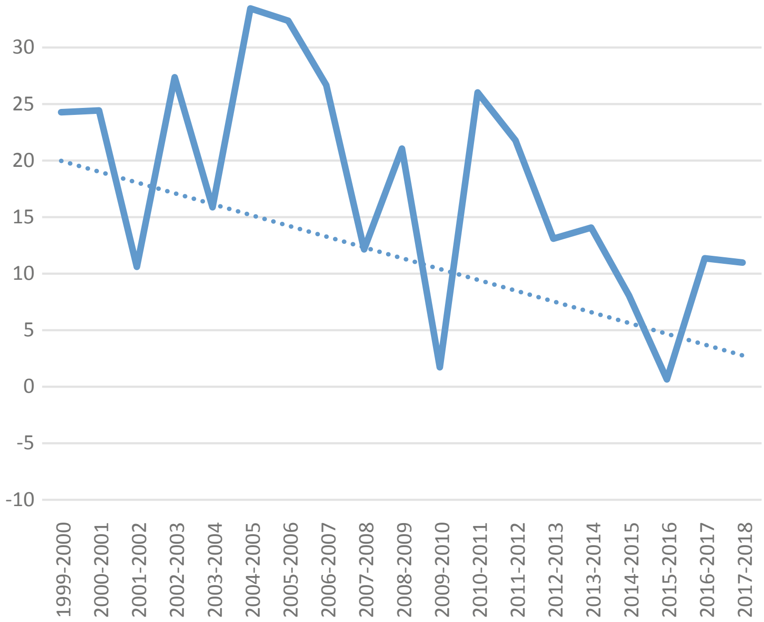


Fig. 7 Real export (Rupee values): year-on-year growth (2011–12 base). *Source* Authors calculations based on CEIC India Database

3.2 Regression Analysis

In what follows, we report the results from a simple OLS regression model with growth in private investments as the dependent variable and the main explanatory variables drawn on the basis of its determinants as discussed in the literature.

Given that expectations regarding the future appear as an important determinant of investment, both in the theoretical literature as well as in empirical studies on India, the model includes a measure of business expectations, on which data is available only from 1995 to 1996 onwards. As such the time period of the analysis is limited to 1995–96 to 2016–17. Since disaggregated quarterly data on private business investments is not available, the estimation relies on annual data.[12] However, this limits the number of observations and the scope for using advanced time series techniques.

Given the data-related constraints and the limited time period for the study, the simplest possible model is estimated to capture the most important determinants of private investments that emerge from theory. In essence GDP, business expectations and the real interest rate are the most basic determinants of investments (Carlin and Soskice 2006). While GDP captures the impact of demand-side factors, the real interest rate provides a measure of the opportunity cost of investments and the

[12]Quarterly data is available only on total gross fixed capital formation, which includes investments by the government, private businesses as well as households. This is used by Anand and Tulin (2014).

importance of 'future outlook' for current investment decisions is captured by the measure of business expectations.

As such in the OLS model, growth in real private corporate investments is the dependent variable, and real GDP growth and its lag, real interest rate and a measure of business expectations are the explanatory variables. The model is also estimated using the user cost of capital as an explanatory variable in place of the real interest rate and, as expected, results are similar in both these models with respect to the sign and statistical significance of the variable measuring the opportunity cost of investments.

In addition, to assess the importance of external demand and the role of public investments, two dummy variables are also used in the model. This is done especially as the literature indicates that public investments have a complementary role in affecting firms' investment decisions and also emphasises the importance of exports. The first dummy variable takes the value one for periods of relatively high export growth and the second, is one for periods of relatively high rates of growth in public investment and zero otherwise; for both series the cut-off used is median growth rate for the period.[13] Two points may be noted here. Firstly, export growth and growth in public investments are both subsumed within real GDP growth; as such inclusion of these as additional explanatory variables in the presence of GDP growth is not very meaningful. Secondly, dummy variables based on episodes of high growth in these two variables may be meaningfully used to test the hypothesis that on average, any given rate of growth of GDP would be associated with a higher rate of growth of private business investments, when export growth or growth in public investments is relatively high. A positive and statistically significant coefficient of the dummy variables would be indicative of such a conclusion.

As variables are measured in growth rates and in changes, rather than levels, the results provide an estimate of the short-run relation indicating the nature of adjustment in the dependent variable likely to be associated with changes in each of the explanatory variables, holding the others constant. Details regarding data sources and construction of variables are provided in the Appendix. The variables are tested for stationarity (results are reported in the Appendix), and OLS is applied on stationary variables, using Newey–West heteroscedasticity and autocorrelation (HAC) robust standard errors, that allow for efficient estimation under arbitrary forms of autocorrelation.

It may be noted that with investment growth as dependent variable and growth in real GDP as an explanatory variable, there is a potential problem that GDP growth may be an endogenous regressor. Therefore, before proceeding with OLS estimation, an instrument variables (IV) approach was adopted, and a two-stage least squares model was also estimated using HAC robust standard errors. Growth in real government consumption and the lagged dependent variable were used as instruments for growth in real GDP following the literature, wherein similar variables have been previously used (Athukorala and Sen 2003). However, the result of the Durbin–Wu–Hausman test showed, the null hypothesis that real GDP growth

[13] Similar results are obtained by using the mean value as the cut-off for both the series.

is an exogenous variable could not be rejected at standard levels of significance.[14] Since the hypotheses test showed no evidence of endogeneity of the regressor, it is valid to proceed with OLS estimation, which is more efficient compared to the IV estimator and based on which causal inferences could be drawn in the absence of endogeneity.

Change in real exchange rates, change in the user cost of capital and volatility in expectations have the expected negative sign and are statistically significant. The coefficient of GDP growth is positive and significant, while that of its first lag is positive but statistically insignificant. This is fairly common in models which include variables in current and lag forms. The presence of multicollinearity makes it difficult to estimate the impact multipliers efficiently. A *t*-test is carried out to test for the statistical significance of the sum of the coefficients of GDP growth and its lag, showed it to be significant at slightly higher than the 10% level of significance.[15] The coefficients of both the dummy variables are positive and statistically significant, indicating thereby private investment growth would be higher if any given rate of growth of GDP (holding real interest rates and volatility in expectations the same) is associated with significantly higher rates of growth of exports or public investments. From the magnitude of coefficients, growth in external demand appears relatively more important compared to growth in public investments in terms of its impact on private investments. It may be mentioned here that inclusion of a dummy to control for years when growth in government consumption was relatively high (higher than the median for the period) leads to a different result—the coefficient is statistically insignificant and has a negative sign.[16] The results, therefore, serve to underscore the importance of government investment, rather than consumption for private investments in the Indian context.

Since the variables are scaled differently, standardised coefficients are used to assess which one has a relatively larger impact on the dependent variable. Results (Table 1) show that volatility in business expectations has by far the largest impact on investment growth, followed by growth in GDP and real interest rates and the user cost of capital. The importance of real interest rates and expectations is also found by Anand and Tulin (2014), who use quarterly data for their study, but do not control for

[14]Details of the test for endogeneity following 2SLS are as follows:

Value of the HAC score chi2(1) test statistic = 0.85918 (*p*-value = 0.3540).

Optimal GMM method was also used, and details of the test for endogeneity following optimal GMM are as follows: Value of the GMM C statistic chi2(1) = 0.00615 (*p*-value = 0.9375).

The high *p*-values (based on HAC robust standard errors) indicated the null hypothesis that the regressor is exogenous could not be rejected.

[15]The null hypothesis is that sum of the coefficients of GDP growth and its lag is zero; and the value of the test statistic, based on t-distribution (using Newey–West HAC standard errors): (a) in Model 1 was 2.57, statistically significant at 11% level; (b) in Model 2 was 2.71, statistically significant at 10.3% level. This provides limited support for the hypothesis that a permanent increase in GDP growth rate by 1 percentage point, *ceteris paribus*, is associated with more than 2.5 percentage point increase in private investment growth after one year.

[16]The estimated coefficient of growth in government consumption is −5.49 with *p*-value of 0.184 when it is included in Model 1.

Table 1 Results of OLS regression

Dependent variable: Pvt_Inv_Gr	Estimated coefficients			
	Model 1	Standardised coefficients	Model 2	Standardised coefficients
R_GDP_Gr (t)	2.30* (0.055)	0.255	2.45** (0.032)	0.271
R_GDP_Gr ($t-1$)	0.271 (0.803)	0.029	0.266 (0.820)	0.029
LI_SD (t)	−3.19*** (0.000)	−0.394	−3.11*** (0.000)	−0.385
R_Int_Ch	−2.11* (0.092)	−0.230		
UC_Ch			−2.43* (0.099)	−0.198
X_Gr_Dum	17.90* (0.066)		18.24* (0.066)	
PubInv_Gr _Dum	9.20* (0.058)		8.15* (0.089)	
Constant	−5.608 (0.644)			
Number of Observations	21		21	
Adjusted R^2	0.56		0.54	

Note Figures in parentheses are p-values based on heteroscedasticity and autocorrelation (HAC) consistent Newey–West standard errors; for each variable, the standardised coefficient indicates the change in by how many standard deviation units the dependent variable would change when the explanatory variable changes by one standard deviation. ***, ** and * denote statistical significance at 1percent, 5 percent and 10 percent levels respectively

Abbreviations

Pvt_Inv_Gr: Year-on-year growth in real private corporate GFCF
R_GDP_Gr: Year-on-year growth in real GDP
LI_SD: Standard deviation of the CEIC Leading Indicator index
R_Int_Ch: Change (first difference) in real interest rate
X_Gr_Dum: Dummy variable which is 1 for years of high growth in exports, 0 otherwise
PbInv_Gr_Dum: Dummy which is 1 for years of high growth in public investments, 0 otherwise
UC_Ch: Change (first difference) in user cost of capital

domestic demand-related factors in their model. These results additionally indicate the importance of demand-side variables in the Indian context.

In the context of the current slowdown, these results primarily emphasise the potential role of uncertainty with respect to the future outlook that would have a negative impact via business expectations. The role of demand-side variables, especially related to the slowdown in external demand and public investments, is also highlighted along with that of high real interest rates and user cost of capital.

This is a limited empirical exercise especially owing to data-related constraints. However, the signs and statistical significance of the variables are as expected and robust. Future work aims to extend the simple analysis carried out in this paper, especially by creating a dataset spanning a longer time period.

4 Conclusion

In the context of the current slowdown in private business investments, existing studies have highlighted the role of a number of macroeconomic policy-related factors as well as the operation of certain firm-level constraints, especially related to the financing of investments. The expository analysis in this paper finds evidence corroborating these factors. The tight money policy in the latter half of the 2000s decade possibly contributed to raising the opportunity cost of investments, while the focus on fiscal consolidation is likely to have played a role in dampening domestic demand, in an environment of subdued external demand. Uncertainty in the policy environment and the role of institutional factors have also been highlighted; these have been manifested in episodes of 'policy paralysis', cases of corruption and time-consuming litigation, all of which are especially likely to affect business confidence and have repercussions on firms' decision to expand capacity. On the financing side, an increasing amount of corporate debt as well as rising burden of NPAs of the banking sector is also likely to have adversely affected business investments; however, only detailed firm-level analyses would provide deeper insights into this aspect, as compared to studies based on aggregate data.

The simple estimation exercise, carried out in this paper based on aggregate data, highlights the role of demand-side factors, especially external demand, real interest rates and business expectations in significantly affecting private investments. The findings also emphasise the importance of public investments, which are likely to have a positive impact on private business investments via demand-side factors in the short run and by alleviating supply-side bottlenecks in the medium to longer term. Specifically, in the context of the current slowdown, the results indicate that strong export growth will play a crucial role in reviving private investment growth. However, a pickup in export growth may well take time, especially with imminent signs of global trade disruptions centring around complex trade negotiations between the USA and its developing trade partners as well as major events like Brexit. In the meantime, the policy focus on enhancing the rate of public investments is likely to be rewarding in terms of ushering a revival in private investment growth in India.

Appendix

The data on the real interest rate is accessed from the World Development Indicators (WDI), 2018. For the remaining variables, i.e. Real GFCF in the private corporate sector, Real GFCF in the public sector, Real GDP, Real exports (in Rupees), NAS data is accessed through the CEIC India Database.

For real GDP at 2011–2012 prices, we spliced the two GDP series at different base years, i.e. 2004–2005 and 2011–2012, to get one long GDP series at 2011–2012 prices. There are methodological differences between the 2011–12 and 2004–05 aggregate macroeconomic series, and therefore, to check the consistency of the data, the growth rates in the series generated through simple splicing were compared to the methodology used in Ahmad et al. (2018) and the results were very similar.[17] Although this method is imperfect, there are no other options for generating a long enough time series for carrying out an estimation exercise. Similarly, the data series on real public gross fixed capital formation, available in 2011–12, 2004–05, 1999–2000 and 1993–94 base years and on real exports (in rupees) available in 2004–05 and 2011–12 base years, were spliced to generate a single series at 2011–12 prices. These series were used to calculate year-on-year growth in each of the variables for the period 1989–2017. However, estimation could be carried out only for the sub-period 1995–96 to 2016–17 owing to limitation related to the measure of business expectations.

The measure of business expectation was calculated based on monthly data on the CEIC Leading Indicator Index available from November 1995 onwards. Standard deviation in this index, calculated for each fiscal year (from April, current year to March, next year), is taken as a measure of uncertainty in the business and overall macroeconomic environment. A high value of this variable indicates greater uncertainty with potentially negative impact on business expectations.

The export growth dummy takes the value one for the years when export growth was higher than 11% (its median value during this period). The public investment growth dummy takes the value one for the years when growth in public investment was higher than 4.5% (its median value during this period).

The user cost of capital (UC) is calculated as UC = external nominal rate of return + nominal depreciation (i.e. real rate of return * investment price $t - 1$ + depreciation rate * investment price t).[18] Nominal depreciation was estimated by calculating capital consumption allowance as a percentage of capital stock and the implicit deflator for GFCF was used to denote investment price—all NAS data accessed from CEIC database.

The table below shows all the growth series are stationary at standard levels of significance, except for real interest rates and user cost, which were therefore first-differenced to render them stationary.

[17] Ahmad et al. (2018) splice the 2004–05 and 2011–12 series backward using the following back-casting method that maintains a growth rate in the new series that is consistent with the old series so that the resulting spliced series is essentially a level shift to the old series with equivalent growth rates:

For the variable X_t that needs to be spliced, the new series is denoted as X^* and X is the variable value in the old series. With data in the new series starting from period t, the value of X^*_{t-1} is calculated as $X^*_{t-1} = (X_{t-1}/X_t)\mathrm{X}^*_{t-1}$.

[18] This is in line with the methodology adopted in RBI, KLEMS India Project (https://rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=894).

Variable	ADF test of H_0: $I(1)$ versus H_1: $I(0)$	Whether $I(0)$ or $I(1)$
Pvt_Inv_Gr	−3.595*** (0.0059)	$I(0)$
R_GDP_Gr	−4.100*** (0.0010)	$I(0)$
LI_SD	−3.500*** (0.0080)	$I(0)$
R_Int	−2.319 (0.1658)	$I(1)$
R_Int_Ch	−7.370*** (0.0000)	$I(0)$
UC	−1.126 (0.7045)	$I(1)$
UC_Ch	−3.235** (0.0181)	$I(0)$

Note Statistical significance at 1% level is denoted by *** and at 5% level by **; all the variables are stationary with and without trend (results reported above)

References

Ahmad, J., Blum, F., Gupta, P., & Jain, D. (2018, July 10–11). *India's Growth Story*. NCAER India Policy Forum. Mimeo.

Anand, R., & Tulin, V. (2014). *Disentangling India's investment slowdown*. IMF Working Paper: Asia and Pacific Department. www.imf.org.

Athey, M., & Reeser, W. (2000). Asymmetric information, industrial policy and corporate investment in India. *Oxford Bulletin of Economics and Statistics, 62,* 267–292.

Athukorala, P., & Sen, K. (2003). Liberalisation and business investment in India. In R. Jha (Ed.), *Indian economic reform*. New York: Palgrave Macmillan.

Carlin, W., & Soskice, D. (2006). *Macroeconomics: Imperfections, institutions and policies*. New Delhi: Oxford University Press.

Dastidar, A. G. (2015). India's experience with export-led growth. In J. Ghosh (Ed.), *India and the international economy, ICSSR research survey and explorations in economics* (Vol. 2). New Delhi: Oxford University Press.

Goldar, B. N. (1997). Econometrics of Indian industry. In K. L. Krishna (Ed.), *Econometric applications in India*. New Delhi: Oxford University Press.

Krishnamurty, K., & Sastry, D. U. (1975). *Investment and financing in the corporate sector in India*. New Delhi: Tata McGraw-Hill.

Nagaraj, R. (2013). India's dream run, 2003–08: Understanding the boom and its aftermath. *Economic and Political Weekly, 48,* 39–51.

Patnaik, I. (2019, March 23). Pick-up in quarterly sales reflects a definite upswing. *Hindustan Times*.

Rajakumar, D. (2005, September 17–23). Corporate financing and investment behaviour in India. *Economic and Political Weekly*, 4159–4165.

Rakshit, M. (2016, April). India's post-crisis macroeconomic slowdown: Some perspectives. *Money & Finance, ICRA Bulletin*, 49–84.

Sarma, J. V. M. (1988, October 22). Investment allowance and private corporate investment in India. *Economic and Political Weekly*, 2235–2238.

Sen, K., Kar, S., & Sahu, J. (2014). *The political economy of economic growth in India 1993–2013*. ESID Working Paper No. 44. Manchester, UK: University of Manchester. www.effective-states.org.

Sen, S., & Dasgupta, Z. (2018). Financialisation and corporate investments: The Indian case. *Review of Keynesian Economics, 6*(1), 96–113.

Sikdar, S. (2006). *Principles of macroeconomics*. New Delhi: Oxford University Press.

On the Lead–Lag Relationship Between Market Capitalization Ratio and Per Capita Growth

Agnirup Sarkar and Abhirup Sarkar

Abstract The paper looks at the short-to-medium term lead–lag relationship between stock market performance, as represented by market capitalization to GDP ratio and the performance of the real economy, as represented by per capita GDP growth. The lead–lag relationship is first established theoretically in an asset pricing framework with production and accumulation and then empirically by separately looking at lagged correlations, Granger causality and variance decomposition using data from 35 countries over the period 1988–2012.

Keywords Growth · Market capitalization

1 Introduction

The purpose of the present paper is to explore and establish a lead–lag relationship between stock market performance and the performance of the real economy. There is a substantial literature concerned with possible causal relationships between different measures of stock market performance and real GDP or its growth both in the long run and in the short run. The present paper differs from this literature in that it does not attempt to look at any causal relationships. In contrast, it considers a scenario where an outside shock affects both variables in such a way that a change in one of the variables in the current period is followed by a change in the other in subsequent period or periods. Representing stock market performance by the ratio of market capitalization to GDP and real performance of the economy by the growth of per capita income,

The authors thank an anonymous referee for helpful comments on an earlier draft.

A. Sarkar
Indian Institute of Technology Guwahati, Guwahati, India
e-mail: agnirupsarkar@iitg.ac.in

A. Sarkar (✉)
Indian Statistical Institute Kolkata, Kolkata, India
e-mail: abhirupsarkar.sarkar@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_4

the paper demonstrates both theoretically and empirically that a change in the market capitalization ratio is followed by a change in the growth rate of per capita income.

There is old literature, both theoretical and empirical, which emphasizes financial intermediation in general, and stock market developments in particular, as major factors behind *long-run* growth. The theoretical literature goes as far back as Schumpeter (1934) and Hicks (1969) and culminates in more recent works of Jacklin (1987), Gorton and Pennacchi (1990), Greenwood and Jovanovic (1990), Bencivenga and Smith (1991), Levine (1991), Japelli and Pagano (1994) and Bencivenga et al. (1995) among others. In this literature, there are primarily two ways in which stock markets help growth and development. First, it acts as an efficient bridge between savers and investors, thereby channelizing funds for investment, capital accumulation and growth. Second, a vibrant secondary market in stocks makes stocks more liquid and therefore increases the incentive of agents to invest in the primary market.

There is a second strand of literature which links stock markets with growth or levels of per capita income through reverse causation. This second strand of literature, the origin of which can be traced back to the efficient market hypothesis of Fama (1965) and was first fully developed by Lucas (1978), visualizes per capita income determining stock prices. The pioneering work of Lucas (1978) led to subsequent theoretical work by Abel (1988), Cochrane (1991) and others. In these theoretical models, stocks are taken to be a major instrument of savings. If there is a rise in income in the current period, consumers, eager to smooth out their consumption over time, would tend to save more. This is likely to increase the demand for stocks and hence their prices. Consequently, following a rise in per capita income, stock prices and the market capitalization ratio, defined as the ratio of the value of total stocks in the market to GDP, are likely to go up. There is, however, an opposite effect. If income effects are strong, a rise in current per capita income might overwhelmingly increase current consumption and not so much savings. Consequently, the rise in stock prices and hence market capitalization will be small and the ratio of market capitalization to per capita income will fall. Thus, in the second stream of literature, per capita income determines the market capitalization ratio, that is, the direction of causality flows from per capita income to market capitalization. But it is not clear whether one can get an unambiguous signal about the current state of the economy from the stock market.

There is a third stream of literature. Brock (1982) developed an intertemporal general equilibrium theory of capital asset pricing which was an extension of the earlier stochastic growth model of Brock and Mirman (1972). In this model, asset prices and outputs are both subject to stochastic shocks and are jointly determined from the model. The present paper is closest to this approach. We extend the Lucas (1978) framework in the spirit of Brock (1982) by introducing production and accumulation. Savings are partly used for accumulation of physical capital and partly to buy stocks. A stochastic shock to production affects both output and the price of stocks. Consequently, both the market capitalization ratio and per capita growth are affected.

The theoretical literature is accompanied by vast empirical literature. This literature can be divided into two parts. One part deals with the question as to how the financial fabric in general and financial intermediation in particular affect long-run growth. The other part addresses the question of causality between financial intermediation and growth and the direction in which it moves. Pioneered by Goldsmith (1969), the empirical literature includes the works of King and Levine (1993a, b, c), Levine and Zervos (1996), Rajan and Zingales (1998), Beck et al. (2000) and Fase and Abma (2003) among others.

In this paper, we establish a lead–lag relationship between market capitalization ratio and per capita growth with the former taking the lead and the later following. Our theoretical model implies that the reverse, that is per capita growth taking the lead and market capitalization ratio following, is not true. Next, this lead–lag relationship is established empirically. Our empirical exercises also establish that the reverse lead–lag relationship is not valid. A substantial literature has emerged around the question of whether and to what extent movements in financial indicators predict future movements in real economic activity. Earlier works include Mills (1988), Estrella and Hardouvelis (1991), Choi et al. (1999), Aylward and Glen (2000) and others. The question assumed renewed interest in the aftermath of the subprime crisis leading to a more recent literature including Reinhart and Rogoff (2009), Apergis et al. (2015), Levanon et al. (2015) and Chen et al. (2018). The present paper attempts to contribute to this literature.

There are two distinct features of the present model which differentiates it from the above-mentioned literature. First, while the above-mentioned literature is basically empirical, the present paper first presents a theoretical framework which predicts that an increase in stock prices relative to GDP leads to real per capita growth in future. Then, as a next step, this theoretical prediction is tested empirically. Thus, apart from empirically establishing the lead–lag relationship, the present paper attempts to put forward a mechanism through which the lead–lag relationship works. Second, it specifically shows that the reverse mechanism, that is, growth preceding the stock market boom is not valid. This strengthens the understanding of the relationship between financial market indicators and real economic activity.

In what follows, in Sect. 2, we show that consequent to a shock, a change in the market capitalization ratio in the current period is followed by a change in subsequent per capita growth. In Sect. 3, this theoretical result is vindicated by three alternative empirical exercises. We separately look at lagged correlations, Granger causality and variance decomposition to establish our result. Section 4 concludes the paper.

2 Asset Pricing with Production and Investment

We consider an economy consisting of infinitely many identical competitive firms and infinitely many identical households. The representative firm is indexed by $i \in [0, 1]$, and the representative household is indexed by $j \in [0, 1]$. Thus, the total size of households and the total size of firms are normalized to unity. Each firm produces a homogeneous output using capital and a linear production function. A firm invests a part of its output, thereby augmenting next period's capital stock and distributes the remaining as dividends to the households who are the owners of the firm. For the representative household, dividend income is the only source of income, a part of which goes into consumption and the rest into buying new stocks.

2.1 Market Capitalization Ratio and Growth

The representative household maximizes its expected utility over an infinite horizon. Time is discrete. The household's utility $u\,(c_t)$ is a function of its consumption c_t alone. Taking β as the household's discount factor, the household's objective function can be formally written as

$$\text{Max: } E_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{1}$$

$$\text{s.t.: } \int_0^1 p_t^z(i)\left[z_{t+1}(i) - z_t(i)\right] di + c_t = \int_0^1 d_t(i) z_t(i) di \tag{2}$$

Equation (2) represents the household's resource constraint, where $p_t^z(i)$ and $d_t(i)$ are the price and dividend of the stock of the ith firm at period t and $z_t(i)$ is the quantity of the stock of the ith firm held by the representative household at period t. Consequently, the right-hand side of Eq. (2) represents the total (dividend) income of the representative household. On the left-hand side of Eq. (2) the first term represents expenditure of the household at period t to acquire additional assets. This, added with the consumption c_t, exhausts the household's total income.

The representative household maximizes (1) subject to (2) by choosing c_t and $z_{t+1}(i)$. The first-order condition with respect to these choice variables establish the following Euler equation:

$$u'(c_t) p_t^z(i) = \beta E_t u'(c_{t+1}) \left(d_{t+1}(i) + p_{t+1}^z(i)\right) \tag{3}$$

Since firms are identical, dividends and share prices are the same across all firms, which means that $d_t(i) = d_t$ and $p_t^z(i) = p_t^z$ for all i. Using this, (3) can be written as

$$u'(c_t) p_t^z = \beta E_t u'(c_{t+1}) \left(d_{t+1} + p_{t+1}^z\right) \tag{4}$$

Without any loss of generality, we assume that the total number of shares of a firm is a unity and this remains unchanged over time. Since the measure of firms is unity, the total number of stocks in the economy is also unity. This is distributed equally among households. Since the measure of households is also unity, each household has one unit of stock. This makes the right-hand side of (2) equal to d_t. Again, since the total number of firms remain unchanged over time, the first term on the left-hand side of Eq. (2) becomes zero which makes the left-hand side of Eq. (2) equal to c_t.

Therefore, Eq. (2) reduces to

$$c_t = d_t \tag{5}$$

Next, we assume that the utility function is logarithmic, i.e. $u(c_t) = \ln c_t$. The stock Euler equation in (3), then becomes

$$\frac{p_t^z}{c_t} = \beta E_t \left(\frac{c_{t+1} + p_{t+1}^z}{c_{t+1}} \right) \tag{6}$$

Solving recursively, the above equation becomes

$$\frac{p_t^z}{c_t} = \frac{\beta}{1-\beta} + \beta E_t \left\{ \lim_{n\to\infty} \beta^{n-1} \left(\frac{p_{t+n}^z}{c_{t+n}} \right) \right\} \tag{7}$$

Assuming that the term $\frac{p_{t+n}^z}{c_{t+n}}$ is bounded above for all n, the limit term in Eq. (7) goes to zero. This means that the equilibrium asset price becomes

$$p_t^z = \frac{\beta}{1-\beta} c_t \tag{8}$$

This price is determined in such a way that in equilibrium, each period, the representative household would not want either to increase or to decrease his holding of assets. This is guaranteed by the logarithmic utility function where income and substitution effects, which are of opposite signs, are of the same magnitude and offset each other.

In an economy described in Lucas (1978), dividend (described as fruit falling from trees in the Lucas asset pricing framework) arrives without any deliberate effort on the part of the consumers and is referred to as an endowment economy or exchange economy. There is no provision of storage and production in this kind of an economy. In the present theoretical framework, on the other hand, we allow for investment in physical capital and output production in each period by identical firms owned by the representative household.

The representative firm manufactures its product (y_t) using capital (k_t) as its only source of input, with the help of a linear production technology given by

$$y_t = \epsilon_t k_t \tag{9}$$

where ϵ_t denotes the total factor productivity (TFP) shock which influences output production in time period t.[1] We assume that ϵ_t is an iid shock. In time period t, the firm invests a part of its produce and distributes the rest as dividend to the household. The firm invests an amount i_t which gives rise to new accumulated capital for period $t+1$, given by k_{t+1}. The investment process for the firm is represented

[1]Implicitly, the production function is $y_t = \epsilon_t k_t l_t$, where $l_t = 1 \forall t$, that is, there is no labour-leisure choice and labour is supplied by the households inelastically at the unit level in each period.

by the following equation.

$$k_{t+1} = (1-\delta)k_t + i_t \tag{10}$$

where $(1-\delta)k_t$ stands for undepreciated capital stock at time period t. δ represents the rate of capital depreciation. The firm maximizes discounted stream of future dividends, where dividend at time period t is given by

$$d_t = y_t - i_t \tag{11}$$

Thus, the time t objective function of the firm is given by

$$\text{Max: } E_t \sum_{s=0}^{\infty} m_{t,t+s} d_{t+s} = E_t \sum_{s=0}^{\infty} m_{t,t+s} \left[\epsilon_{t+s} k_{t+s} - \{k_{t+s+1} - (1-\delta)k_{t+s}\}\right] \tag{12}$$

with k_{t+1}as the firm's choice variable. $m_{t,t+s}$ denotes the representative household's stochastic discount factor

$$m_{t,t+s} = \frac{\beta u'(c_{t+s})}{u'(c_t)} \tag{13}$$

As the firm maximizes its dividends on behalf of the household, it uses the latter's marginal rate of substitution or stochastic discount factor $m_{t,t+s}$ in its dividend maximization problem.

Since in equilibrium the household does not increase or decrease its holding of assets, the equilibrium resource constraint can be written as

$$c_t = \epsilon_t k_t - \left[k_{t+1} - (1-\delta)k_t\right] \tag{14}$$

The left-hand side of Eq. (14) represents the representative household's dividend income which is entirely consumed in equilibrium. Taking the first-order condition of the firm's maximization problem w.r.t. k_{t+1} and combining it with the equilibrium resource constraint, we can derive equilibrium consumption (c_t) and capital accumulation (k_{t+1}) expressions as

$$c_t = (1-\beta)\,(\epsilon_t + 1 - \delta)\,k_t \tag{15}$$

and

$$k_{t+1} = \beta\,(\epsilon_t + 1 - \delta)\,k_t \tag{16}$$

Utilizing the equilibrium consumption and capital accumulation expressions from the above equations in (15) and (16) along with the equilibrium asset price in (8), next we derive equilibrium expressions for market capitalization as a ratio of output (mk_t) and output growth (yg_t).

For period t, we define market capitalization ratio as $\frac{p_t^z}{y_t}$ and growth as $\frac{y_t}{y_{t-1}}$. From the model, the solutions to these variables are obtained as

$$mk_t = \beta \left[1 + \frac{(1-\delta)}{\epsilon_t} \right] \tag{17}$$

and

$$yg_t = \beta \epsilon_t \left[1 + \frac{(1-\delta)}{\epsilon_{t-1}} \right] \tag{18}$$

From Eqs. (17) and (18), we get

$$yg_{t+1} = \epsilon_{t+1} mk_t \tag{19}$$

Equation (19) gives a lead–lag relationship between market capitalization and growth. More specifically, market capitalization ratio in period t has a positive impact on growth in period $t+1$, that is, an increase in the former leads to an increase in the latter after a one-period lag. This is precisely the relationship we have been looking for. The intuition behind the lead–lag relationship is straightforward. An increase in ϵ_t from its steady-state value increases both y_t and k_{t+1} [see Eqs. (9) and (16)]. The increase in k_{t+1} in turn increases y_{t+1}. However, the proportionate increase in y_t is greater than the proportionate increase in y_{t+1}.[2] Hence, there is a fall in the rate of growth in period $t+1$. On the other hand, from (17), it directly follows that a positive productivity shock reduces the market capitalization ratio.[3] We wish to point out that the lead–lag relationship of Eq. (19) does not imply any causation. It is induced by a shock which is exogenous to the system. We summarize our findings in the following proposition:

Proposition *In an asset pricing model with production, capital accumulation and growth and an exogenous iid productivity shock, there is a lead–lag relationship between market capitalization ratio and growth in the sense that a rise in the market capitalization ratio in period t leads to a rise in the rate of growth in period* $t+1$.

The proposition must be qualified by an important caveat. In our model, we have assumed that the utility function is logarithmic. It is well known that in a logarithmic utility function, income effects and intertemporal substitution effects exactly cancel out. In a Lucas (1978) asset pricing model, where there is no production or capital accumulation, a logarithmic utility function yields a market capitalization ratio which is constant in equilibrium. In particular, it is independent of any outside shock.[4] This is because log utility implies a proportionate increase in savings and consumption

---

[2]This may be verified by differentiating Eqs. (9) and (16) proportionately, noting that the only perturbation to the system comes through a change in ϵ_t which induces changes in y_t and k_{t+1}.

[3]A positive productivity shock increases the stock price but also increases output in a greater proportion. Hence, the fall in the market capitalization ratio.

[4]This result is reported in Ljungqvist and Sargent (2004).

following an increase in income. But in the Lucas (1978) model, goods cannot be stored. Hence, consumption rises in the same proportion as income and with log utility so does savings. Since the only vehicle of savings is stocks and the number of stocks is fixed at unity a rise in savings is translated into a proportionate rise in the price of stocks. All this, taken together, implies that in the Lucas (1978) model with a logarithmic utility function, the market capitalization ratio is independent of income.

This independence breaks down once accumulation is allowed. With capital accumulation, a part of the increased savings, following a shock, is invested in physical capital. Hence, the demand for stocks increases by less and so does the price of stocks. As a result, the market capitalization ratio falls as shown in Eq. (17). The positive lead–lag relationship follows from this. But this unambiguous relationship is dependent on logarithmic utility. We have shown elsewhere[5] that if more general utility functions are allowed, the relationship between market capitalization ratio and income or its growth becomes ambiguous. Hence, in the empirical exercise that follows, we shall try to see if the market capitalization ratio influences growth in subsequent periods. Simple correlations indicate that this influence is positive.

3 Empirical Analysis

3.1 Motivation

The existing empirical literature does not throw much light on the lead–lag nature of the relationship between the stock market and growth. In the real world, a macroeconomic shock can influence both the stock market and growth, but the change in behaviour of any one of the variables due to the shock can take some time to get translated to the other. Since no clear conclusion can be drawn from the existing literature about the direction of causality between stock market capitalization and growth, in order to figure out whether the present stock market activities depend upon past values of growth or whether the past values of stock market capitalization determine current growth, we empirically investigate this first by calculating simple lead–lag correlations between market capitalization ratio in year t and growth in year $t + 1$. To check the reverse possibility of lead–lag, we also calculate the correlations between growth in year t and market capitalization ratio in year $t + 1$. Next, we run a Granger causality test, using time series data on market capitalization—GDP ratio and growth of per capita GDP from different countries and country groups. This is done despite the fact that no clear causality might actually exist between the stock market activities on the one hand and economic growth on the other and both can be influenced by independent exogenous factors. The sole purpose of the Granger causality exercise is to develop an understanding of nature of lead–lag relationship

[5] See Sarkar and Sarkar (2019).

between market capitalization ratio and growth, if there exists any such relationship. Finally, within a VAR structure, we perform a variance decomposition analysis of stock market capitalization and growth in order to compare the percentage of short-run fluctuations in future market capitalization driven by GDP growth with the percentage fluctuations of GDP growth driven by market capitalization. In other words, a variance decomposition helps understand whether market capitalization is the chief determinant of growth during later time periods or whether the opposite is true.

3.2 Data

We have already mentioned that as an indicator of the stock market activities, we use market capitalization as a proportion of output. To understand the relationship between market capitalization as a proportion of output and growth of output over time, we look at annual data on these two variables for 35 countries and four country groups for the time period covering 1988–2012. Since we deal with data on market capitalization, it is understandable that the countries in question are mostly high income and middle income. The data used for our analysis are secondary.[6] The time period and the choice of countries have been dictated by the availability of data.

In order to find a measure of output growth, we have first collected data on GDP per capita, which is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2005 US dollars. From this series, we have constructed a gross growth of per capita GDP series for each country for the given time span of 25 years.

As a measure of stock market performance, we have used data on market capitalization as a percentage of GDP for each country for 25 years, where market capitalization (also known as market value) is the share price times the number of shares outstanding. Listed domestic companies are the domestically incorporated companies listed on the country's stock exchanges at the end of the year. Listed companies do not include investment companies, mutual funds or other collective investment vehicles.

In the tables summarizing the empirical results, we use **MK** to denote market capitalization as a percentage of GDP and **PCG** to denote growth of per capita GDP over time. To ensure linearity, natural logs of both series are taken while performing empirical operations.

[6]Data Source: World Development Indicators (www.data.worldbank.org).

3.3 Cross-Country Evidence of Lead–Lag Relationship Between Market Capitalization Ratio and Growth

3.3.1 Lead–Lag Correlations

To set our initial ideas, we calculate two sets of correlations. First, we calculate the correlation, for each of the 35 countries, between market capitalization ratio in year t and growth in year $t + 1$. We have data for 1988–2012, that is, for 25 years. So we consider the market capitalization ratio for the period 1988–2011 and growth for 1989–2012. This gives us 24 data points for each country. Next, we calculate the reverse correlations, that is, between growth in year t and market capitalization ratio for year $t + 1$.

Correlation coefficients between MK_t and PCG_{t+1} show the extent to which current market capitalization ratio influences future growth. If this correlation is positive, one can argue that there is a lead–lag relationship between market capitalization ratio and growth with the former taking the lead and the latter following with a lag. These correlation coefficients are presented in Table 1. Out of the 35 countries reported in the table, 22 countries exhibit lead–lag correlations between MK_t and PCG_{t+1} significant at 5% level. Another three countries exhibit significant correlations at 10%. It is to be noted that all correlations are positive. This means that an increase in the market capitalization ratio in the current year *positively* influences growth in the next year.

We also calculate the reverse correlations, that is, correlations between PCG_t and MK_{t+1}. It turns out that for *all* the 35 countries this correlation is insignificant. Therefore, we can conclude that growth in the current year has no significant influence on next year's market capitalization ratio. Per capita growth cannot lead to a lagged boom in the stock market. Since correlations between PCG_t and MK_{t+1} are insignificant for all countries, we do not report these numbers.

To summarize, our calculations indicate that more than 60% of the 35 countries exhibit a lead–lag relationship between market capitalization ratio and growth, but not between growth and market capitalization.

3.3.2 Granger Causality

We perform a Granger causality exercise for market capitalization ratio and growth for all the 35 countries and four country groups. For each of the different countries and country groups, data for each of the two variables are available for 25 years. Denoting market capitalization ratio by MK_t and per capita growth by PCG_t and after ensuring that the two series are stationary, I conduct a joint Granger causality exercise. To test the null hypothesis that MK_t does not Granger cause PCG_t, consider the following autoregression

Table 1 Market capitalization growth lead–lag correlation

Countries	Lead–lag correlation
Australia	0.12
Austria	0.02
Bangladesh	0.42**
Belgium	0.18
Brazil	0.73**
Canada	0.65**
China	0.3*
Columbia	0.41**
Denmark	0.11
Egypt	0.78**
Finland	0.41**
France	0.25
Germany	0.43**
Greece	0.55**
India	0.64**
Italy	0.68**
Malaysia	0.44**
Netherlands	0.21
New Zealand	0.32*
Nigeria	0.41*
Norway	0.18
Pakistan	0.39**
Philippines	0.48**
Portugal	0.11
Russia	0.9**
Singapore	0.1
South Africa	0.52**
Spain	0.37**
Sri Lanka	0.56**
Sweden	0.63**
Switzerland	0.42**
Thailand	0.24
Trinidad and Tobago	0.45**
UK	0.44**
USA	0.31*

**Significant at 5%, *Significant at 10%

$$\begin{aligned}\mathrm{PCG}_t &= a_0 + a_1\mathrm{PCG}_{t-1} + a_2\mathrm{PCG}_{t-2} + \cdots + \mathrm{PCG}_{t-m} \\ &\quad + b_1\mathrm{MK}_{t-1} + b_2\mathrm{MK}_{t-2} + \cdots + b_n\mathrm{MK}_{t-n} + \epsilon_{pcg\ t}\end{aligned} \tag{20}$$

In the regression in Eq. (20), all lagged values of MK_t that are individually significant provided that collectively they add explanatory power to the regression according to an F test whose null hypothesis is no explanatory power jointly added by the different lagged values of MK_t. The null hypothesis that MK_t does not Granger cause PCG_t is not rejected if and only if no lagged values of MK_t are retained in the regression.

Similarly, to test the null hypothesis that PCG_t does not Granger cause MK_t, in the following autoregression

$$\begin{aligned}\mathrm{MK}_t &= \alpha_0 + \alpha_1\mathrm{MK}_{t-1} + \alpha_2\mathrm{MK}_{t-2} + \cdots + \alpha_p\mathrm{MK}_{t-p} \\ &\quad + \beta_1\mathrm{PCG}_{t-1} + \beta_2\mathrm{PCG}_{t-2} + \cdots + \beta_q\mathrm{PCG}_{t-q} + \epsilon_{\mathrm{MK}\,t}\end{aligned} \tag{21}$$

if and only if no lagged values of PCG_t are retained in the regression, the null hypothesis cannot be rejected and it is established that PCG_t does not, indeed, Granger cause MK_t.

In our analysis, we reject the null hypothesis in favour of the alternative hypothesis only if the p value falls below 0.05, in which case causality is established at 5% level of significance. The optimal lag length for each country is chosen by the Akaike Information Criterion. Table 2 summarises the direction of causality between MK_t and PCG_t for each country.

From Table 2, it is evident that for 23 countries (Austria, Belgium, Brazil, Columbia, Denmark, Egypt, Finland, France, Germany, Greece, India, Italy, Malaysia, Netherlands, Norway, Pakistan, Philippines, Portugal, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Trinidad and Tobago, UK and the USA) out of 35 countries, market capitalization Granger causes per capita growth. For the remaining 12 (Australia, Bangladesh, Canada, China, Columbia, Greece, India, New Zealand, Nigeria, Russia, Singapore and South Africa) countries, none of the null hypothesis can be rejected, i.e. for these 12 countries, neither market capitalization Granger causes growth, nor growth Granger causes market capitalization. For three (Euro-area, High-income non-OECD and High-income OECD) of the four country groups, the direction of causality runs from market capitalization to growth, while only for South Asia, no causality is found to exist either from market capitalization to growth or the other way round. For none of the countries and country groups, causality is observed to run from growth to market capitalization ratio. From these findings, it can be inferred that for the majority of the countries and country groups, the direction of causality is found to exist from market capitalization ratio to growth and not the other way round.

Table 2 Granger causality analysis for different countries and country groups

Country/country groups	Direction of causality between MK and PCG
Australia	No Causality
Austria	**MK⟶PCG**
Bangladesh	No causality
Belgium	**MK⟶PCG**
Brazil	**MK⟶PCG**
Canada	No causality
China	No causality
Columbia	No causality
Denmark	**MK⟶PCG**
Egypt	**MK⟶PCG**
Finland	**MK⟶PCG**
France	**MK⟶PCG**
Germany	**MK⟶PCG**
Greece	No causality
India	No causality
Italy	**MK⟶PCG**
Malaysia	**MK⟶PCG**
Netherlands	**MK⟶PCG**
New Zealand	No causality
Nigeria	No causality
Norway	**MK⟶PCG**
Pakistan	**MK⟶PCG**
Philippines	**MK⟶PCG**
Portugal	**MK⟶PCG**
Russia	No causality
Singapore	No causality
South Africa	No causality
Spain	**MK⟶PCG**
Sri Lanka	**MK⟶PCG**
Sweden	**MK⟶PCG**
Switzerland	**MK⟶PCG**
Thailand	**MK⟶PCG**
Trinidad and Tobago	**MK⟶PCG**
UK	**MK⟶PCG**
USA	**MK⟶PCG**
Euro-area	**MK⟶PCG**
South Asia	No Causality
High-income non-OECD	**MK⟶PCG**
High-income OECD	**MK⟶PCG**

Countries in bold letters support the lead–lag relationship between market capitalization and growth

3.3.3 Variance Decomposition Analysis

Most of the theoretical literature on asset pricing looks into the contemporaneous relationship between stock prices and growth. However, in reality the short-run relationship between market capitalization and growth can be of a lead–lag nature, i.e. due to a one-time shock, the change in the behaviour of one of the variables in a certain time period can also have an influence on the other's behavioural pattern during later time periods. The relationship between the short-run fluctuations in market capitalization ratio and growth, as represented in Eq. (19), provides a suitable foundation to test empirically the existence of a lead–lag relationship between market capitalization ratio and growth. Our analysis of Granger causality indicates that the 'lead' role, if any, is likely to be taken by the market capitalization ratio with per capita growth responding with a 'lag'. To empirically test whether there really exists a lead–lag relationship between market capitalization ratio and growth, we conduct a variance decomposition exercise with all the countries within a VAR set-up. In this way, we compare the relative influence of market capitalization in determining future fluctuations of per capita growth with the role played by per capita growth in driving the future fluctuations of market capitalization ratio.

In order to understand whether there exists a lead–lag relationship between market capitalization ratio and growth, we first propose a vector autoregression (VAR) setting. With the help of a variance decomposition analysis, we find the influence of a one-time shock to market capitalization on the short-run fluctuations of growth and also how much a one-time shock to growth accounts for a change in short-run fluctuations of market capitalization.

When it is unclear whether a variable is actually exogenous or not, a natural extension of transfer function analysis is to treat each variable symmetrically. A VAR framework helps us to understand how each variable depends on its lagged (past) values as well as the lagged values of the other variables. Denoting market capitalization as a ratio of GDP at time t by MK_t and per capita growth at time t by PCG_t, the time path of MK_t can be allowed to be affected by current and past realizations of the sequence of PCG_t, and likewise, the PCG_t sequence can be affected by current and past realizations of the MK_t sequence in the following pth order VAR set-up:

$$\begin{aligned}\mathrm{MK}_t &= c_{mk} - a^0_{mk\,pcg}\mathrm{PCG}_t + a^1_{mk\,mk}\mathrm{MK}_{t-1} + a^1_{mk\,pcg}\mathrm{PCG}_{t-1} + a^2_{mk\,mk}\mathrm{MK}_{t-2} \\ &\quad + a^2_{mk\,pcg}\mathrm{PCG}_{t-2} + \cdots + a^p_{mk\,mk}\mathrm{MK}_{t-p} + a^p_{mk\,pcg}\mathrm{PCG}_{t-p} + \epsilon_{mk\,t} \end{aligned} \tag{22}$$

$$\begin{aligned}\mathrm{PCG}_t &= c_{pcg} - a^0_{pcg\,mk}\mathrm{MK}_t + a^1_{pcg\,mk}\mathrm{MK}_{t-1} + a^1_{pcg\,pcg}\mathrm{PCG}_{t-1} + a^2_{pcg\,mk}\mathrm{MK}_{t-2} \\ &\quad + a^2_{pcg\,pcg}\mathrm{PCG}_{t-2} + \cdots + a^p_{pcg\,mk}\mathrm{MK}_{t-p} + a^p_{pcg\,pcg}\mathrm{PCG}_{t-p} + \epsilon_{pcg\,t} \end{aligned} \tag{23}$$

The first model denoted by Eq. (22) describes how market capitalization as a ratio of GDP at time t (MK_t) depends on its own past values up to p lags as well as the past values of per capita growth PCG also up to p lags. MK_{t-p} denotes the value of market capitalization as a ratio of GDP p periods before time period t. Similarly, PCG_{t-p} denotes per capita growth p periods before time period t. In a similar manner, the second model denoted by Eq. (23) describes how per capita growth at time period t (PCG_t) depends on its own past values up to p lags as well as the past values of market capitalization ratio MK also up to p lags.

c_{mk} and c_{pcg} denote the constant terms, and the coefficient $a^k_{i,j}$ denotes the measure of the dependence of i on j, where $i = mk,\ pcg$ and $j = mk,\ pcg$.

It is assumed that both MK_t and PCG_t are stationary, $\epsilon_{mk\,t}$ and $\epsilon_{pcg\,t}$ are white noise disturbances with standard deviations σ_{mk} and σ_{pcg}, respectively, and $\{\epsilon_{mk\,t}\}$ and $\left\{\epsilon_{pcg\,t}\right\}$ are uncorrelated white noise disturbances.

The structure of the system represented by Eqs. (22) and (23) incorporates feedback because MK_t and PCG_t are allowed to affect each other. For example, $a^0_{mk\,pcg}$ is the contemporaneous effect of a unit change of PCG_t on MK_t and $a^1_{mk\,pcg}$ is the effect of a unit change in PCG_{t-1}on MK_t. The terms $\epsilon_{mk\,t}$ and $\epsilon_{pcg\,t}$ are pure innovations or shocks in MK_t and PCG_t, respectively. Thus, assuming coefficients $a^k_{i,j} \neq 0$, a shock to market capitalization ratio (per capita growth) in time period t will indirectly affect per capita growth (market capitalization ratio) contemporaneously in time period t as well as in future time periods.

Now, in order to investigate whether a lead–lag relationship exits between market capitalization ratio and growth, we have to figure out how market capitalization in a certain time period affects growth in later time periods and also how growth affects market capitalization in future, both in the short run and in the long run. For this, we examine the forecast error variance decompositions, which indicate the percentage of the variance in one variable that is due to errors in forecasting itself as well as the other variable. In other words, the forecast error variance decomposition is used to understand the proportion of movements in a sequence due to its own shocks versus shocks to the other variable. If $\epsilon_{pcg\,t}$ shocks explain none of the forecast error variance of MK_t at all forecast horizons, it can be inferred that the $\{\mathrm{MK}_t\}$sequence is exogenous. In this case, MK_t evolves independently of the $\epsilon_{pcg\,t}$ shocks and of the $\{\mathrm{PCG}_t\}$ sequence. At the other extreme, $\epsilon_{pcg\,t}$ could explain all the forecast error variance in the $\{\mathrm{MK}_t\}$ sequence at all forecast horizons, making MK_t entirely endogenous.

However, a variance decomposition analysis of market capitalization as a ratio of GDP, i.e. MK_t, and growth of per capita GDP, i.e. PCG_t, establish that for all 35 countries and four country groups, neither MK_t nor PCG_t is entirely exogenous or endogenous, i.e. forecast error variance in each of these variables is influenced partly due to its own shock and partly as a result of a shock to the other variable. Although in each period, short-run fluctuations of MK_t and PCG_t are caused due to shocks in both these variables, in order to clearly understand which of the variables has a greater impact on the other's behaviour during the future time periods, we focus only on the shock impact of a particular variable in explaining the fluctuations of the other variable. In other words, as a result of the variance decomposition analysis, we compare the influence of a one-time shock to MK_t in explaining short-run movements of PCG_t during the later periods with the effect of a one time shock to PCG_t in explaining the future short-run fluctuations in MK_t.

For our analysis, we choose the number of time periods as 10, but report our findings only for period 2 and period 10, with period 2 taken as the short run and period 10 as the comparatively longer run. For all the countries and country groups, the percentage of the total change in fluctuations of PCG_t brought about due to a one-time shock on MK_t, is reported both for period 2 and for period 10. Similarly, the percentage of the variance in fluctuations in MK_t explained by a one-time shock to PCG_t is also reported for both long-run and short-run scenarios. We calculate the business cycle fluctuations using the Christiano–Fitzgerald filter, separating out the cyclical component from the trend component of a series. In this way, for each of the thirty-three countries, we obtain the short-run fluctuations in MK_t as well as PCG_t. These short-run business cycle fluctuations represent a deviation from the trend for each variable. Since we are interested to investigate which variable is the main driving force towards a change in short-run behaviour of the other variable, we choose to carry out a variance decomposition analysis of the short-run fluctuations of MK_t and PCG_t for each of the countries and country groups. The optimal lag length for this analysis is selected according to the Akaike Information Criteria. Tables 3 and 4 report the influence of MK_t on PCG_t and vice versa for the 2nd period and the 10th period respectively, treating MK_t as the exogenous variable in the Cholesky ordering for VAR. The countries for which MK_t is the main determinant of the short-run behaviour of PCG_t during the later periods are highlighted in each of the following tables.

From Table 3, it follows that for all countries except Russia, the variance in fluctuations of per capita growth in the 2nd period due to a one-time shock to market capitalization ratio is far more than the short-run variance in fluctuations of market capitalization due to a one-time shock to per capita growth when the market capitalization ratio is treated as the exogenous variable in the Cholesky ordering of the VAR set-up.

From Table 4, it follows that in the tenth period also, market capitalization ratio plays a much more dominant role in determining the fluctuations of per capita growth, compared to the role played by per capita growth in determining the cyclical behaviour of market capitalization ratio, when market capitalization ratio is treated as the exogenous variable in the Cholesky ordering. However, in the tenth period, there is an increase in the number of countries for which per capita growth plays a much more dominant role in determining the short-run fluctuations of market capitalization ratio. This number increased from one in the second period to six in the tenth period; the six countries being Australia, Austria, Malaysia, Nigeria, Portugal and Russia. From Tables 3 and 4, it is clear that Russia is the only country for which per capita growth as a driving force behind the cyclical fluctuations of market capitalization ratio dominates the market capitalization ratio both in the short run and in the comparatively long run. Australia, Austria, Malaysia, Nigeria and Portugal were the countries for which the percentage of fluctuations in per capita growth as

Table 3 Short-run ($p = 2$) forecast error decomposition of MK as explained by PCG and that of PCG explained by MK (Cholesky ordering: MK, PCG)

Country/groups	% MK variance explained by PCG (%)	% PCG variance explained by MK (%)
Australia	**5.29**	**6.33**
Austria	**13.21**	**14.36**
Bangladesh	**2.13**	**46.26**
Belgium	**9.48**	**46.40**
Brazil	**7.14**	**77.22**
Canada	**4.27**	**49.45**
China	**8.08**	**41.94**
Columbia	**6.44**	**31.32**
Denmark	**0.44**	**79.06**
Egypt	**0.51**	**51.84**
Finland	**3.81**	**63.85**
France	**3.38**	**60.03**
Germany	**4.45**	**67.26**
Greece	**3.63**	**64.16**
India	**2.88**	**37.17**
Italy	**0.25**	**31.80**
Malaysia	**0.28**	**52.64**
Netherlands	**1.82**	**72.65**
New Zealand	**1.83**	**30.18**
Nigeria	**2.16**	**2.18**
Norway	**1.14**	**39.10**
Pakistan	**7.46**	**40.25**
Philippines	**2.87**	**30.98**
Portugal	**2.04**	**69.26**
Russia	25.47	15.82
Singapore	**2.86**	**67.92**
South Africa	**4.49**	**48.32**
Spain	**47.29**	**65.52**
Sri Lanka	**3.26**	**48.49**
Sweden	**1.42**	**72.17**
Switzerland	**1.50**	**68.48**
Thailand	**2.82**	**56.25**
Trinidad and Tobago	**4.04**	**20.22**
UK	**1.30**	**51.51**
USA	**1.44**	**66.47**
Euro-area	**1.11**	**87.95**
South Asia	**0.06**	**34.87**
High-income non-OECD	**0.50**	**83.57**
High-income OECD	**0.11**	**91.45**

Countries in bold letters support the lead–lag relationship between market capitalization and growth

Table 4 Short-run ($p = 10$) forecast error decomposition of MK as explained by PCG and that of PCG explained by MK (Cholesky ordering: MK, PCG)

Country/groups	% MK variance explained by PCG (%)	% PCG variance explained by MK (%)
Australia	9.22	7.57
Austria	49.17	14.69
Bangladesh	**15.43**	**45.33**
Belgium	**20.35**	**58.41**
Brazil	**22.67**	**75.67**
Canada	**5.86**	**66.31**
China	**34.04**	**42.20**
Columbia	**6.43**	**41.38**
Denmark	**6.89**	**82.41**
Egypt	**7.80**	**70.99**
Finland	**21.34**	**65.34**
France	**12.86**	**70.04**
Germany	**14.39**	**67.92**
Greece	**27.20**	**36.83**
India	**6.81**	**61.20**
Italy	**0.43**	**36.59**
Malaysia	44.44	39.19
Netherlands	**14.45**	**69.41**
New Zealand	**3.16**	**28.95**
Nigeria	32.50	8.70
Norway	**10.13**	**50.85**
Pakistan	**13.42**	**40.99**
Philippines	**4.23**	**30.64**
Portugal	51.78	41.87
Russia	36.97	22.48
Singapore	**22.09**	**59.45**
South Africa	**6.58**	**66.76**
Spain	**62.64**	**46.32**
Sri Lanka	**3.58**	**51.09**
Sweden	**3.21**	**80.62**
Switzerland	**4.45**	**72.34**
Thailand	**4.09**	**57.84**
Trinidad and Tobago	**6.50**	**24.59**
UK	**1.64**	**66.97**
USA	**2.88**	**77.93**
Euro-area	**7.82**	**89.19**
South Asia	**6.92**	**34.01**
High-income non-OECD	**3.29**	**91.14**
High-income OECD	**3.14**	**93.75**

Countries in bold letters support the lead–lag relationship between market capitalization and growth

determined by market capitalization ratio is found to dominate those in market capitalization ratio determined by per capita growth only for the shorter run, but not for the comparatively longer run.

However, throughout the above analysis of variance decomposition, market capitalization ratio has been treated as the exogenous variable in the Cholesky ordering. It will be interesting to figure out the extent of forecast error decomposition of market capitalization ratio as explained by per capita growth and that of per capita growth as explained by market capitalization ratio when the Cholesky ordering is reversed, i.e. per capita growth is treated as the exogenous variable and market capitalization ratio is treated as the endogenous variable. Table 5 depicts the short-run effect of a one-time shock to per capita growth on the fluctuations in the market capitalization ratio and vice versa in the second period, and Table 6 depicts the effect of market capitalization on the cyclical behaviour of per capita growth in the tenth period when per capita growth instead of market capitalization ratio is treated as the exogenous variable in the Cholesky ordering.

From Table 5, it is clear that even when per capita growth is treated as the exogenous variable, a one-time shock to market capitalization ratio has more influence on the short-run fluctuations of per capita growth than the impact of a one-time shock to per capita growth on the cyclical behaviour of market capitalization ratio in the second period for the majority of countries and country groups. For 26 out of 35 countries and 2 out of 4 country groups, the percentage of fluctuations of per capita growth as explained by market capitalization ratio is greater than the fluctuations of market capitalization ratio explained by per capita growth. The nine countries for which per capita growth is the main driving force towards the cyclical behaviour of market capitalization ratio during the later time periods are Australia, Greece, India, Netherlands, New Zealand, Norway, Pakistan, Russia and Spain, while among the country groups, this holds true for the two country groups of South Asia and high-income non-OECD.

From Table 6, it is evident that when per capita growth is treated as the exogenous variable in the Cholesky ordering, still for 26 of the 35 countries and 3 out of the 4 country groups, market capitalization ratio can explain more of the short-run fluctuations of per capita growth than the corresponding impact of per capita growth on the cyclical behaviour of market capitalization ratio in the 10th time period. The 9 countries for which per capita growth is the main driving force towards the cyclical behaviour of market capitalization ratio during the later time periods are Australia, Greece, India, Netherlands, New Zealand, Norway, Pakistan, Russia and Spain, while among the country groups, this holds true only for South Asia.

From Tables 5 and 6, it follows that for Greece, India, Netherlands, New Zealand, Norway, Russia, Spain and South Asia, market capitalization is not the chief determinant of fluctuations in per capita growth throughout the initial and later time periods, when per capita growth is treated as the exogenous variable in the Cholesky ordering. For Australia, Pakistan and the high-income non-OECD country group, only for the tenth period and not for the second period, the percentage of fluctuations in per capita growth determined by market capitalization is found to be larger than the percentage of fluctuations in market capitalization ratio determined by per capita

Table 5 Short-run ($p = 2$) forecast error decomposition of MK as explained by PCG and that of PCG explained by MK (Cholesky ordering: PCG, MK)

Country/groups	% MK variance explained by PCG (%)	% PCG variance explained by MK (%)
Australia	6.92	0.83
Austria	**0.34**	**10.22**
Bangladesh	**0.15**	**1.05**
Belgium	**21.13**	**45.69**
Brazil	**33.40**	**48.01**
Canada	**12.38**	**32.23**
China	**4.89**	**55.88**
Columbia	**7.45**	**30.71**
Denmark	**1.88**	**75.41**
Egypt	**13.57**	**54.15**
Finland	**4.65**	**32.54**
France	**8.67**	**53.18**
Germany	**2.94**	**71.62**
Greece	15.27	8.70
India	40.11	1.58
Italy	**6.44**	**36.28**
Malaysia	**8.15**	**59.60**
Netherlands	62.95	1.53
New Zealand	31.65	2.01
Nigeria	**0.68**	**1.78**
Norway	25.25	8.11
Pakistan	20.35	6.68
Philippines	**3.76**	**28.98**
Portugal	**31.09**	**42.13**
Russia	23.36	14.02
Singapore	**15.62**	**19.68**
South Africa	**17.16**	**29.34**
Spain	63.17	13.08
Sri Lanka	**0.49**	**50.44**
Sweden	**20.81**	**51.09**
Switzerland	**25.10**	**52.20**
Thailand	**10.54**	**42.05**
Trinidad and Tobago	**3.69**	**20.09**
UK	**15.49**	**27.88**
USA	**14.67**	**51.97**
Euro-area	**11.30**	**49.69**
South Asia	34.69	1.42
High-income non-OECD	18.96	5.92
High-income OECD	**18.21**	**44.35**

Countries in bold letters support the lead–lag relationship between market capitalization and growth

Table 6 Short-run ($p = 10$) forecast error decomposition of MK as explained by PCG and that of PCG explained by MK (Cholesky ordering: PCG, MK)

Country/groups	% MK variance explained by PCG (%)	% PCG variance explained by MK (%)
Australia	**11.19**	**12.76**
Austria	**0.32**	**16.61**
Bangladesh	**0.33**	**1.25**
Belgium	**24.59**	**53.01**
Brazil	**32.48**	**58.95**
Canada	**13.57**	**50.79**
China	**7.65**	**57.56**
Columbia	**20.07**	**39.83**
Denmark	**6.76**	**83.60**
Egypt	**13.12**	**60.67**
Finland	**5.86**	**53.78**
France	**17.08**	**66.27**
Germany	**16.06**	**67.03**
Greece	20.35	10.94
India	40.33	4.75
Italy	**7.14**	**42.68**
Malaysia	**8.21**	**64.04**
Netherlands	39.96	6.86
New Zealand	32.19	15.40
Nigeria	4.06	2.21
Norway	26.58	24.20
Pakistan	**20.05**	**23.57**
Philippines	**5.13**	**28.67**
Portugal	51.78	41.87
Russia	39.95	24.45
Singapore	**0.48**	**23.28**
South Africa	**17.25**	**57.55**
Spain	51.29	25.88
Sri Lanka	**3.27**	**53.02**
Sweden	**20.76**	**62.19**
Switzerland	**26.49**	**66.25**
Thailand	**12.82**	**47.78**
Trinidad and Tobago	**6.17**	**24.58**
UK	**16.87**	**41.54**
USA	**16.03**	**64.28**
Euro-area	**14.17**	**60.01**
South Asia	34.41	1.54
High-income non-OECD	**21.54**	**26.58**
High-income OECD	**18.86**	**60.70**

Countries in bold letters support the lead–lag relationship between market capitalization and growth

growth. On the other hand, for Nigeria and Portugal, the percentage of fluctuations in per capita growth as determined by market capitalization ratio is found to dominate those in market capitalization ratio determined by per capita growth for the second period, but not for the tenth period.

Thus, for eight countries, market capitalization ratio determines per capita growth fluctuations neither in the short run, nor in the long run when per capita growth is treated as the exogenous variable in the Cholesky ordering; these eight countries being Greece, India, Netherlands, New Zealand, Norway, Russia, Spain and South Asia. However, this number is only one (Russia) when market capitalization ratio is treated as exogenous in the ordering.

On the whole, irrespective of the nature of Cholesky ordering, the variance decomposition analysis establishes that for most countries, market capitalization is the main determinant of future fluctuations of per capita growth both in the short run and in the comparatively long run. This further establishes the lead–lag relationship between market capitalization ratio and per capita growth.

4 Concluding Remarks

While the extant literature is concerned with the causal relationship between stock market development and growth of per capita income, the present paper looks at a possible lead–lag relationship between the two variables. It is shown, both theoretically and empirically, that there is indeed a lead–lag relationship with the market capitalization ratio representing stock market behaviour taking the lead and per capita growth following with a lag. The implication of this analysis is that variations in per capita income can be partly explained by variations in the lagged values of the market capitalization ratio. This, in turn, provides a new dimension to explain per capita income variations, without specifying causality.

A substantial literature has evolved around the question as to whether movements in current financial variables can predict movements in future real economic activity. The present paper attempts to contribute to this literature. Apart from providing empirical supports to the hypothesis that current movements in financial variables reflect future movements of real variables, it provides a theoretical framework through which this might happen.

References

Abel, A. B. (1988, October). Stock prices under time varying dividend risk: An exact solution in an infinite horizon general equilibrium model. *Journal of Monetary Economics*. (Elsevier Science Publishers B.V., North-Holland).

Apergis, N., Artikis, P., & Kyriazis, D. (2015). Does stock market liquidity explain real economic activity? New evidence from two large European stock markets. *Journal of International Financial Markets, Institutions and Money*, *38*, 42–64.

Aylward, A., & Glen, J. (2000). Some international evidence on stock prices as leading indicators of economic activity. *Applied Financial Economics*, *10*, 1–14.

Beck, T., Demirguc-Kunt, A., & Levine, R. (2000). A new database on the structure and development of the financial sector. *World Bank Economic Review*, *14*(3), 597–605.

Bencivenga, V. R., & Smith, B. D. (1991). Financial intermediation and endogenous growth. *The Review of Economic Studies*, *58*(2), 195–209.

Bencivenga, V. R., Smith, B. D., & Starr, R. M. (1995). *Equity markets, transaction costs and capital accumulation*. Policy Research Working Paper Series 1456, The World Bank.

Brock, W. A. (1982). Asset prices in a production economy, UMI. In J. McCall (Ed.), *The economics of information and uncertainty* (pp. 1–46).

Brock, W. A., & Mirman, L. J. (1972). Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory*, *4*, 479–513.

Chen, Y., Eaton, G., & Paye, B. (2018). Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity. *Journal of Financial Economics*, *130*(1), 48–73.

Choi, J., Hauser, S., & Kopecky, K. (1999). Does stock market predict real economic activity? Time series from G7 countries. *Journal of Banking and Finance*, *23*, 1771–1792.

Cochrane, J. H. (1991). Production-based asset pricing and the link between stock returns and economic fluctuations. *The Journal of Finance*, *46*(1), 209–237.

Estrella, A., & Hardouvelis, G. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *XLVI*(2), 555–576.

Fama, E. F. (1965). The behaviour of stock market prices. *The Journal of Business*, *38*(1), 34–105.

Fase, M. M. G., & Abma, R. C. N. (2003). Financial development and economic growth in selected Asian countries. *Journal of Asian Economics*, *14*(2003), 11–21.

Goldsmith, R. W. (1969). *Financial structure and development*. New Haven, CT: Yale University Press.

Gorton, G., & Pennacchi, G. (1990). Financial intermediaries and liquidity creation. *Journal of Finance*, *45*(1), 49–71.

Greenwood, J., & Jovanovic, B. (1990). Financial development, growth and the distribution of income. *Journal of Political Economy*, *98*(5(1)), 1076–1107.

Hicks, J. (1969). *A theory of economic history*. Oxford: Clarendon Press.

Jacklin, C. J. (1987). Demand deposits, trading restrictions, and risk sharing. In E. Prescott & N. Wallace (Eds.), *Contractual arrangements for international trade* (pp. 26–47). Minneapolis, MN: University of Minnesota Press.

Japelli, T., & Pagano, M. (1994). Saving, growth, and liquidity constraints. *The Quarterly Journal of Economics*, *109*(1), 83–109.

King, R. G., & Levine, R. (1993a). Financial intermediation and economic development. In C. Mayer & X. Vives (Eds.), *Financial intermediation in the construction of Europe* (pp. 156–89). London: Centre for Economic Policy Research.

King, R. G., & Levine, R. (1993b). Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics*, *108*(3), 717–37.

King, R. G., & Levine, R. (1993c). Finance entrepreneurship and growth: Theory and evidence. *Journal of Monetary Economics*, *32*(3), 513–42.

Levanon, G., Manini, J., Ozyildrim, A., & Schaitkin, B. (2015). Using financial indicators to predict turning points in business cycle: The case of the leading economic index for the United States. *International Journal of Forecasting*, *31*(2), 426–445.

Levine, R. (1991). Stock markets, growth and tax policy. *The Journal of Finance*, *46*(4), 1445–1465.

Levine, R., & Zervos, S. (1996). Stock market development and long run growth. *The World Bank Economic Review*, *10*(3), 323–339.

Ljungqvist, L., & Sargent, T. (2004). *Recursive macroeconomic theory* (2nd ed.). Cambridge, MA: MIT Press.

Lucas, R. E., Jr. (1978, November). Asset prices in an exchange economy. *Econometrica*, *46*(6), 1429–1445.

Mills, L. (1988). Can stock prices reliably predict recessions? *Business Review, Federal Reserve Bank of Philadelphia*, 1–12.

Rajan, R. G., & Zingales, L. (1998). Financial systems, industrial structure and growth. *Oxford Review of Economic Policy*, *17*(4), 467–482.

Reinhart, C., & Rogoff, K. (2009). The aftermath of financial crisis. *American Economic Review*, *99*(2), 466–72.

Sarkar, A., & Sarkar, A. (2019). Do stock markets signal real economic performance? A theoretical and empirical note. In S. Sikdar (Ed.), *Forthcoming in a Conference Volume of IIM Calcutta*.

Schumpeter, J. A. (1934) [1912]. *The theory of economic development* (R. Opie, Trans.). Cambridge, MA: Harvard University Press.

A Study of Volatility of Five Major Stock Indices of Indian Stock Market

Ayanangshu Sarkar and Malabika Roy

Abstract This paper is an attempt to understand how far the volatility in the world stock market has penetrated Indian stock market and the nature of volatility of stock indices at the sectoral level during a period covering the subprime crisis. Most of the previous studies on sectoral returns are related to diversification and optimum portfolio management. But it is also important to understand the characteristics of the volatility of the individual sectoral returns as well as their relationship with market returns. This paper empirically investigates the pattern of volatility at the sectoral level in the Indian stock market during April 1, 2006, to March 31, 2011, and any 'spillover effect' between the domestic sectors and the US stock markets, which gives us an idea about how deep the penetration of the subprime crisis was in the Indian stock market and how persistent was its effect. The study focuses on the time-varying nature of volatility and the presence of characteristics like volatility clustering. The disaggregated study gives us an idea of how far the different segments of stock markets are integrated with the international stock markets. It also addresses the issue of leverage effect for all the indices to check whether a negative shock is creating more volatility than positive shocks. It also gives us the opportunity to compare the level of persistence among pre-crisis, crisis and post-crisis period.

1 Introduction

Indian capital market has experienced tremendous changes since 1991 when the government adopted New Economic Policy, which was based on LPG model, that is liberalization, privatization and globalization of industries, trade and more importantly financial markets. The extent of integration of Indian stock market with the

A. Sarkar
Department of Business Administration, Pailan College of Management & Technology, Kolkata, India
e-mail: ayanangshusarkar2@gmail.com

M. Roy (✉)
Department of Economics, Jadavpur University, Kolkata, India
e-mail: malabikar@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_5

world market increased as a result. Indian stock market, being one of the emerging markets in the world, showed remarkable growth. From 3739.69 points on March 31, 1999, Bombay Stock Exchange (BSE) Sensitivity Index (SENSEX) reached to 21,000 level points in January 2008.

This paper is an attempt to understand how deep the penetration of the world market volatility was as a result of integration with world market. To this end, we study the variance of returns at the sectoral level and the extent of spillover from the world market during, before and after the subprime crisis. The previous studies on sectoral returns are related to portfolio diversification and optimum portfolio management. But it is also important to understand the characteristics of the volatility of the sectoral returns individually and together with market returns. The present paper is an attempt to fill up this lacuna.

First, this paper empirically investigates the pattern of volatility at the sectoral level in the Indian stock market covering a pre-crisis, crisis and post-crisis period. To this end, five major sectors were chosen: banking (BANKEX), information technology (IT), health care (HC), fast-moving consumer goods (FMCG), oil and gas (OILGAS) during April 1, 2006, to March 31, 2011. These sectors were chosen for two reasons: First, the stocks of the firms belonging to these sectors cover a large part of Indian stock market. Secondly, economically speaking, these sectors are more integrated with the world market. The study focuses on the time-varying nature and presence of characteristics like volatility clustering. Secondly, this paper estimates whether there was any 'spillover effect' between the domestic sectors and the US stock market during the subprime crisis as well as in the pre-crisis and post-crisis period and if so, whether the nature of spill over effect changed.

Choice of the US stock market as a representative stock market needs some justifications. Table 1 gives some idea of the flow of portfolio investment to and from the USA. For most of the years, the USA held the second/third position in terms of flow

Table 1 (Values are in million USA $s): Portfolio investment to and from the USA

Year	Investment in the USA by India	Total portfolio investment by India	Percentage (%)	Investment from the USA in India	Total portfolio investment coming to India	Percentage (%)
2011	143	1194	12	58,594	251,875	23
2010	379	1583	24	91,486	370,818	25
2009	291	1366	24	65,495	325,696	20
2008	116	949	12	31,917	198,402	16
2007	54	1104	4.90	85,149	359,418	23
2006	31	284	11	49,231	183,115	27
2005	27	81	33	33,226	117,509	28

Source Constructed from Coordinated Portfolio Investment Survey (IMF): 2005–2011 of portfolio investment.[1] Table 1 shows how very important a position USA enjoys if we consider the flow of portfolio investment to and from India.

The other reason for choosing the US stock market was that subprime crisis, which was the major financial crisis of the period under study started in the US economy. As one of our aims was to study the sectoral penetration of the subprime crisis in the Indian stock market, the US stock market seems to be the obvious choice.

The disaggregated study gives us an idea to what extent the different segments of stock markets are integrated with the US stock market. The paper also addresses the issue of leverage effect in case of all the indices to check whether a negative shock is creating more volatility than positive shock at the sectoral level. Calculation of the half-life of a shock in different sectors is another objective of this study. This exercise gives us an idea about how persistent the effect of the shock was at a disaggregated level during pre-crisis, crisis and post-crisis period and compared the level of persistence across the three periods. The paper is organized as follows: Sect. 2 provides the survey of the literature. In Sect. 3, the methodology, descriptive statistics and diagnostic tests are described. In Sect. 4, we model sectoral volatility using GARCH technique. Section 5 discusses the volatility spillover and leverage effects. Section 6 concludes.

2 Survey of Literature

The study of sectoral returns related to diversification and optimum portfolio management as the capital markets gets more integrated has interested the researchers for a long time. However, the study of volatility of sectoral returns has been limited. Studies are usually focused on diversification of portfolio either by investing in the different sectoral indices in the same market or by investing in different global markets under the same sector. The study by Meric et al. (2008) found that the global diversification in the bull market of the same sector is better compared to diversification in an individual market across different sectors. Especially in the bear market, all sectors in the local market are correlated and hence global diversification would provide better returns. Although our study does not focus on this aspect, this result can be tested in light of our result if specific sector proved to be more integrated internationally in crisis or post-crisis period.

Demirer and Lien (2005) worked on the correlation between the different related sectoral indices with market movement in either direction that is while the stock index is increasing vis-a-vis while it is falling. The study found that the sectoral correlation is higher in the upside movement of the market or in other words integration gets better in the upswing of stock index.

The study by Wang et al. (2005) showed that the information on Chinese stock markets creates higher dependence of stock exchanges on each other or in simple

[1]The first/second position is held by countries like Mauritius and Luxemburg where the portfolio investment is driven by tax considerations.

words availability of information is the binding force for stock index. Industry sector was found to be most integrated with the impact on each other due to the information flow, while the finance sector has minimum integration with other sectors.

You and Daigler (2010) argued that the decision on the basis of simple constant correlation to diversify portfolios is misleading. Instead, dynamic nature of the correlation between the different indices in the markets of the world would provide a better analysis. The use of the conditional correlation, eventual tail risks, skewness and the standard deviation analysis would result in better diversification decision.

There are some studies, which cover the same ground as our study. Malik and Ewing (2009) focused on the weekly returns for fifteen years in different sectors and the transmission of volatility. They employed bivariate GARCH models to simultaneously estimate the mean and conditional variance between five different US sector indices and oil prices. The study established the relation between the different sectors. It was found that if a shock is introduced in the oil sector, the volatility will be transmitted to other sectors. That is volatility of one sector influenced the volatility of dependent sectors.

Patra and Poshakwale (2008) studied the sectoral indices data of the Athens stock exchange. It was estimated that there was weaker relation in the sectoral returns in the long term. However, there was a significant impact of the banking sector on the variance and return of other sectoral indices. This paper suggested the changes and information of the banking sector could be used to predict the returns of the other sectoral indices in the short term. This research can be further investigated as the banking sector is one of the five major sectors we have selected for investigation.

A similar study was conducted by Lafuente and Ruiz (2004) on the data of the Spanish new market index. The paper focuses on the return of a sector against volatility of that sectoral index. The study found a significant impact of the volatility of the new market index on the other sectors like finance, industry etc.

Mazouz and Saadouni (2007) worked on the FTSE 100 index. They employed GARCH and residual variance of the single index model. Their focus was to study the impact on liquidity characteristics in the time of index revision by addition or deletion of new and old stocks, respectively. This study found that the added or deleted stocks reflected permanent change in terms of price and volume.

As volatility can be considered as a proxy of risk so the research of McMillan and Kambouroudis (2009) is of interest to this study. They used GARCH and APARCH models to determine the forecasted values for the risk parameters in G7 and the Asian countries. According to them, GARCH models provided superior prediction compared to VAR forecasting methods.

Joseph (2003) studied the impact of changes in the FOREX market and interest rates on stock market returns and found there was no impact of Forex and interest rate changes on stock market returns. However, volatility or variance in the sectoral indices was found to influence future performances.

The existing literature gives conflicting results in terms of spillover, so in our study we have introduced one-period lag return of Dow Jones in the EGARCH equation to test the possibility of direct and differential impact of spillover in domestic subsectors.

3 Methodology, Descriptive Statistics and Diagnostic Tests

In this section, we discuss the methodology adopted for the study and discuss the descriptive statistics and the results of diagnostic tests. Before analyzing the nature of volatility characteristics, we have converted daily stock prices to daily returns. The present study uses the logarithmic difference of sectoral indices of two successive periods for the calculation of rate of return for each sub-sectors. That is, if I_t be the closing level of a sectoral index on date t and I_{t-1} be the same for its previous business day, i.e., omitting intervening weekend or stock exchange holidays, then the one day return on the market portfolio is calculated as

$$r_t = \ln\left(\frac{I_t}{I_{t-1}}\right) * 100$$

where $\ln(z)$ is the natural logarithm of 'z.' The data of sectoral returns are taken from the Bombay Stock Exchange as quoted in Yahoo Finance Historical Price Index. ADF and Phillips–Perron test are used to check the stationarity. Modeling volatility characteristic utilizes ARCH and GARCH family of model.

Descriptive statistics of return series are reported in Table 2. BANKEX is found to have the highest average intraday return during the period under study, while health care has minimum return. The range shows that OILGAS sector experienced highest variation in return in terms of range, and the highest intraday return for the oil–gas sector was 7.59% while the lowest was recorded −7.04%. If standard deviation is considered as the proxy of historical volatility, then banking sector showed the highest volatility with sd 1.08 followed by OILGAS and IT. This gives a rough idea that these three sectors have the highest integration with the rest of the world as the

Table 2 Descriptive statistics of the domestic indices[a]

	RTBANK	RTFMCG	RTHC	RTIT	RTOG
Mean	0.032285	0.016356	0.014990	0.015929	0.025050
Median	0.064175	0.017895	0.036310	0.011615	0.024539
Maximum	7.621140	5.760416	3.858558	5.685551	7.593415
Minimum	−5.856387	−4.002194	−3.767577	−4.626202	−7.040383
Std. Dev.	1.083415	0.705897	0.619134	0.918649	0.981039
Skewness	0.133183	0.149200	−0.503147	0.240909	−0.283798
Kurtosis	7.177868	8.791274	9.193537	6.558331	10.38995
Jarque–Bera	895.9930	1719.226	2012.920	659.1989	2808.478
Probability	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	39.61427	20.06849	18.39302	19.54463	30.73644
Sum Sq. Dev.	1439.065	610.9040	469.9595	1034.640	1179.949
Observations	1227	1227	1227	1227	1227

[a]RT stands for return. *Source* Yahoo Finance Historical Price Index

period of study comprises of the period of subprime crisis and world experienced one of the greatest recessions in financial sector, they suffered from the maximum variability.

Skewness and kurtosis are important because few investment returns are normally distributed. Investors often predict future returns based on standard deviation that is volatility, but such predictions assume a normal distribution. In reality, investment returns are not perfectly normally distributed. Skewness and kurtosis measure how its distribution differs from a normal distribution and therefore provide an indication of the reliability of predictions based on the standard deviation. The descriptive statistics show that for BANKEX, FMCG and IT both Skewness and mean are positive indicating better investment areas as positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. In case of health care and oil–gas, although return that is mean is positive, they are negatively skewed. So for these two sectors, there exists possibility of some extreme losses even with positive mean return. (refers to the degree of peak in a distribution). More peak than normal (leptokurtic) means that a distribution also has fatter tails and that there are more chances of extreme outcomes compared to a normal distribution. Stock returns are typically characterized by high kurtosis. Estimates show that for all the indices, kurtosis is much higher than three proving the possibility of extreme outcome of these return of indices in the period concern.

Stationarity of the indices is established by using both augmented Dickey–Fuller test and Phillips–Perron test. The log return series of all the indices are found to be stationary.

4 Volatility Modeling for Sectoral Indices Using GARCH

In econometric literature, usually volatility clustering is modeled by ARCH process. In this section, however, we use GARCH model to justify volatility clustering and to calculate half-life of a shock to the five domestic sectors, namely banking, health care, IT, FMCG and OIL–GAS. We have used GARCH(1,1) model as we have found GARCH(1,1) model better fit BSE SENSEX compared to GARCH(2,1) or GARCH(1,2)[2] for the period under consideration. As these indices of sub-sectors are also listed in BSE, so we use GARCH(1,1) model to explain volatility characteristics.

We estimate volatility both for the entire period of April 2006 to March 2011 as well as for the three sub-periods: pre-crisis, crisis and post-crisis as demarcated by the Chow test (Fig. 1). The purpose is to check whether the pattern of volatility changed during these sub-periods.

The section consists of three subsections: The first section provides the details of diagnostic test as a prerequisite to estimate GARCH. The second section discusses the results of GARCH(1,1) estimation to check volatility clustering. The final section calculates half-life of shock for the five sectors.

[2]Results are available with the authors.

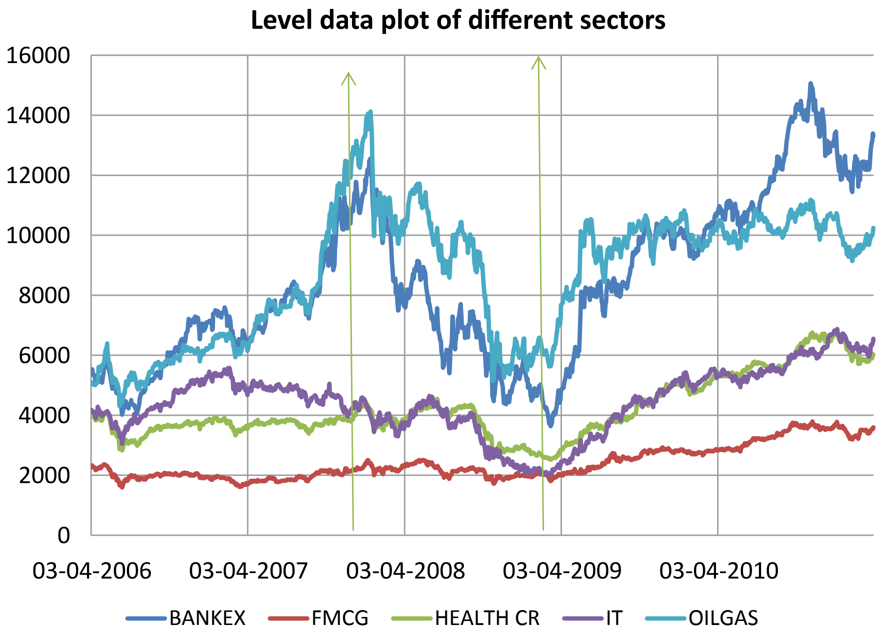


Fig. 1 Level data plots of different sectors. *Source* Constructed from Yahoo Finance Historical Price Index. Level data plots of the five sectors show that banking and oil–gas suffered from the highest fluctuation. Health care and IT have shown more or less similar movement while FMCG suffered least from the impact of the shock

4.1 Diagnostic Test to Run GARCH

As part of the diagnostics, we begin with a visual inspection of the plot of daily returns of BANKEX, FMCG, HC, IT and OG as shown in Fig. 2. It can be seen that returns continuously fluctuate around a mean value that is close to zero, and fluctuations with bigger magnitudes are clustered together while smaller fluctuations are followed by smaller movement. That is larger returns are followed by periods of larger returns, while smaller returns are followed by periods of smaller returns. This is consistent with Fama's (1965) observation that stock returns exhibit volatility clustering, where large returns tend to be followed by large returns and small returns by small returns leading to adjoining periods of volatility and stability.

Another notable observation is that the banking sector, IT sector and oil–gas sector have shown greater fluctuation compared to FMCG and healthcare sector. This intuitively implies that impact of shock creates greater volatility in banking, IT and oil–gas sectors during the period under consideration. The horizontal axis represents the 1250 trading days over a period of 5 years while vertical axis represents percentage returns.

Stock return data are typically characterized by serial correlation in the returns indicating that successive returns are not independent. Serial correlation in the

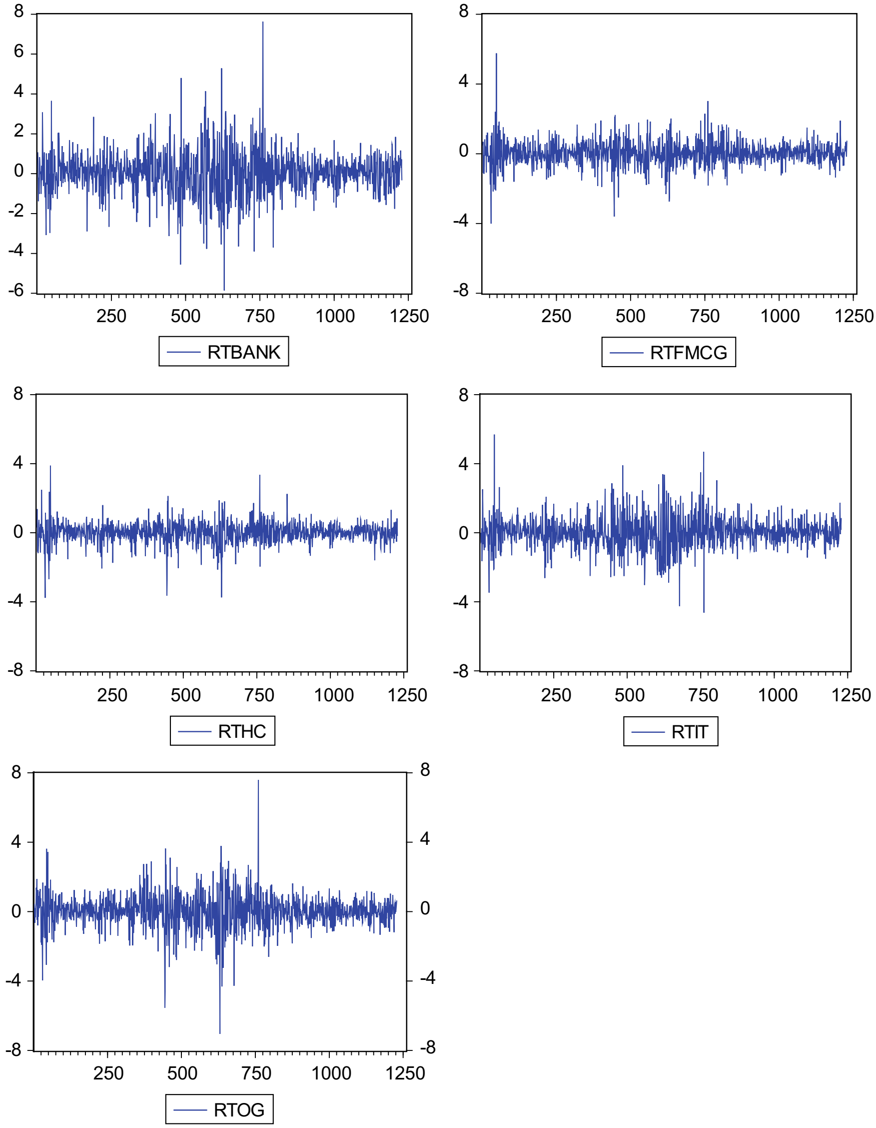


Fig. 2 Plot of daily returns of BANKEX, FMCG, health care, IT and oil–gas. *Source* Constructed from Yahoo Finance Historical Price Index

squares of returns results in distinct periods of high volatility and relative stability, i.e., volatility clustering. The existences of skewness and high kurtosis have already been established in descriptive statistics (Table 1). Still as part of the diagnostic test, we have graphically plotted the asymmetry in the distribution of returns, questioning the assumption of an underlying normal distribution. The distribution of returns also shows leptokurtosis with too many values near the mean and in the tails of the distribution, when compared with the normal distribution (Figs. 3, 4, 5, 6 and 7). We find the kurtosis value is very high and none of the indices is symmetric, which can be confirmed from the value of skewness. As already discussed high kurtosis and wide, then narrow bands in plot are hints of conditional heteroskedasticity.

4.2 GARCH Model Estimation and Volatility Clustering

Bollerslev (1986) developed the GARCH model. We use GARCH(1,1) with the following structure:

$$\text{Mean equation} \quad r_t = \mu + \varepsilon_t$$
$$\text{Variance equation} \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

r_t is the return of the asset at time t, μ is the average return, and ε_t is the residual return. The variance (h_t) is a function of an intercept (α_0), a shock from the prior period (coefficient α_1) that is the residual term ε_t and the variance from last period h_{t-1} (coefficient β_1). The size of parameters α and β determines the short-run dynamics of the volatility in time series.

The parameter estimates of the GARCH(1,1) models are described in Table 3.

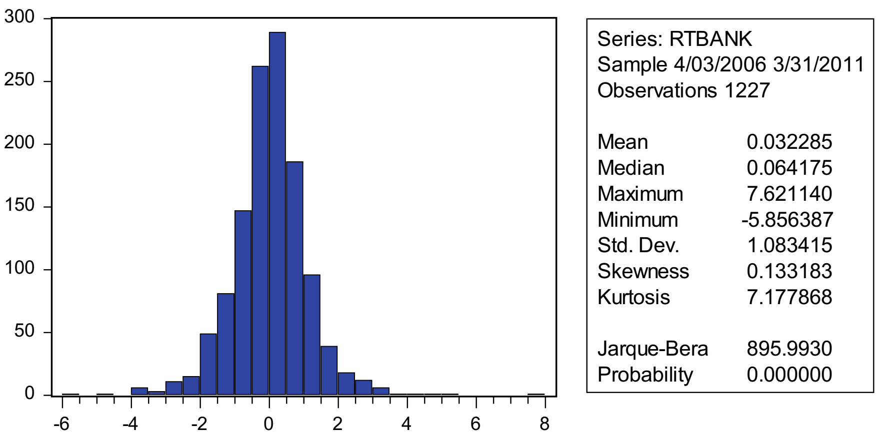


Fig. 3 Normality test plot of BANKEX

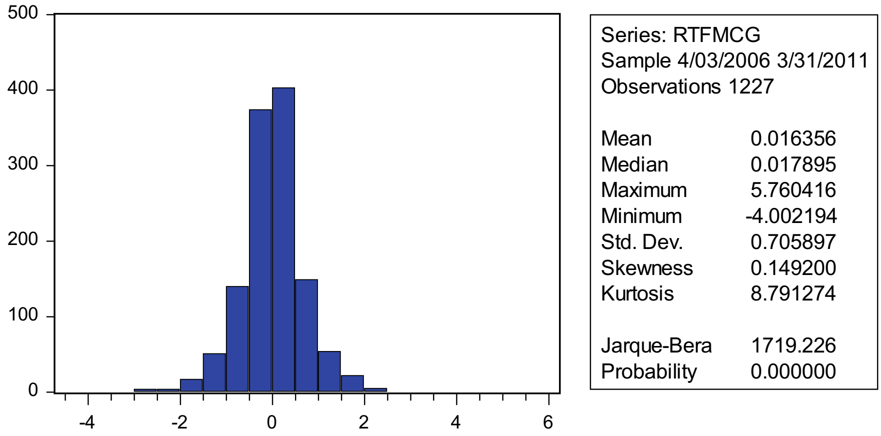


Fig. 4 Normality test plot of FMCG

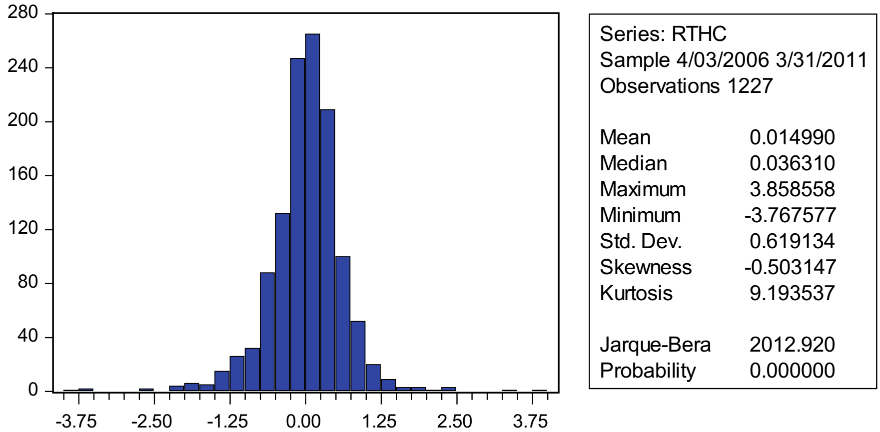


Fig. 5 Normality test plot of health care

The corresponding estimates of SENSEX are also given for reference. The values are all statistically significant. The estimates of β_1 are always markedly greater than those of α_1, and the sum $\beta_1 + \alpha_1$ is very close to but smaller than unity indicating that the stationarity conditions are satisfied. The fact that $\beta_1 + \alpha_1$ is close to unity is useful for purposes of forecasting conditional variances. The sum $\beta_1 + \alpha_1$, however, is rather close to one, which indicates a long persistence of shocks in volatility in all the sectors. Poterba and Summers (1986) have argued that for a long period asset like stocks, persistence of shocks is important in taking decision about whether to change the expectation about the price of the stock. The reason for such an argument is that if shocks to the variance are only transitory in nature, i.e., has only short-term effect, investors will not make any changes in their discounting factor while obtaining the

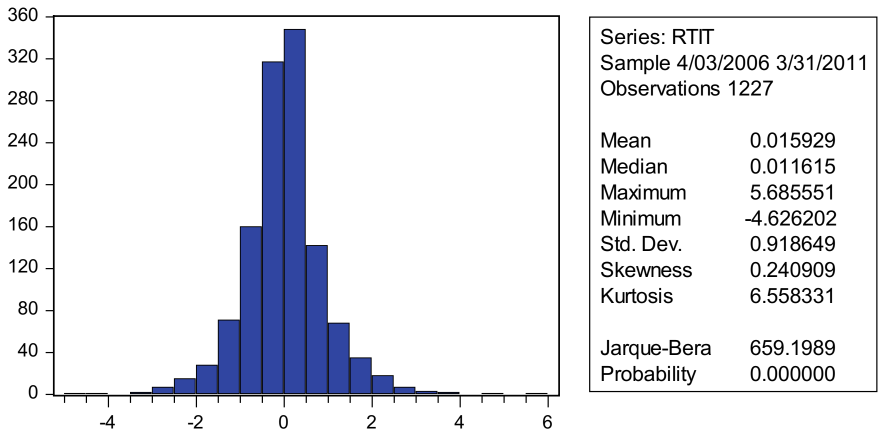


Fig. 6 Normality test plot of IT

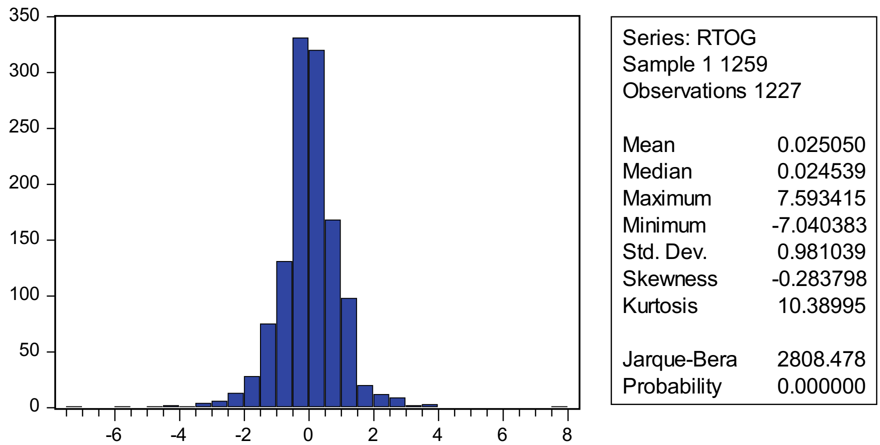


Fig. 7 Normality test plot of oil–gas

present discounted value of the stock and hence its price. But if the shock is going to stay, then discounting factor to value the stock has to be revised.

Estimates confirm that for all the sub-sectors, the shock is going to stay for a longer period, and for the banking and oil–gas sector the persistence is maximum as the value of $\alpha_1 + \beta_1$ is 0.99 for them, which is highest showing a greater vulnerability. (Detail estimation in Appendix Tables 12, 13, 14, 15 and 16.) (Table 3).

The GARCH estimations of each sub-sector as well as the SENSEX for each sub-period are presented in Table 4, where we find a distinct characteristic in the persistence of the shock. For all the sectors considered, we find third sub-period when the economy was recovering from the shock has the highest persistence of

Table 3 GARCH(1,1) estimates: domestic indices, entire period

	RTBANK	RTFMCG	RTHC	RTIT	RTOG	SENSEX
A	0.080541 (0.0006)*	0.039517 (0.0131)	0.036563 (0.0158)	0.061524 (0.0014)**	**0.032097 (0.1226)**	0.054471 (0.0011)
α_0	0.011087 (0.0003)*	0.021126 (0.0000)*	0.015791 (0.0000)*	0.024121 (0.0000)*	0.008908 (0.0000)*	0.009830 (0.0000)*
α_1	0.086678 (0.0000)*	0.162210 (0.0000)*	0.125321 (0.0000)*	0.147438 (0.0000)*	0.107340 (0.0000)*	0.138379 (0.0000)*
β_1	0.906132 (0.0000)*	0.799753 (0.0000)*	0.836183 (0.0000)*	0.829016 (0.0000)*	0.890772 (0.0000)*	0.855913 (0.0000)*
$\alpha_1 + \beta_1$	0.99281	0.96196	0.9615	0.97645	0.998	0.994292

*Significant at 1%. **Significant at 5%

shock. The values of $\alpha_1 + \beta_1$ are highest and close to 1 (more than 0.98) for all the sectors in the recovery phase of the economy. The same result holds for SENSEX as well. Result of SENSEX shows that persistence is estimated to be maximum in post-crisis period. For RTBANK, it is 0.987069 which is greater than the previous two sub-period's values of 0.877149 and 0.944416. This proves that persistence of shock has increased in post-crisis recovery period. Clearly that indicates that the investors have become more cautious in dealing with the capital market, so once a shock is introduced they refrain themselves from taking further investment decision and must be including the volatility in determining the discounting factor to determine expected price of shock. This observation is true for all the five sectors as well as the SENSEX.

Other than the banking sector, all other sectors and SENSEX have the minimum persistence effect during the second sub-period, which was the down swing of the economy or the period of recession. In case of the banking sector, value of $\alpha_1 + \beta_1$ has increased from 0.877149 in the pre-crisis period to 0.944416 in the crisis period implying shock persistence increased in crisis. For other sectors, estimates of $\alpha_1 + \beta_1$ are minimum in the crisis period. The distinct characteristics shown by the sub-sectors in Indian context indicate that shock persistence reduced from the pre-crisis to crisis and again increased significantly in post-crisis recovery period. In case of healthcare sector, the stability condition may be satisfied but the value of $\alpha_1 + \beta_1 = \mathbf{0.644809}$ is much less than 1 implying that during the recessionary period the investors of that sector considered that the shock in healthcare sector is only transitory in nature.

Table 4 GARCH(1,1) estimates to justify volatility clustering of domestic indices in each three sub-periods

		A	α_0	α_1	β_1	$\alpha_1 + \beta_1$
SENSEX	4/04/2006 to 12/14/2007	0.098926 (0.0007)**	0.026160 (0.0002)**	0.180966 (0.0000)*	0.775294 (0.0000)*	0.956260
	12/17/2007 to 4/01/2009	−0.058058 (0.3254)	0.0751856 (0.0571)	0.131614 (0.0072)	0.820384 (0.0000)*	0.951998
	4/02/2009 to 3/31/2011	0.050780 (0.0229)	0.005644 (0.0321)	0.118561 (0.0000)*	0.875519 (0.0000)*	0.994080
RTBANK	4/04/2006 to 12/14/2007	−0.134035 (0.0009)**	0.102949 (0.0100)	0.139307 (0.0043)**	0.737842 (0.0000)*	0.877149
	12/17/2007 to 4/01/2009	0.126447 (0.1166)	0.141315 (0.1787)	0.104988 (0.0088)	0.839428 (0.0000)*	0.944416
	4/02/2009 to 3/31/2011	−0.084308 (0.0062)	0.007498 (0.0153)	0.076840 (0.0000)*	0.910229 (0.0000)*	0.987069
RTFMCG	4/04/2006 to 12/14/2007	−0.044694 (0.1101)	0.028361 (0.0024)**	0.210242 (0.0000)*	0.748879 ((0.0000)*	0.959121
	12/17/2007 to 4/01/2009	−0.008058 (0.8326)	0.077452 (0.0140)	0.344120 (0.0009)**	0.570086 (0.0000)*	0.914206
	4/02/2009 to 3/31/2011	−0.045753 (0.0474)	0.004003 (0.0083)	0.035854 (0.0024)	0.947011 (0.0000)	0.982865
RTHC	4/04/2006 to 12/14/2007	−0.041574 (0.0708)	0.031999 (0.0036)**	0.270634 (0.0000)*	0.674559 (0.0000)*	0.945193
	12/17/2007 to 4/01/2009	−0.007192 (0.0832)	0.194891 (0.0008)**	0.307023 (0.0000)*	0.337786 (0.0157)	**0.644809**
	4/02/2009 to 3/31/2011	−0.052107 (0.0096)	0.001845 (0.0004)**	0.013576 (0.0250)	0.974159 (0.0000)*	0.987735
RTIT	4/04/2006 to 12/14/2007	−0.057681 (0.0683)	0.065656 (0.0003)**	0.260893 (0.0000)*	0.659540 (0.0000)*	0.920433
	12/17/2007 to 4/01/2009	0.032191 (0.6105)	0.140804 (0.1710)	0.112961 (0.0437)	0.799952 (0.0000)*	0.912913

(continued)

Table 4 (continued)

		A	α_0	α_1	β_1	$\alpha_1 + \beta_1$
	4/02/2009 to 3/31/2011	−0.057589 (0.0334)	0.004069 (0.0154)	0.013960 (0.0070)	0.972453 (0.0000)*	0.986413
RTOG	4/04/2006 to 12/14/2007	−0.092805 (0.0055)	0.024668 (0.0093)	0.126577 (0.0000)	0.840146 (0.0000)*	0.966723
	12/17/2007 to 4/01/2009	0.027242 (0.6972)	0.185556 (0.0321)	0.212809 (0.0000)*	0.700875 (0.0000)*	0.913684
	4/02/2009 to 3/31/2011	−0.010908 (0.6734)	0.003215 (0.0000)*	−0.021312 (0.0000)*	1.008629 (0.0000)*	0.987317

*Significant at 1%, **Significant at 5%

4.3 Half-Life Estimation for the Entire Horizon and Sub-Periods

Next, we calculate the half-life developed by Lamoureux and Lastrapes (1990). Half-life period is that period in which the shock diminishes to half of its original size. The half-life for GARCH(1,1) process is $1 - [\log e2/\log e(\alpha_1 + \beta_1)]$.

The calculations of half life for sectoral indices for the entire period is reported in Table 5. From Table 5, we see that the effect of a shock to the volatility process of daily return takes about maximum 368 days to diminish by half its original impact in the case of OILGAS and minimum 19 days for health care. Apparently, oil and gas sector is found to have maximum vulnerability in face of a shock followed by banking sector. This corroborates with earlier estimates of descriptive statistics which show banking and oil–gas sector experience the biggest range of variability in expected returns. Also in case of oil–gas, estimates show that distribution has positive mean with negative skewness proving a possibility of few small gains with some extreme loses. The chance of extreme loses is also indicated by the high kurtosis. It is above ten for oil–gas. The reason may be these industries have maximum vertical integration with other industries in the economy. It appears that any bad or good news does have a significant and long-lasting impact on the volatility of the stock prices.

Table 6 shows the results of half-life calculation for the SENSEX in each sub

Table 5 Half-life of shock calculation for the domestic indices

	BANKEX	FMCF	HC	IT	OILGAS
α_1	0.08668	0.16221	0.12532	0.14744	0.10734
β_1	0.90613	0.79975	0.83618	0.82902	0.89077
Half-life	97.05735	18.87416	18.65685	30.09005	367.7864

Table 6 Half-life of shock calculation for three sub-periods of SENSEX

	Period-1: March 4, 2006, to December 14, 2007	Period-2: 17th 2007 to April 1, 2009	Period-3: April 2 to March 31, 2011
α_1	0.180966	0.131614	0.118561
β_1	0.775294	0.820384	0.875519
Half-life	16.50	15.09	117.74

period, and Table 7 gives the same results for sectoral indices. The results at disaggregated level show that as sub-sector banking and oil–gas sector have maximum similarities with composite index of SENSEX. For banking and oil–gas sector, the persistence of shock in post-crisis period is showing a similar pattern to that of SENSEX. That indicates these two sectors have maximum integration with the rest of the world during the period of study and shows similar behavior.

When we calculate the half-life of shocks for three sub-periods separately, we observe the third sub-period or the post-recession recovery period has the longest half-life period of the shock for all the sectors taken into consideration. The length of the third post-crisis period is about 490 trading days. The half-life calculation shows during that period the half-life of all the sectors increased significantly and ranged between 40 and 60 days. During the recession, the half-life of shock is at its lowest. This may be the outcome of our earlier GARCH finding that volatility persistence is highest for the post-recession recovery period showing that given the experience of the recession, bad or good news does have a significant and long-lasting impact

Table 7 Half-life of shock calculation of the domestic indices of each three sub-periods

	Period	α_1	β_1	Half-life
BANKEX	4/04/2006 to 12/14/2007	0.139307	0.737842	6.288035
	12/17/2007 to 4/01/2009	0.104988	0.839428	13.12039
	4/02/2009 to 3/31/2011	0.07684	0.910229	54.2562
FMCG	4/04/2006 to 12/14/2007	0.210242	0.748879	17.60709
	12/17/2007 to 4/01/2009	0.34412	0.570086	8.727449
	4/02/2009 to 3/31/2011	0.035854	0.947011	41.10455
HC	4/04/2006 to 12/14/2007	0.270634	0.674559	13.29723
	12/17/2007 to 4/01/2009	0.307023	0.337786	2.579639
	4/02/2009 to 3/31/2011	0.013576	0.974159	57.16696
IT	4/04/2006 to 12/14/2007	0.260893	0.65954	9.360129
	12/17/2007 to 4/01/2009	0.112961	0.799952	8.607414
	4/02/2009 to 3/31/2011	0.01396	0.972453	51.66811
OILGAS	4/04/2006 to 12/14/2007	0.126577	0.840146	21.48109
	12/17/2007 to 4/01/2009	0.212809	0.700875	8.678557
	4/02/2009 to 3/31/2011	−0.02131	1.008629	55.30436

on the volatility of the stock prices. The significant shift of shock persistence is recorded for healthcare sector. GARCH estimates in Sect. 2. Table 3 shows that $\alpha_1 + \beta_1$ is 0.644809 for RTHC in crisis period, which changes significantly to 0.987735 in post-shock recovery period. The inevitable outcome is reflected here in terms of half-life of shock. In the healthcare sector, half-life increases from 2.5 days in the crisis period to 57 days in the post-crisis period.

The result shows that before or during crisis, healthcare sector has the lowest persistence of shock compared to other sectors. As per Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers, India's pharmaceutical industry had a total turnover of US$21.04 billion between 2008 and September 2009, whereas the domestic market was worth US$12.26 billion.[3] The Indian pharmaceutical sector was not affected adversely due to the dependence on the domestic market and generic drugs. India is one of the largest producers of generics in the world, as majority of people prefer cheap medications over branded ones. The cost of production is less and manpower is abundant due to which the MNCs have good advantage to increase their revenue and profit by setting up industries in India. But after the crisis period, like all other sectors, health care also had become cautious and all the listed pharmaceuticals company's shares suffer from the same kind of uncertainty as can be confirmed by the sudden increase in shock persistence and half-life of that sector.

5 Leverage Effect and Volatility Spillover from the USA

This section investigates the existence of leverage effect to find out whether negative news causes more volatility than positive news for these five domestic sectors and the spillover effect at the sectoral level from the US stock market. The purpose of this estimation is to check the level of penetration of a matured market like the USA in domestic sectors of an emerging economy during a crucial period, which covers the period of subprime crisis. Section 5.1 discusses the leverage effect, and Sect. 5.2 discusses the spillover effect.

5.1 EGARCH Estimates: Leverage Effect

EGARCH model is estimated to investigate the existence of leverage effect for both the entire time horizon and for the sub-periods. The purpose is to check whether the pattern of impact of bad news changes overtime across the pre-crisis, crisis and post-crisis period.

EGARCH model specifies the conditional variance in the following way:

$$\log \sigma_t^2 = \omega + \beta \log\left(\sigma_{t-1}^2\right) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

[3]Bhatt and Panigrahi (2014).

where ω, β, γ, α are the parameters to be estimated. The ω parameter represents a magnitude effect or the symmetric effect of the model—the 'GARCH' effect. β measures the persistence in conditional volatility irrespective of anything happening in the market. When β is relatively large, then volatility takes a long time to die out following a crisis in the market. The γ parameter measures the asymmetry or the leverage effect. If $\gamma = 0$, then the model is symmetric. When $\gamma < 0$, then positive shocks (good news) generate less volatility than negative shocks (bad news). When $\gamma > 0$, it implies that positive innovations are more destabilizing than negative innovations.

Results of the EGARCH model estimation are reported in Tables 8 and 9. We only report the results of the parameters, which are of interest to us.

We find that all the coefficients are significant for all the indices. The asymmetry coefficient γ that is leverage term is negative in all cases indicating the existence of the leverage effect for the stock market returns during the sample period. Negative leverage effect implies that the variance goes up more after negative residuals or returns than after positive returns. That means good news generates less volatility than the bad news in Indian sub-markets. The persistence parameter β is very large, implying that variance moves slowly through time.

When we considered the leverage effect of the three sub-periods separately (Table 9), we find that leverage effect is positive and significant for all the sectors for the first two periods. In the third post-recession recovery period, it is not significant except in the oil and gas sector. That implies that good or bad news is having less impact during the recovery. The only explanation maybe investors have become more risk averse and more observant to bad news from their past experience which made them less vulnerable. These results relate to our earlier result of half-life calculation. Greater the half-life, more is the persistence of shock. We found in the disaggregated level, half-life of the shocks increases in the post-crisis period. That in turn establishes the fact that leverage effects are insignificant in the post-crisis period because given the persistence of shock, temporary good or bad news has insignificant impact on stock prices. The persistence parameter is highest in the third sub-period for all sectors—nearly 1. This implies that volatility was most persistent in the third sub-period.

Only in the oil and gas sector still leverage effect exists in the third period, given the fact that India being the principal oil importing country. Even our earlier estimates

Table 8 Leverage effect of domestic indices

	RTBANKEX	RTFMCG	RTHC	RTIT	RTOG
γ	−0.074834 (0.0000)*	−0.114238 (0.0000)*	−0.066558 (0.0000)*	−0.065836 (0.0000)*	−0.041001 (0.0000)*
β	0.981670 (0.0000)*	0.934618 (0.0000)*	0.955219 (0.0000)*	0.964341 (0.0000)*	0.978392 (0.0000)*

The detailed estimates can be obtained from the authors on request
*Significant at 1%

Table 9 Leverage effect of domestic indices for each three sub-periods

	Leverage effect for different sub-periods					
		RTBANKEX	RTFMCG	RTHC	RTIT	RTOG
γ	First subsample 4/04/2006 14/12/2007	0.203456 (0.0000)*	0.171486 (0.0000)*	0.159267 (0.0001)**	0.176982 (0.0000)*	0.035240 (0.0269)
	Second subsample 17/12/2007 1/04/2009	0.076834 (0.0074)	0.184885 (0.0019)**	0.198216 (0.0000)*	0.137352 (0.0016)**	0.130056 (0.0002)**
	Third subsample 2/04/2009 31/3/2011	**0.023089 (0.2989)**	**0.007835 (0.5922)**	**0.009102 (0.4571)**	**−0.005985 (0.6600)**	0.048745 (0.0181)
β	First subsample 4/04/2006 14/12/2007	0.898440 (0.0000)*	0.933730 (0.0000)*	0.861398 (0.0000)*	0.862702 (0.0000)*	0.948145 (0.0000)*
	Second subsample 17/12/2007 1/04/2009	0.950496 (0.0000)*	0.857400 (0.0000)*	0.892589 (0.0000)*	0.947349 (0.0000)*	0.930345 (0.0000)*
	Third subsample 2/04/2009 31/3/2011	0.990382 (0.0000)*	0.990187 (0.0000)*	0.991481 (0.0000)*	0.992378 (0.0000)*	0.993957 (0.0000)*

*Significant at 1%, **Significant at 5%

show highest shock persistence and highest half-life, and oil and gas sector still got affected by good and bad news even in the post-crisis period.

5.2 EGARCH (Asymmetric—GARCH) Estimates: Spillover Effect

In 1993, when foreign institutional investors (FII) were allowed to invest in the Indian equity market, the fate of the domestic investors started becoming increasingly integrated with the ups and downs in the other stock markets of the world. Over time, foreign portfolio investments have increased. The trend specially gathered momentum during the Information and Communications Technology (**ICT**) boom of 1998–2001. The simultaneous listing of a number of Indian companies with large market capitalization in the ICT sector on the Indian stock markets and NASDAQ was then expected to generate sympathetic movements or return and volatility ‘spillover’ across the two markets. As found by Galbraith and Hale (2004), 'From January 1994 to February 2000, the NASDAQ composite index rose from 776.80 to 4696.69, a 605% increase, heavily influenced by prices of high-technology stocks.' These stocks also had significant weight in both the domestic indices, viz., SENSEX and NIFTY.

In this section, we apply asymmetric GARCH model to study the relationship between the USA and the domestic segments of markets of India represented by BANKEX, FMCG, IT, health care and OILGAS. We use the Dow Jones (DJ) index as the proxy for the US market. We compute the daily return series on Dow Jones in the same way as we have done before. This series is then introduced as explanatory variables first in the mean equation and then in the variance equation. Thus, the EGARCH(1,1) models to be estimated for the different indices have now become:

Mean equation:

$$r_t = \mu + \varepsilon_t + \theta(r_{t-1}dj)$$

Variance equation:

$$\log \sigma_t^2 = \omega + \beta \log\left(\sigma_{t-1}^2\right) + \gamma \frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] + \psi(r_{t-1}dj)$$

Here $r_{t-1}dj$ is the lagged return calculated using Dow Jones index. The interpretations of the other variables remain the same. This approach is followed to verify (a) the presence of significant correlation between the two markets and (b) whether the correlations between the US stock market and sectoral indices were any different for different sectors due to recessionary shock. Again, we report the estimated values of the coefficients we are interested in.

From the results, we can draw the following inferences: With respect to returns, Dow Jones exhibits no significant positive correlation with any of the selected sectors, throughout the periods. The picture is not different for volatility either. No positive correlation is seen between the US index and any of the other indices taken into consideration. Oil and gas sector does receive some spillover effect from Dow Jones but only at 10% level of significance (Table 10).

Next, we checked the spillover for three sub-periods using the same EGARCH model, again with a lag return of Dow Jones as regressor in variance equation. Like the previous analysis, we found no spillover in mean return of any of the sectors. Coefficient of return of Dow Jones across the sub-periods was found to be not significant in the mean equation as can be seen from Table 11.

But when we break the period in relatively smaller sub-periods, we find spillover in volatility (Table 11). In the first pre-crisis period, spillover in volatility is recorded for healthcare and oil and gas sectors. During the crisis period, mean return of Dow Jones was found to contribute to the volatility of banking sector index as well as healthcare sector and oil–gas. More specifically in the third sub-period, we found spillover

Table 10 EGARCH estimates to volatility spillover in domestic sub-sectors

	RTBANKEX	RTFMCG	RTHC	RTIT	RTOG
	Mean equation				
θ (coefficient of $r_{t-1}dj$)	−0.049714 (0.2108)	−0.012651 (0.5956)	−0.018621 (0.4303)	−0.042085 (0.2193)	0.004732 (0.8925)
	Variance equation				
Ψ (coefficient of $r_{t-1}dj$)	−0.007916 (0.7518)	0.028479 (0.3882)	0.005028 (0.8555)	0.043266 (0.1798)	**0.042990 (0.0846)**
γ	−0.114591 (0.0000)*	−0.114591 (0.0000)*	−0.066550 (0.0000)*	−0.070434 (0.0000)*	−0.040187 (0.0000)*
β	0.937186 (0.0000)*	0.937186 (0.0000)*	0.955449 (0.0000)*	0.966197 (0.0000)*	0.981652 (0.0000)*

Details results are available with the authors and can be provided on request
*Significant at 1%

in all the sectors, implying investors have become more observant and thus more integrated with the world market. Return in these domestic sectors is not influenced by the return of Dow Jones as can be further confirmed by the fact that leverage effect is not significant in the third sub-period for most of the sectors. That implies that good or bad news is having less impact during the recovery. But volatility is found to have a significant coefficient that is investors do take into account the variation in the return of Dow Jones in determining the variation of stock index depending on the signal from Dow Jones. Or more precisely investors believe that the variation in Dow Jones will contribute to the volatility but will not determine the price of the stock. Oil and gas sector is the only sector found to be continuously integrated with matured market USA as spillover in volatility is recorded to be significant in all the three sub-periods.

6 Conclusion

For the sectoral indices, GARCH estimates confirm volatility clustering for all the indices and the sum of GARCH coefficient, $\beta_1 + \alpha_1$, is close to one, which indicates a long persistence of shocks in volatility. The half-life calculation showed that the effect of a shock to the volatility process of daily return takes about maximum 368 days to diminish by half its original impact in case of oil and gas and minimum 19 days for health care. For all the other sectors considered, we find third sub-period when the economy was recovering from the shock has the highest persistence of shock.

When we calculate the half-life of shocks for three sub-periods differently, we observe the third sub-period or the post-recession recovery period has the highest half-life period of the shock for all the sectors taken into consideration. EGARCH model estimation shows leverage term is positive and statistically different from zero

Table 11 Volatility spillover into domestic sectors in each three sub-periods

Spillover effect from an external market to different sectors at different subsamples						
		RTBANKEX	RTFMCG	RTHC	RTIT	RTOG
Mean equation						
θ	First subsample 4/04/2006 to 14/12/2007	0.188347 (0.0628)	0.024245 (0.7244)	0.072283 (0.2485)	0.109950 (0.1797)	0.063760 (0.4148)
	Second subsample 17/12/2007 to 1/04/2009	−0.001525 (0.9877)	−0.000560 (0.9891)	0.009333 (0.8290)	−0.113803 (0.1432)	0.092326 (0.2411)
	Third subsample 2/04/2009 to 31/3/2011	0.063133 (0.2566)	−0.016794 (0.7421)	0.029588 (0.4453)	0.078951 (0.1215)	0.024075 (0.5739)
Variance equation						
ψ	First subsample 4/04/2006 to 14/12/2007	0.129505 (0.3038)	0.196842 (0.1942)	**0.300217 (0.0120)**	0.086535 (0.4762)	**0.369670 (0.0005)****
	Second subsample 17/12/2007 to 1/04/2009	**−0.188166 (0.0027)****	0.003555 (0.9666)	**−0.203961 (0.0032)****	−0.063624 (0.3428)	**−0.166054 (0.0154)**
	Third subsample 2/04/2009 to 31/3/2011	**0.322419 (0.0000)***	**0.115596 (0.0283)**	**0.208682 (0.0000)**	**0.157600 (0.0000)**	**0.114776 (0.0000)**
γ	First subsample 4/04/2006 to 4/12/2007	0.192956 (0.0000)*	0.178863 (0.0000)*	0.171670 (0.0001)*	0.179278 (0.0000)*	0.041509 (0.1493)
	Second subsample 17/12/2007 to 1/04/2009	0.084177 (0.0052)	0.184220 (0.0067)	0.212382 (0.0000)*	0.136169 (0.0028)**	0.158524 (0.0000)*
	Third subsample 2/04/2009 to 31/3/2011	0.046867 (0.0118)	−0.001253 (0.9412)	0.027461 (0.0047)	−0.033165 (0.1119)	0.031605 (0.0641)
β	First subsample 4/04/2006 to 14/12/2007	0.910971 (0.0000)*	0.943008 (0.0000)*	0.863165 (0.0000)*	0.871183 (0.0000)*	0.964178 (0.0000)*

(continued)

Table 11 (continued)

Spillover effect from an external market to different sectors at different subsamples						
	Second subsample 17/12/2007 to 1/04/2009	0.926611 (0.0000)*	0.838730 (0.0000)*	0.858189 (0.0000)*	0.939767 (0.0000)*	0.912791 (0.0000)*
	Third subsample 2/04/2009 to 31/3/2011	0.976514 (0.0000)*	0.980019 (0.0000)*	0.980460 (0.0000)*	0.972887 (0.0000)*	0.989135 (0.0000)*

*Significant at 1%, **Significant at 5%

indicating the existence of the leverage effect for the stock market returns during the sample period for all the indices. When we considered the leverage effect of the three sub-periods separately, we find that leverage effect is positive and significant for all the sectors for the first two periods. In the third post-recession recovery period, it is not significant. That implies that good or bad news is having less impact during the recovery.

We found no significant spillover is taking place from the USA directly to these domestic sub-sectors of India either in mean or in variance, with an exception of oil and gas sector, where some spillover may be taking place as the coefficient is found to be significant at 10% level. However, when we break the period in relatively smaller sub-period, we find spillover in volatility. More specifically in the third sub-period, we found spillover in all the sectors, implying investors have become more observant thus less vulnerable to bad news which can be further confirmed by the fact that leverage effect is not significant in the third sub-period.

Taking the entire analysis into account, we can conclude that the evidence shows that there has been capital market integration at the sectoral level. Investors' reactions to subprime crisis were manifested at the sectoral level as well as at the level of SENSEX. However, the extent of integration was not the same across sectors. Generally speaking, sectors like banking and oil and gas turned out to be more susceptible than other sectors.

Appendix

See Tables 12, 13, 14, 15 and 16.

Table 12 GARCH(1,1) for entire period of RTBANK

Dependent variable: RTBANK				
Method: ML—ARCH (Marquardt)—Normal distribution				
Sample (adjusted): 4/04/2006–3/31/2011				
Included observations: 1227 after adjustments				
Convergence achieved after 15 iterations				
Variance backcast: ON				
GARCH = $C(2) + C(3)$ * RESID$(-1)^2 + C(4)$ * GARCH(-1)				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.080541	0.023515	3.425018	0.0006
	Variance equation			
C	0.011087	0.003035	3.652936	0.0003
RESID$(-1)^2$	0.086678	0.011522	7.522818	0.0000
GARCH(-1)	0.906132	0.011201	80.89979	0.0000
R-squared	−0.001985	Mean dependent var		0.032285
Adjusted R-squared	−0.004443	S.D. dependent var		1.083415
S.E. of regression	1.085820	Akaike info criterion		2.734445
Sum squared resid	1441.922	Schwarz criterion		2.751111
Log likelihood	−1673.582	Durbin–Watson stat		1.755322

Table 13 GARCH estimate for entire period of RTFMCG

Dependent variable: RTFMCG				
Method: ML—ARCH (Marquardt)—Normal distribution				
Date: 02/13/13 Time: 13:12				
Sample (adjusted): 4/04/2006–3/31/2011				
Included observations: 1227 after adjustments				
Convergence achieved after 14 iterations				
Variance backcast: ON				
GARCH = $C(2) + C(3)$ * RESID$(-1)^2 + C(4)$ * GARCH(-1)				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.039517	0.015928	2.480870	0.0131
	Variance equation			
C	0.021126	0.004074	5.185453	0.0000
RESID$(-1)^2$	0.162210	0.018479	8.778315	0.0000
GARCH(-1)	0.799753	0.019711	40.57324	0.0000
R-squared	−0.001077	Mean dependent var		0.016356
Adjusted R-squared	−0.003533	S.D. dependent var		0.705897
S.E. of regression	0.707143	Akaike info criterion		1.900939
Sum squared resid	611.5622	Schwarz criterion		1.917605
Log likelihood	−1162.226	Durbin–Watson stat		2.003647

Table 14 GARCH estimate for entire period of RTHC

Dependent variable: RTHC				
Method: ML—ARCH (Marquardt)—Normal distribution				
Date: 02/13/13 Time: 13:15				
Sample (adjusted): 4/04/2006–3/31/2011				
Included observations: 1227 after adjustments				
Convergence achieved after 20 iterations				
Variance backcast: ON				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.036563	0.015146	2.414050	0.0158
	Variance equation			
C	0.015791	0.003234	4.882409	0.0000
RESID$(-1)^2$	0.125321	0.014379	8.715802	0.0000
GARCH(-1)	0.836183	0.018874	44.30337	0.0000
R-squared	−0.001215	Mean dependent var		0.014990
Adjusted R-squared	−0.003671	S.D. dependent var		0.619134
S.E. of regression	0.620270	Akaike info criterion		1.664304
Sum squared resid	470.5305	Schwarz criterion		1.680970
Log likelihood	−1017.051	Durbin–Watson stat		1.893584

Table 15 GARCH estimate for entire period of RTIT

Dependent variable: RTIT				
Method: ML—ARCH (Marquardt)—Normal distribution				
Date: 02/13/13 Time: 13:17				
Sample (adjusted): 4/04/2006–3/31/2011				
Included observations: 1227 after adjustments				
Convergence achieved after 14 iterations				
Variance backcast: ON				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.061524	0.019200	3.204418	0.0014
	Variance equation			
C	0.024121	0.005281	4.567261	0.0000
RESID$(-1)^2$	0.147438	0.017831	8.268720	0.0000
GARCH(-1)	0.829016	0.019171	43.24270	0.0000
R-squared	−0.002465	Mean dependent var		0.015929
Adjusted R-squared	−0.004924	S.D. dependent var		0.918649
S.E. of regression	0.920908	Akaike info criterion		2.414996
Sum squared resid	1037.191	Schwarz criterion		2.431662
Log likelihood	−1477.600	Durbin–Watson stat		2.012576

Table 16 GARCH estimate for entire period of RTOG

Dependent variable: RTOG				
Method: ML—ARCH (Marquardt)—Normal distribution				
Date: 02/13/13 Time: 13:17				
Sample (adjusted): 4/04/2006–3/31/2011				
Included observations: 1227 after adjustments				
Convergence achieved after 18 iterations				
Variance backcast: ON				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.032097	0.020789	1.543953	0.1226
	Variance equation			
C	0.008908	0.002647	3.365362	0.0008
RESID(−1)^2	0.107340	0.010723	10.01025	0.0000
GARCH(−1)	0.890772	0.009643	92.37028	0.0000
R-squared	−0.000052	Mean dependent var		0.025050
Adjusted *R*-squared	−0.002505	S.D. dependent var		0.981039
S.E. of regression	0.982267	Akaike info criterion		2.470056
Sum squared resid	1180.010	Schwarz criterion		2.486722
Log likelihood	−1511.380	Durbin–Watson stat		1.876110

References

Agarwal, R. N. (2000). *Capital market development, corporate financing pattern and economic growth in India*. Institute of Economic Growth Discussion Paper No. 20.

Arshanapalli, B., & Doukas, J. (1993). International stock market linkages: Evidence from the pre- and post-October 1987 period. *Journal of Banking and Finance, 17,* 193–208.

Ballie, R. T., & DeGennaro, R. P. (1990). Stock returns and volatility. *Journal of Financial and Quantitative Analysis, 25*(2), 203–214.

Bhatt, S., & Panigrahi, A. K. (2014, November). Impact of recession on Indian pharma sector. *International Journal of Research in Commerce, Economics and Management, 4*(11).

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.

Chaudhry, T. (1997). Stochastic trends in stock prices: Evidence from Latin American Markets. *Journal of Macroeconomics, 19,* 285–304.

Chen, J., Hong, H. G., & Stein, J. C. (2002, November). Breadth of ownership and stock returns. *Journal of Financial Economics, 66*(2–3), 171–205.

Demirer, R., & Lien, D. (2005). Correlation and return dispersion dynamics in Chinese markets. *International Review of Financial Analysis, 14*(4), 477–491.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74,* 427–431.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49,* 1057–1072.

Enders, W. (2004). *Applied economic time series* (2nd ed.). Chichester: Wiley.

Engle, R., & White, H. (1999). *Cointegration, causality and forecasting*. Oxford: Oxford University Press.

Galbraith, J. K., & Hale, T. (2004). *Income distribution and the information technology bubble*. http://utip.lbj.utexas.edu/papers/utip_27.pdf.

Huang, B., Yang, C., & Hu, J. W. (2000). Causality and cointegration of stock markets among the United States, Japan, and the South China Growth Triangle. *International Review of Financial Analysis, 9,* 281–297.

Lafuente, J. Á., & Ruiz, J. (2004). The New Market effect on return and volatility of Spanish stock indexes. *Applied Financial Economics, 14*(18), 1343–1350.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Heterskedasticity in stock return data: Volume versus GARCH effect. *Journal of Finance, 45*(1), 221–229.

Ljung, G., & Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika, 66,* 67–72.

Malik, F., & Ewing, B. T. (2009). Volatility transmission between oil prices and equity sector returns. *International Review of Financial Analysis, 18*(3), 95–100.

Mazouz, K., & Saadouni, B. (2007). New evidence on the price and liquidity effects of the FTSE 100 index revisions. *International Review of Financial Analysis, 16*(3), 223–241.

McMillan, D. G., & Kambouroudis, D. (2009). Are RiskMetrics forecasts good enough? Evidence from 31 stock markets. *International Review of Financial Analysis, 18*(3), 117–124.

Meric, I., Ratner, M., & Meric, G. (2008). Co-movements of sector index returns in the world's major stock markets in bull and bear markets: Portfolio diversification implications. *International Review of Financial Analysis, 17*(1), 156–177.

Patra, T., & Poshakwale, S. S. (2008). Long-run and short-run relationship between the main stock indexes: Evidence from the Athens stock exchange. *Applied Financial Economics, 18*(17), 1401–1410.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75,* 335–346.

Poterba, J. M., & Summers, L. H. (1986, December). The persistence of … stock market fluctuations. *American Economic Review*, 1142–1152.

Wang, Z., Kutan, A. M., & Yang, J. (2005). Information flows within and across sectors in Chinese stock markets. *Quarterly Review of Economics and Finance, 45*(4–5), 767–780.

You, L., & Daigler, K. (2010). Is international diversification really beneficial? *Journal of Banking & Finance, 34,* 163–173.

Institutions and Development

Uncertain R&D Outcomes and Cooperation in R&D

Abhishek Kabiraj and Tarun Kabiraj

Abstract The present paper provides a brief survey of some of the papers dealing with R&D uncertainty. This helps us identify which factors are more favorable for cooperative R&D and which factors are not. The paper provides the analysis under a unified framework. We take the classic paper by Marjit (Econ Lett 37:187–191, 1991) as the benchmark case and then proceed to examine whether, or to what extent, Marjit result will undergo a change with respect to different assumptions related to R&D investment.

1 Introduction

Economic growth and development of a country depend, to a large extent, on the country's capability of doing research and development (R&D) successfully. But R&D activity involves a huge expenditure in setting a research laboratory, installing scientific instruments and recruiting scientific personnel. It requires well-directed and well-coordinated efforts. Even after such investment, an R&D firm does not know a priori whether it will come out with a successful innovation. This means R&D outcome is uncertain. Often a success comes only after many failures. There is also uncertainty in commercializing and marketing the innovation. Even when success occurs, the innovator does not know whether it will be able to appropriate the required amount of profits before the innovation becomes obsolete. This is because

This is our humble contribution to this volume to honor Professor Sarmila Banerjee (Mallik). Both myself and Abhishek were her direct students, myself being one of her first batch students and Abhishek being her student during the last year of her formal teaching. She has always been a great teacher. Overtime, she has been our family friend and we have depended on her for taking every crucial decision in life. From the core of our hearts, we cherish our love and respect for her.

A. Kabiraj
HSBC Pvt. Ltd., Kolkata, India

T. Kabiraj (✉)
Economic Research Unit, Indian Statistical Institute,
203 B. T. Road, Kolkata 700108, India
e-mail: tarunkabiraj@hotmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_6

of the problem of spillovers, free riding, and imitation of the R&D outcome by the rivals who become competitive at the market place. All these problems lead to under-investment in R&D.[1] So, it is an important concern of the policy makers about how to provide sufficient incentives to the private firms for doing R&D.

Given the concern for under-investment in R&D, most of the governments extend fiscal support to the investors in the form of R&D subsidy, cheap credit facility, and tax concessions. To the problem of imitation and leaking out of knowledge, the government of a country provides patent protection to the new innovations so that the imitators cannot use or copy the innovation. However, it all depends on the effectiveness and enforcement of patent laws. Weak patent protection can hardly protect innovations from copying. Patent protection cannot fully protect the innovation; it raises the cost of imitation.[2]

Under these circumstances, cooperative R&D is suggested to be the way out of the problem of high R&D cost, uncertainty, and spillovers of knowledge. By cooperating the firms can avoid duplicating scarce resources and share research cost and output as well as uncertainty (Katz 1986). Cooperative R&D is an *ex ante* agreement among the member firms on sharing R&D expenses and results. Most of the countries nowadays promote cooperative research.[3] Most popular form of cooperative research is research joint venture or RJV. Under RJV, the firms conduct research in a single laboratory and share R&D cost and output. The firms can also write a contract to do R&D independently in their own laboratories, but share the R&D outcome of any lab.

This has led to the question of whether R&D will be cooperative or non-cooperative. This is the problem of the choice of R&D institution or organization. A substantial literature has emerged discussing this question. The pioneering work in this field is d'Aspremont and Jacquemin (1988). The paper focuses on R&D spillovers and studies whether R&D investment will be cooperative or non-cooperative. When spillovers are high, firms generally go for cooperative research. The work has been extended by Kamien et al. (1992) to the case of differentiated duopoly and by Suzumura (1992) to the case of oligopoly. Motta (1992) discussed the choice under vertical product differentiation. Amir et al. (2003) study the problem when spillovers are endogenous.

These works, however, do not consider uncertainty in R&D outcome. Marjit (1991) was pioneer to show that uncertainty alone can result in cooperative research. It was shown that if probability of success in R&D is either high or low, cooperative research should occur, otherwise non-cooperative R&D would be preferred. Then following Marjit (1991), a number of papers have been contributed to the literature examining the choice problem (see, e.g.,, Combs 1992; Choi 1992, 1993; Kabiraj 2006, 2007; Mukherjee and Marjit 2004; Kabiraj and Chattopadhyay 2015; Kabi-

[1]There are theoretical and empirical literatures showing under-investment in R&D due to spillovers of knowledge. See, for instance, Spence (1984), Jaffe (1986), Ornaghi (2006), and Bakhtiari and Breunig (2018).

[2]Griliches (1990) found to have inverse relation between the degree of patent protection and the level of spillovers.

[3]In the USA, the National Cooperative Research Act 1984 was passed, and following this, a large number of cooperative ventures had been registered (Vonortas 1997).

raj and Kabiraj 2019).[4] All these papers have assumed that the product market is non-cooperative. However, cooperation in production may give additional incentive to R&D (d'Aspremont and Jacquimin 1988; Kamient et al. 1992). In a three-firm framework, Kabiraj and Mukherjee (2000) have studied whether cooperation in production will induce the firms to cooperate in R&D, and vice versa. Silipo and Weiss (2005) have studied the choice between cooperative and non-cooperative R&D in the presence of both spillovers and uncertainty.

The purpose of the present paper is to provide a brief survey of some of those works dealing with R&D uncertainty. This will help us identify which factors are more favorable for cooperative R&D and which factors are not. We provide an analysis in the same frame. We consider Marjit (1991) as the benchmark case and then examine whether, or to what extent, Marjit result will undergo a change with respect to different assumptions related to R&D investment. In this sense, the present paper is a review of Marjit (1991). In the context of Marjit (1991) model, we consider various scenarios like: product or process innovation; drastic or non-drastic innovation; patent protection may or may not be available; imitation may or may not be possible; technology transfer may or may not be allowed; cooperative R&D be conducted in a single or more than one laboratory; possibility of more than one innovation; and incomplete information. We show that patent protection and technology transfer will increase incentives for non-cooperative research, whereas imitation and incomplete information will tilt the choice toward cooperative R&D. However, the qualitative result of Marjit in these cases will remain unaltered. On the other hand, duplicating research under cooperation will substantially change the Marjit result. We also see how the size of the innovation, whether small or large, may affect the choice of R&D organization. Readers, perhaps, will understand how tinkering one or the other assumption might generate a different result.

The plan of the paper is the following. In Sect. 2, we first provide Marjit (1991) model as the benchmark case, and then in a number of subsections, we study the effect of the change of one or the other assumption underlying the model and examine to what extent Marjit (1991) results are robust. In Sect. 3, we suggest some possible extensions for future research. Finally, Sect. 4 concludes the paper.

2 Framework of Analysis

Consider, initially, a symmetric duopoly market for a homogeneous product. Two firms, firm 1 and firm 2, simultaneously interact in R&D and production. They play a two-stage game. Assuming that non-cooperative research is always profitable, the firms in the first stage decide whether to conduct R&D cooperatively or

[4] Mukherjee and Ray (2009) discussed the problem when there is uncertainty in patent approval, but the R&D outcome is certain. Kabiraj (2018) studied the problem in a three-firm framework when cooperative research takes the form of RJV or knowledge sharing.

non-cooperatively. Then in the second stage, subject to the realization of the R&D outcome, they play a la Cournot and compete non-cooperatively in the product market.

Each research laboratory requires an investment $R > 0$ targeting a specified (product or process) innovation, but the research outcome is uncertain, that is, research may lead to a success or failure. When R is invested, the probability of success of a research laboratory is ρ, $0 < \rho < 1$. If both firms come up with the innovation, the market will be symmetric duopoly, but if only one firm comes up with the innovation, the market will be either monopoly or asymmetric duopoly depending on the size and nature of innovation. Below, we first provide briefly the Marjit (1991) model as the benchmark case and identify some assumptions underlying the model stated implicitly or explicitly. Then relaxing one or the other assumption, we examine to what extent Marjit (1991) results are sensitive to a particular assumption.

2.1 Benchmark Case: Marjit (1991) Model

Marjit (1991) assumes process innovation that reduces the unit cost of production. Further, it is assumed that the innovation is 'drastic' (or major) in the sense that the firm which alone adopts the innovation emerges as a monopolist in the product market and the other firm ceases to operate. Finally, it is assumed that under cooperative R&D firms conduct research in a single laboratory, sharing both R&D cost and R&D result, hence they form an RJV. Denoting the pre- and post-R&D symmetric duopoly profits of a firm by π and π^d, respectively and monopoly profit by π^m,[5] the expected payoffs of a firm under cooperative (C) and non-cooperative (NC) R&D will be given by,

$$E(C0) = \rho\pi^d + (1-\rho)\pi - \frac{R}{2} \tag{1}$$

$$E(\mathrm{NC}0) = \rho^2\pi^d + \rho(1-\rho)\pi^m + (1-\rho)^2\pi - R \tag{2}$$

Therefore, cooperative R&D is to be preferred to non-cooperative R&D if and only if

$$E(C0) > E(\mathrm{NC}0) \Leftrightarrow \rho(1-\rho) < \frac{R}{2\left[\pi^m - \left(\pi^d + \pi\right)\right]} \tag{3}$$

Clearly, the RHS of (3) is positive and constant, and the LHS, $\rho(1-\rho)$, is strictly inverted U-shaped with a unique maximum at $\rho = \left(\frac{1}{2}\right)$ and the value of the function is 0 both at $\rho = 0$ and $\rho = 1$. Then Marjit (1991) result can be stated in the following proposition.

[5] In the paper, as we can see, no result will change if $\pi \geq 0$.

Proposition 1 [Marjit (1991) Result] *If R is not very large,* $\exists \underline{\rho} \& \overline{\rho}, 0 < \underline{\rho} < \overline{\rho} < 1$, *such that cooperative R&D is preferred to non-cooperative R&D* $\forall \rho \epsilon \left(0, \underline{\rho}\right) \cup (\overline{\rho}, 1)$*; otherwise, non-cooperative R&D is preferred.*

Thus, Marjit (1991) result states that cooperative R&D will occur if the probability of success is either small or large. When the probability of success is in an intermediate range, non-cooperative R&D will occur.

It can be noted that although Marjit (1991) assumed (drastic) process innovation, it can easily be interpreted as product innovation. We may assume that initially production is not viable because each firm's marginal cost of production is sufficiently high (so, $\pi = 0$). But the process innovation reduces costs and makes the production viable; hence, this is equivalent to product innovation.

In general, irrespective of whether process innovation is drastic or non-drastic, let π^{SS} denote the payoff of a firm when both the firms adopt the successful innovation and π^{FF} denote the payoff when none has successful innovation. Similarly, when only one firm has the successful innovation, its payoff is π^{SF} and the other firm's payoff is π^{FS}. Then under Cournot competition, we must have:

$$\pi^{SF} > \pi^{SS} > \pi^{FF} > \pi^{FS}$$

Clearly, when innovation is drastic, we have: $\pi^{SF} = \pi^m, \pi^{SS} = \pi^d, \pi^{FS} = 0$, and $\pi^{FF} = \pi$, and if the innovation is non-drastic, then $\pi^{SF} < \pi^m$ and $\pi^{FS} > 0$. So under process innovation, (1), (2), and (3) can be rewritten, more generally as:

$$E(C*) = \rho\pi^{SS} + (1-\rho)\pi^{FF} - \frac{R}{2} \tag{1*}$$

$$E(\text{NC}*) = \rho^2\pi^{SS} + \rho(1-\rho)\left(\pi^{SF} + \pi^{FS}\right) + (1-\rho)^2\pi^{FF} - R \tag{2*}$$

$$E(C*) > E(\text{NC}*) \Leftrightarrow \rho(1-\rho) < \frac{R}{2\left[\pi^{SF} + \pi^{FS} - \pi^{SS} - \pi^{FF}\right]} \tag{3*}$$

Here, we assume that $[\pi^{SF} + \pi^{FS} - \pi^{SS} - \pi^{FF}] > 0$.[6] So Marjit result, as stated in Proposition 1, holds for non-drastic innovation.

In the following analysis, we focus on the following assumptions stated explicitly or implicitly in Marjit (1991) model.

(A1) Single process innovation;
(A2) Cooperative R&D occurring in a single lab;
(A3) Neither patent protection available nor imitation of technology possible;
(A4) Technology transfer under non-cooperative R&D not allowed;
(A5) There is no asymmetry of information.

[6]For the linear demand function, we have necessarily $\left[\pi^{SF} + \pi^{FS} - \pi^{SS} - \pi^{FF}\right] > 0$.

We shall relax one or the other assumption stated above and examine whether, or to what extent, Marjit (1991) results will undergo a change. In particular, we show that the qualitative result of Marjit (1991) will remain unchanged even if we introduce patent protection, imitation, technology transfer, or incomplete information in the analysis although incentives for cooperative or non-cooperative research will be affected.[7] On the other hand, if we consider cooperative research in two labs (hence, duplicating research), the result of Marjit (1991) will substantially change. We shall also compare Marjit (1991) result with the case when there is more than one conceivable innovation.

2.2 *Patent Protection and Imitation*

First, consider that patent protection for new innovation is available, hence it prevents imitation. Further, note that patent protection cannot affect the expected cooperative payoff because both firms have always the same access to information—patent is granted to the R&D cooperation.[8] Under non-cooperative R&D, if only one firm is successful to innovate, it gets patent protection. But when both the firms are successful, we assume that each firm will get patent protection with probability ½. This partly takes care of the fact that getting a patent protection may sometimes be uncertain (Mukherjee and Ray 2009). Assuming drastic innovation along with patent protection, the expected payoff under cooperative R&D will be as usual given by (1), but the expression of the expected payoff under non-cooperative R&D will be accordingly modified to get:

$$E(\mathrm{NC1}(a)) = \left[\frac{\rho^2}{2} + \rho(1-\rho)\right]\pi^m + (1-\rho)^2\pi - R \tag{4}$$

Then comparing (1) and (4), cooperative R&D is to be preferred to non-cooperative R&D if and only if $E(C0) > E(\mathrm{NC1}(a))$, that is, if and only if

$$\rho(1-\rho) < \frac{R}{2\left[\pi^m - \left(\pi^d + \pi\right)\right]} - \frac{\left[\pi^m - 2\pi^d\right]}{2\left[\pi^m - \left(\pi^d + \pi\right)\right]}\rho^2 \tag{5}$$

Since the RHS of (5) is a falling (and concave) function of ρ, comparing Marjit (1991) [hence (3)], one can easily see that the probability interval for non-cooperative R&D goes up, therefore, patent protection gives a larger incentive for non-cooperative R&D. Note that

[7] In an interesting paper, Bandyopadhyay and Mukherjee (2014) have shown that the possibility of entry by a non-innovating firm may also affect the incentive for cooperative R&D depending on the extent of spillover of knowledge.

[8] Given the possibility of infringement of patents, Marjit et al. (2001) have shown that the patent infringement agreements between the innovating firms may act as cooperation in R&D.

$$\text{RHS}(\rho = 1) < 0 \quad \text{for} \quad R < \pi^m - 2\pi^d$$

Therefore, when $R < \pi^m - 2\pi^d$, $\exists\tilde{\rho}$ such that non-cooperative R&D is preferred $\forall\rho \in (\tilde{\rho}, 1)$.[9]

Now consider that patent protection is not available and each firm is capable to imitate the other's innovation perfectly. Hence, in this case under non-cooperative R&D, the market will always be duopoly. Then, the expected payoff of a firm under non-cooperative R&D will be:

$$E(\text{NC1}(b)) = \left[\rho^2 + 2\rho(1-\rho)\right]\pi^d + (1-\rho)^2\pi - R \tag{6}$$

Hence,

$$E(C0) > E(\text{NC}(1(b)) \Leftrightarrow \rho(1-\rho) < \frac{R}{2\left(\pi^d - \pi\right)} \tag{7}$$

Thus, even when imitation is possible, the basic qualitative result of Marjit (1991) remains unaltered. But comparing the RHS of (3) and (7), we see that $\frac{R}{2(\pi^d - \pi)} > \frac{R}{2[\pi^m - (\pi^d + \pi)]}$, this means the relevant probability interval for cooperative R&D goes up, hence imitation increases incentives for cooperative research.

So, we can write the following proposition.

Proposition 2 *Patent protection in Marjit (1991) model enhances R&D incentives for non-cooperative R&D whereas imitation does reduce it.*

2.3 Technology Transfer Under Non-cooperative R&D

The possibility of technology transfer under non-cooperative R&D arises when only one firm comes up with a non-drastic process innovation successfully[10] and the innovator shares its knowledge with the other firm against some payment. Further, assume that there is neither patent protection nor imitation of the innovation. We consider fee licensing, that is, licensing of the superior technology under a fixed fee contract. The licensing contract takes place only on *ex post* innovation when it is profitable.

[9] We show later that in Combs (1992) non-cooperative R&D is preferred to cooperative R&D only for small values of the probability of success.

[10] Mukherjee and Marjit (2004) have introduced technology transfer in Combs (1992) framework and Kabiraj and Kabiraj (2019) in Marjit (1991) framework. The latter paper studies the choice of R&D organization when the form of R&D cooperation is also endogenously determined.

Following Marjit (1990) and others, fee licensing in a duopoly is profitable if and only if the post-licensing industry profit is larger than the pre-licensing industry profit,[11] that is,

$$2\pi^{SS} > \pi^{SF} + \pi^{FS}$$

For the linear demand function and also for the general demand function (but with some restriction), the condition will hold if and only if $\varepsilon < \varepsilon^0$, that is, the size of the innovation is below a critical level. In the remaining analysis, we continue to assume that

$$\pi^{SF} + \pi^{FS} \gtreqless 2\pi^{SS} \Leftrightarrow \varepsilon \gtreqless \varepsilon^0$$

Therefore, given that the size of the innovation is small, the expected payoff of a firm under non-cooperative R&D (with technology transfer) is:

$$E(\text{NC2}) = \rho^2\pi^{SS} + \rho(1-\rho)\left(\pi^{SS} + F\right) + (1-\rho)\rho\left(\pi^{SS} - F\right) + (1-\rho)^2\pi^{FF} - R$$

where F is the license fee given by $F = \pi^{SS} - \pi^{FS}$. Hence,

$$E(\text{NC2}) = \rho(2-\rho)\pi^{SS} + (1-\rho)^2\pi^{FF} - R \tag{8}$$

Since we have considered non-drastic process innovation, the expected payoff under cooperative R&D will be given by the expression of $E(C*)$ in (1*).

Therefore, given that technology transfer is profitable (i.e., $\varepsilon < \varepsilon^0$), cooperative R&D will occur if and only if,

$$E(C*) > E(\text{NC2}) \Leftrightarrow \rho(1-\rho) < \frac{R}{2\left(\pi^{SS} - \pi^{FF}\right)} \tag{9}$$

This means the qualitative result of Marjit (1991) remains unchanged even if we allow the possibility of technology transfer. Comparing the RHS of (3*) and (9), we see that

$\frac{R}{2\left[\pi^{SF}+\pi^{FS}-\pi^{SS}-\pi^{FF}\right]} > \frac{R}{2\left(\pi^{SS}-\pi^{FF}\right)}$, since $\varepsilon < \varepsilon^0$.

This means the probability interval of cooperative research falls under technology transfer; hence, technology transfer increases incentives for non-cooperative research.[12]

Proposition 3 *Possibility of technology transfer increases the incentive for non-cooperative R&D.*

---

[11] If, in this structure, we include the possibility of a royalty contract, then royalty contract will strictly dominate fee contract, and royalty contract is always profitable (see Wang 1998).

[12] See also Mukherjee (2005) for similar result.

2.4 Cooperative Research in Two Labs

Marjit (1991) assumed that cooperative research occurs in one lab. By this, the firms can save their R&D costs. In Combs (1992), cooperative R&D occurs in each of their labs. Doing research in two labs implies duplicating the research effort. There is no R&D cost saving. Yet in the context of Marjit (1991) model, it can be shown that doing cooperative research in two labs independently and sharing the results of R&D can sometimes be a better option.[13] Under cooperative agreement, if at least one firm is successful, both the firms will have access to the innovation. This occurs with probability $\rho(2-\rho) > \rho$.

Assuming drastic innovation, the expected payoffs of each firm under cooperative R&D will be:

$$E(C3) = \rho(2-\rho)\pi^d + (1-\rho)^2\pi - R \tag{10}$$

The expected payoff under non-cooperative R&D is given by $E(\text{NC0})$ in (2). We can check that in this case

$$E(\text{NC0}) > E(C3)\forall\rho \tag{11}$$

because $\pi^m > 2\pi^d$ for homogeneous good. Thus, if cooperative research is to conduct in two laboratories, then cooperative research will never occur. Although under cooperative research success occurs with a higher probability, this gain is not sufficient to compensate for the loss due to having no cost-saving under cooperative R&D. Therefore, duplicating research with drastic innovation does not yield any incentive for cooperative research vis-à-vis non-cooperative research.[14]

Proposition 4 *When cooperative research is to occur in two labs, non-cooperative R&D will strictly dominate cooperative R&D.*

It will be interesting in this context to compare the results with the case of Combs (1992) type research. Combs (1992) assumed to have multiple research projects (say, $m \geq 2$) out of which only one project is successful. Hence, under non-cooperative R&D, the probability of success is $\rho = (1/m)$. Under cooperative research, the R&D cooperation will select two projects, one for each lab; hence, the probability of success under cooperative research is $2\rho = (2/m)$.

At this moment without restricting to drastic or non-drastic innovation, the expected payoff of a firm under cooperative R&D will be:

$$E(C3a) = 2\rho\pi^{SS} + (1-2\rho)\pi^{FF} - R \tag{12}$$

[13] This can be the case when the probability of success lies in an intermediate interval. See Kabiraj and Kabiraj (2019).

[14] Note that 'drastic' innovation means innovation size is sufficiently large and we get the result that non-cooperative research strictly dominates cooperative research. If, instead of drastic innovation, we consider non-drastic innovation, then cooperative research will be the choice if and only if the size of the innovation is sufficiently small.

But the expected payoff under non-cooperative, R&D will be given by $E(\mathrm{NC}*)$ in (2*).

Therefore,

$$E(C3a) > E(\mathrm{NC}*) \Leftrightarrow \rho > \frac{\pi^{SF} + \pi^{FS} - 2\pi^{SS}}{\pi^{SF} + \pi^{FS} - \left(\pi^{SS} + \pi^{FF}\right)} \equiv \rho^0 < 1 \tag{13}$$

Combs (1992) assumed product innovation (or drastic process innovation), i.e., $\pi^{SF} = \pi^m$ and $\pi^{FS} = 0$. Further, in Combs (1992), $\pi^{FF} = 0$, (although it is not required for the result). Then (13) is reduced to the following Combs (1992) condition:

$$\rho > \frac{\pi^m - 2\pi^d}{\pi^m - \pi^d} \equiv \rho^c; \quad 0 < \rho^c < 1 \tag{14}$$

This gives that cooperative R&D occurs when the probability of success is above a critical level, i.e., $\rho \in (\rho^c, 1]$. In contrary, in the context of Marjit (1991) with cooperative research in two labs, cooperative R&D never occurs [see (11)].

Now, consider non-drastic innovation, hence consider (13). We have already taken that $\pi^{SF} + \pi^{FS} \gtreqless 2\pi^{SS}$ according to $\varepsilon \gtreqless \varepsilon^0$. Therefore, if $\varepsilon > \varepsilon^0$ (including the case of drastic innovation), (13) yields $0 < \rho^0 < 1$ and we shall get back Combs (1992) result. On the other hand, if we assume $\varepsilon \leq \varepsilon^0$, we will have $\rho^0 \leq 0$. In this case, cooperative R&D is always preferred to non-cooperative R&D irrespective of the size of ρ. The size of the innovation determines the critical value of ρ.

2.5 R&D Under Incomplete Information

In Marjit (1991), there is no asymmetry of information. So, we can think asymmetric information about the success and failure of R&D. Under cooperative R&D, this will not matter because the firms do research jointly. But under non-cooperative R&D, assume that the firms have asymmetric information, that is, whether or not a firm is successful to innovate is private information. However, the rival has a prior belief about the outcome. Details of the model can be found in Kabiraj and Chatterjee (2015). It is shown that incomplete information about the R&D outcome reduces the expected payoff of a firm under non-cooperative R&D; therefore, compared to complete information, incomplete information increases incentives for cooperative research. However, the qualitative result of Marjit (1991) goes through.[15]

[15] Chattopadhyay and Kabiraj (2015) have shown that under incomplete information about the size of the innovation, the qualitative result of Marjit (1991) will hold irrespective of whether it is quantity or price competition in the product market.

2.6 Conceivable Two Innovations

In this subsection, we consider conceivable two products (say X and Y) which could be innovated by the firms. To simplify the algebra, assume that products are independent and market demands for the products are identical.

As before, innovation of each such good involves an investment $R > 0$, and innovation is uncertain. Further, one firm cannot take more than one research project at a time. We continue to assume no patent protection and no imitation,[16] as in Marjit (1991). While the probability of success of one research laboratory is ρ, but a firm chooses X or Y with probability ½.

Then expected payoff under non-cooperative R&D will be given by,

$$E(\text{NC5}) = \left(\frac{1}{2}\right)\rho\pi^m + \left(\frac{1}{2}\right)\left[\rho^2\pi^d + \rho(1-\rho)\pi^m\right] - R \tag{15}$$

Now, consider cooperative R&D. If it occurs in a single laboratory, the firms can choose any product and then the expected payoff of a firm under cooperative R&D will be

$$E(C5(i)) = \rho\pi^d - \frac{R}{2}.$$

Then
$E(C5(i)) > E(\text{NC5})$ if and only if $\frac{R}{2(\pi^m - \pi^d)} > \rho - (\rho^2/2)$.

One can check that the RHS is an increasing function of ρ, hence cooperative R&D is to be preferred if and only if the probability of success is small, otherwise non-cooperative R&D is preferred. This is reverse of Combs (1992) result.

On the other hand, if cooperative research occurs in two laboratories, they can take the same research project in both laboratories or different research projects in different labs. The corresponding expected payoffs of a firm will be, respectively,

$$E(C5(ii)) = [\rho^2 + 2\rho(1-\rho)]\pi^d - R = \rho(2-\rho)\pi^d - R$$
$$E(C5(iii)) = [2\rho^2 + 2\rho(1-\rho)]\pi^d - R = 2\rho\pi^d - R$$

Since we have E$(C5(iii)) > E(C5(ii))$, the research cooperation will take different research projects in different labs. Then cooperative R&D will be preferred to non-cooperative R&D if and only if

$$E(C5(iii)) > E(\text{NC5}) \Leftrightarrow \rho > \rho^* \equiv \frac{2\left[\pi^m - 2\pi^d\right]}{\pi^m - \pi^d} \tag{16}$$

[16] For the case of patent protection or imitation and for further details, see Kabiraj (2006).

Now,

$$\rho^* < 1 \quad \text{iff} \quad \pi^m < 3\pi^d \tag{17}$$

Therefore, if $\pi^m < 3\pi^d$, cooperative R&D will dominate non-cooperative R&D for all $\rho \in (\rho^*, 1]$. This is similar to Combs (1992) result. On the contrary, if $\pi^m \geq 3\pi^d$, (so that $\rho^* \geq 1$), non-cooperative R&D will dominate cooperative R&D.

3 Possible Extensions

In our brief survey, on the choice of R&D institution or organization, we have restricted to Marjit (1991) framework and discussed whether, or to what extent, Marjit result will undergo a change with respect to different assumptions related to R&D investment. However, there are some issues which are important but not yet fully explored in the literature. The future researchers and scholars may find it interesting to explore these issues further.

We have shown that the size of the innovation is an important determinant of the choice between cooperative and non-cooperative research. The present work assumes that the level of R&D investment, and hence the size of the innovation, is exogenous. In our context, therefore, one may endogenously determine the size of the innovation and then examine the choice of R&D organization in the presence of uncertainty in R&D. It might also be interesting to study the implication of the R&D cost having its variable component.

In our analysis, R&D success is uncertain, but the probability of success is assumed constant and exogenously fixed. However, probability of success, to a large extent, depends on the R&D organization itself and on the level of R&D investment. This means, probability of success in cooperative R&D might not only differ from that in the non-cooperative R&D, but, more importantly, it can be endogenously determined. To the extent there is coordination problem under cooperative research, the same level of R&D investment will lead to different probabilities of success under cooperative and non-cooperative R&D. This consideration will perhaps provide a new dimension and further insight into the choice between cooperative and non-cooperative R&D.

4 Conclusion

The present paper has reviewed the choice between cooperative and non-cooperative R&D under various scenarios when R&D success is uncertain. In particular, we have studied the effect of the change of one or the other assumption underlying Marjit (1991) model. Generally, low probability of success or large uncertainty induces the choice in favor of cooperative research. Similarly, imitation possibility and incomplete information about the size of the innovation tilt the choice toward cooperative R&D. On the other hand, availability of patent protection or the compulsion of doing research in multiple labs will reduce the incentive for cooperative research.

We have also suggested some possible extensions of the present work. Research in this direction is likely to give a further insight into the problem of R&D and the choice of R&D organization.

References

Amir, R., Evstigneev, I., & Wooders, J. (2003). Non-cooperative versus cooperative R&D with endogenous spillover rates. *Games and Economic Behavior, 42,* 183–207.

Bakhtiari, S., & Breunig, R. (2018). The role of spillovers in research and development expenditure in Australian industries. *Economics of Innovation and New Technology, 27,* 14–38.

Bandyopadhyay, S., & Mukherjee, A. (2014). R&D cooperation with entry. *The Manchester School, 82,* 52–70.

Chattopadhyay, S., & Kabiraj, T. (2015). Incomplete information and R&D organization. *Economics Bulletin, 35,* A3.

Choi, J. P. (1992). Cooperative R&D with moral hazard. *Economics Letters, 39*, 485–491.

Choi, J. P. (1993). Cooperative R&D with product market competition.*International Journal of Industrial Organization, 11*, 553–571.

Combs, K. L. (1992). Cost sharing versus multiple research project in cooperative R&D. *Economics Letters, 39,* 353–357.

d'Aspremont, C., & Jacquemin, A. (1988). Cooperative and non-cooperative R&D in duopoly with spillovers. *American Economic Review, 5,* 1133–1137.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 18,* 1661–1707.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. *American Economic Review, 76,* 984–1001.

Kabiraj, A. (2018). R&D cooperation in a three firm Cournot industry. *Economics Bulletin, 38,* 1541–1546.

Kabiraj, T. (2006). On the incentives for cooperative and non-cooperative R&D in duopoly. *Arthaniti, 5,* 24–33.

Kabiraj, T. (2007). On the incentives for cooperative research. *Research in Economics, 61,* 17–23.

Kabiraj, T., & Chattopadhyay, S. (2015). Cooperative versus non-cooperative R&D incentives under incomplete information. *Economics of Innovation and New Technology, 24,* 624–632.

Kabiraj, A., & Kabiraj, T. (2019). Cooperative versus non-cooperative R&D in a Cournot duopoly under the possibility of research duplication and technology transfer. *Keio Economic Studies* (forthcoming).

Kabiraj, T., & Mukherjee, A. (2000). Cooperation in R&D and production: A three-firm analysis. *Journal of Economics (Zeitschrift), 71,* 281–304.

Kamien, M., Muller, E., & Zang, I. (1992). Research joint ventures and R&D cartels. *American Economic Review, 82,* 1293–1306.

Katz, M. (1986). An analysis of cooperative research and development. *RAND Journal of Economics, 17,* 527–543.

Marjit, S. (1990). On a non-cooperative theory of technology transfer. *Economics Letters, 33,* 293–298.

Marjit, S. (1991). Incentives for cooperative and non-cooperative R and D in duopoly. *Economics Letters, 37,* 187–191.

Marjit, S., Mukherjee, A., & Shi, H. L. (2001). Cooperation in R&D: The case of patent infringement agreements. *Journal of Economic Behavior & Organization, 45,* 383–401.
Motta, M. (1992). Cooperative R&D and vertical product differentiation. *International Journal of Industrial Organization, 10*, 643–661.
Mukherjee, A. (2005). Innovation, licensing and welfare. *The Manchester School, 73,* 29–39.
Mukherjee, A., & Marjit, S. (2004). R&D organization and technology transfer. *Group Decision and Negotiation, 13,* 243–258.
Mukherjee, A., & Ray, A. (2009). Unsuccessful patent application and cooperative R&D. *Journal of Economics, 97,* 251–263.
Ornaghi, C. (2006). Spillovers in product and process innovation: Evidence from manufacturing firms. *International Journal of Industrial Organization, 24,* 349–380.
Silipo, D. B., & Weiss, A. (2005). Cooperation and competition in an R&D market with spillovers. *Research in Economics, 59,* 41–57.
Spence, M. (1984). Cost reduction, competition and industry performance. *Econometrica, 52*, 101–121.
Suzumura, K. (1992). Cooperative and non-cooperative R&D in an oligopoly with spillovers. *American Economic Review, 82,* 1307–1320.
Vonortas, N. S. (1997). Research joint ventures in the US. *Research Policy, 26,* 577–595.
Wang, X. H. (1998). Fee versus royalty licensing in a Cournot duopoly model. *Economics Letters, 60,* 55–62.

Informal Sector: Progression or Persistence?

A Study of Four Traditional Clusters of West Bengal, India

Abhik Mukherjee, Nirupam Saha, Pratip Kumar Datta and Saumya Chakrabarti

Abstract In the context of development, the informal sector assumes importance owing to its large workforce; all the more so, because it is a major site of exclusion. A study of this sector has, therefore, become essential from the standpoint of the mainstream strategy of inclusive growth. It is opined by the dominant discourse that a major part of this informal sector could, in fact, act as dynamic micro-capital and thereby, could engender an inclusive growth process. In this context, we have tried to understand, whether the informal firms have the potential to survive and grow as micro-capital or they are just subsisting and even decaying as non-capitalistic petty firms, using quantitative as also qualitative information collected through primary surveys, focus group discussions and individual interviews on a variety of unorganised/informal manufacturing units engaged in iron forging and fabrication and in handloom weaving. We have tried to understand these tendencies focusing on various aspects like production organisation, especially owner–labourer relation, market structure, especially hierarchies and bottlenecks and on abilities/intensions of the firms to survive/grow. We have also considered certain firm-characteristics as well as socio-economic-cultural features of the surveyed locations to analyse how firm behaviours (as micro-capital or non-capital) are influenced by these firm-traits and environment. Summarising our results and engaging with the literature, we could say that, although, a few firms may be behaving like dynamic micro-capital (Bardhan in Econ Polit Wkly 44:31–36, 2009; Marjit and Kar in The outsiders: economic reform and informal labour in a developing economy, Oxford University Press, New Delhi, 2011) having symptoms of progression and some control over markets, overwhelming majority, without much of owner–worker separation and almost no control over markets, is able and/or interested in maintaining their existence only (Sanyal in Rethinking capitalist development: primitive accumulation, governmentality and post-colonial capitalism, Routledge, New Delhi, 2007; Chatterjee in Econ Polit Wkly 19:53–62, 2008).

A. Mukherjee · N. Saha · S. Chakrabarti (✉)
Department of Economics and Politics, Visva-Bharati (University), Santiniketan 731235, India
e-mail: saumya_chakrabarti@biari.brown.edu

P. K. Datta
Rajatpur Indranarayan Vidyapith, Rajatpur, Birbhum, West Bengal 731204, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_7

Keywords Production organisation · Owner–worker relation · Market structure · Market dependence · Micro-capital · Non-capital · Accumulation/survival

1 Introduction

1.1 Prologue

In the context of development, the non-farm informal sector assumes importance owing to its large workforce (La Porta and Shleifer 2014); all the more so, because it is a major site of exclusion (Chen 2007). Consequently, a sizeable literature has emerged on this informal sector.

It is advocated as a probable engine of progressive transformation by the mainstream researchers and policy makers. And the informal firms are projected mostly as micro-capital, linked with the global economy in one way or the other and as one of the most dynamic economic entities (Ranis and Stewart 1993; De Soto 2000; Bardhan 2009; Marjit and Kar 2011).[1]

However, heterodox researchers are critical of such optimism. The structuralists (Moser 1978; Mezzadri 2010; Breman 2013) see the informal sector as a site of petty commodity production with either absence or very little separation between capital/owner and labour/hired workers. These researchers also look at this sector as an appendage to the formal sector that exploits it through a variety of market and non-market institutions and perpetuates its existence. On the other hand, postcolonial researchers describe the informal sector as a 'need economy', whose primary objective is not accumulation of capital and hence, growth, rather fulfilment of need (Sanyal 2007; Chatterjee 2008).[2]

Contrarily, in the Indian context, it is shown that, while a part of the informal sector prospers, a large section remains stagnant. Further, researchers show that linkages with the formal sector, in general and local linkages, in particular, play a significant role (Kundu 1993; Mitra 1994,1998; Liebl and Roy 2003; Mukherjee 2004a, b; Kundu et al. 2005; Mitra and Mitra 2005). On the other hand, a group of researchers posit the informal sector as a composite of capitalistic and non-capitalistic firms. While few firms behave like micro-capital, use hired labour, participate in a variety of relationships with corporate capital and accumulate; majority of the firms have non-capitalist production structures, are mostly dependent on household labour, exploited by the formal sector through a variety of market and non-market relationships and able to produce either very little or no economic surplus for re-investment and accumulation (Khasnabis and Nag 2001; Kundu and Chakrabarti

---

[1] "In fact, both informal manufacturing units and self-employed units accumulate fixed assets, invest and prosper, and they may do so even at a time when their formal counterparts show much less dynamism." (Marjit and Kar 2011, pp. 6).

[2] "I conceptualise it as a need economy. I see it as an ensemble of economic activities undertaken for the purpose of meeting needs, as distinct from activities driven by an impersonal force of systemic accumulation." (Sanyal 2007, pp. 209).

2008; Bhattacharya et al. 2013; Basole et al. 2015; Chakrabarti 2016; Raj and Sen 2016; Bhattacharya and Kesar 2018).

Finally, the following definition of the Indian informal sector could act as a launching pad for our own intervention.

> Informal sector may be broadly characterized as consisting of units engaged in the production of goods or services with the *primary objective of generating employment and incomes* to the persons concerned. These units typically operate at *low level of organisation*, with *little or no division between labour and capital* as factors of production and on a *small scale*. Labour relations, where they exist, are based mostly on casual employment, kinship, or personal or social relations rather than contractual arrangements with *formal* guarantees. (NSSO 2001, report no: 459, pp. 1; emphasis ours)

Thus, the Indian definition too indicates at the informal sector as an ensemble of non-capitalist producers, with the absence of capital–labour dichotomy, and is mainly guided by the objectives of generating employment and basic income for both the owner and the worker.

1.2 Our Interventions

Taking cue from the heterodox literature as mentioned just above, we can hypothesise the following:

1. While a small section of informal firms behaves like dynamic ('progressive') micro-capital, a very large section persists essentially as non-capital.
2. For these non-capitalistic enterprises:

    (a) there is a little separation (if any) between hired worker and owner (in terms of pattern and duration of work, skill and to some extent income as well); there is a significant extent of community participation and a variety of non-hierarchical processes determine the wage and work of the hired worker (if any);

    (b) there is drain of surplus due to a variety of input–output market distortions (due to market and non-market power) and other problems (related to market size, market access, raw material, labour supply and increasing competition);

    (c) the surplus is mostly used for an improvement of standard of living; thus, satisfaction of need, rather than accumulation, is the primary objective of such firms.

1.3 Methodology and Data Source

This study is a primary data-based work on a cross-section of informal/unorganised[3]/micro-firms of few traditional/natural clusters of West Bengal. As we have focused on clusters of micro-firms, rather than stray and extremely heterogenous service units scattered across rural–urban spaces, we have chosen manufacturing as our target of analysis. Further, so far as micro/unincorporated manufacturing sector is concerned, West Bengal's share is overwhelmingly larger than any other state (NSSO 2017, report no: 582).

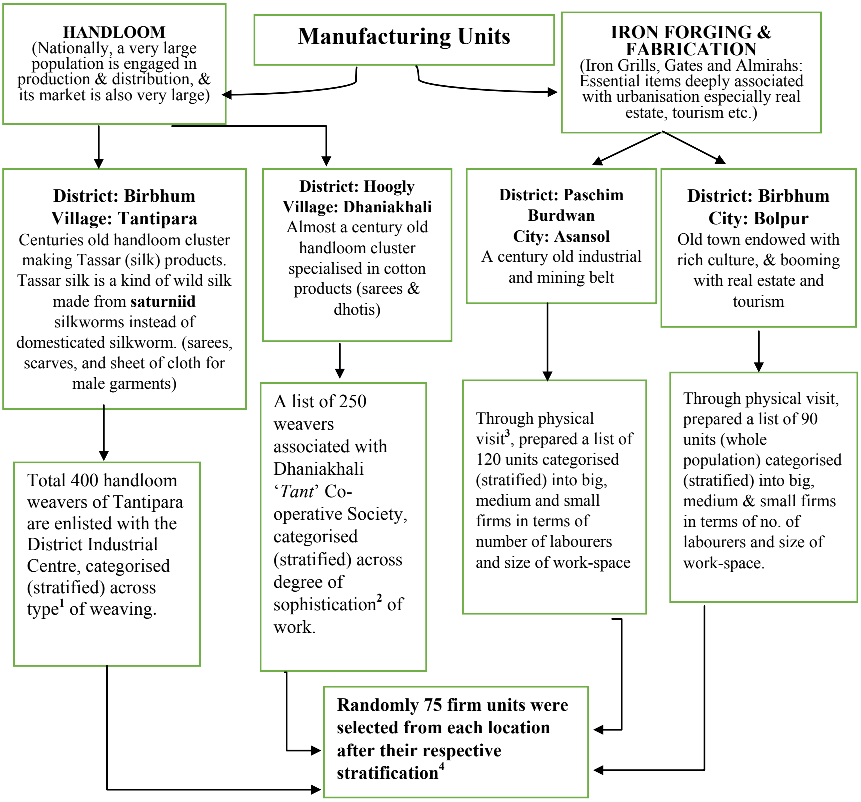
Local bodies and community-based organisations are the sources of additional extra-firm information. Data are collected via structured questionnaire, focus group discussions (FGDs) and individual interviews during August 2017 to February 2019. Some important FGDs, personal interviews and representative photographs could be found at: https://drive.google.com/open?id=1gG1tISMk3NHc1naz1nfibvY6UdbVSqpN. A total of 302 firm-units have been surveyed following a stratified random sampling technique sometimes supplemented with snowball method and complete enumeration as well, wherever necessary. We have chosen four locations from three districts for two specific products which are shown in the following way (Chart 1). The rationale of choosing the specific products and corresponding locations are presented in Chart 1, along with the rest of the sampling process. See Chart 1.

1.4 Some Sample Characteristics

Table 1 represents a profile of our sample firms across survey locations:

(a) while handloom is dominated by the Hindu owners, forging and fabrication units of Asansol and Bolpur are mostly owned by Muslims; (b) despite the presence of regulatory institutions and government agencies, Tantipara weavers and many of the producers of Bolpur and Asansol are not having trade licence, neither their units are registered; (c) the units are pretty old and duration of operation is almost similar across all the segments, with slight difference across handloom units and forging and fabrication units, as the former segment is fully household based, while the latter uses a somewhat separate production space; (d) average net income per firm, average value of asset per firm and per capita income of family labour as well as combined (family plus hired) value of labour productivity of the fabrication units are much more than that of the handloom units; (e) despite spending higher amount on raw material and labour, the firms of Bolpur are unable to generate even equal amount of net income vis-à-vis that of Asansol (elaborated subsequently); (f) the fabrication

[3] "In the unorganised sector, in addition to the unincorporated proprietary or partnership enterprises (i.e. informal enterprises), enterprises run by cooperative societies, trusts, private and public limited companies (Non-ASI) are also covered. The informal sector can therefore be considered as a subset of the unorganised sector" (NSSO 2001, report no: 459, pp. 3).



1. There are generally two types of tassar products woven in Tantipara - pure tassar and blend of pure and synthetic (Korea-made).
2. There are 3 types of sarees weaved in Dhaniakhali - 100s x 100s (most sophisticated), 80s x 80s and 60s x 60s. If a saree is M x N, it means there should be M and N number of threads in the 'Reed' and 'Pick' respectively in a square inch.
3. Physical visit was done in the main town of Asansol and surrounding areas like Burnpur, Riverside, Kulti, Dishergarsh and Neamatpur. A stratified random sampling based on complete enumeration of the units of Asansol, Burnpur, Riverside, Kulti and Dishergarh was done. We were unable to access the units in the interior of Neamatpur due to security reasons; moreover, we have not included the units of Kulti and Dishergarh due to sparsity and distance.
4. Supplemented by Snowball Method: replacing the units where owners were not available at the time of survey and where the units went out of business within the period from sampling to data collection (around 3 months).

Chart 1 Sampling process

units use both family and hired labour, whereas the handloom units are dependent on family and community labour; (g) family labourers work more (per day) in the fabrication units than the handloom units, although, the unit owners of Tantipara work most intensively; (h) in Asansol and Bolpur, both family labourer and hired worker work for around 8 h a day (without much of difference in terms of type of work as well, as will be elaborated subsequently); (i) average family labour income is around three times compared to skilled wage in fabrication (however, this difference is considerably less for smaller firms, as will be seen subsequently); (j) there are a variety of institutional arrangements in input and output markets; however, whatever be these institutional arrangements, the firms across all the locations suffer from a variety of market-related problems, as will be discussed subsequently; (k) while Bolpur and

Table 1 Sample profile across locations

Product categories	Handloom		Iron forging and fabrication	
Locations	Tantipara	Dhaniakhali	Asansol	Bolpur
Sample size	75	79	73	75
Sex of owner	All male	87% male	All male	All male
Religion of owner	100% Hindu	100% Hindu	50.68% (Muslim)	41.33% (Muslim)
Caste of owner (Majority)	OBC	OBC	General	General
Trade Licence (no. of units)	0	0	44	63
Registration (no. of units) (Non-SSI/Handloom/Artisan)	1	78 (cooperative)	23	30
Average age of unit (months)	321	454	153	158
Duration of operation (Days per month)	28.81	27.69	23.67	24.48
Duration of operation (Months per year)	11.94	12	10.31	10.92
Average turnover per firm (Rs./Month)	15,910	2957	138,509	149,585
Average net firm-income (Rs./Month)	6773.279	2647.363	25,126.46	24,442.23
Net income per family labourer (Rs./Month) = firm-income/family-labourer	2855.046	1179.233	23,278.32	22,464.62
Average no. of hired skilled labourer	–	–	2.0	2.1
Average skilled wage (Rs./Month)	–	–	7078.266	7980.597
Average no. of hired unskilled labourer	–	–	1.7	1.4
Average unskilled wage (Rs./Month)	–	–	3294.05	4240.31
Labourer productivity (Rs./Month) = total-revenue/total-labourer	6966.77	1332.652	26,716.07	29,999.3

(continued)

Table 1 (continued)

Product categories	Handloom		Iron forging and fabrication	
Average hours of work by family labour per day	6.6	6.6	8.1	7.8
Average hours of work by owner per day	9.1	7.8	8.1	8.1
Average hours of work by hired labour per day	–	–	7.94	7.83
Major input sources	Aratdars and Mahajans	cooperative	Wholesalers	Wholesalers
Major output destinations	Traders, contractors and Mahajans	cooperative	Individual buyers and contractors	Individual buyers
Average value of asset per firm (Rs.)	14,821.3	10,968.4	73,942.5	62,736.7
Average cost of raw material per firm (Rs./Month)	8884	214 (raw material from cooperative)	91,076	95,678
Average cost of labour per firm (Rs./Month)	–	–	20,242	23,718
Average cost on repair of machineries, etc. (depreciation) (Rs./year)	1200	9825	12,183	13,056
Average investment (last year in Rs.)	0	580 (30 units did not invest and 18 units invested Rs. 800 each)	14 (only 1 firm invested)	4693 (45 units did not invest at all)

Source Authors' compilation from field survey

Asansol spend substantial amount on hired labour, Tantipara and Dhaniakhali use only household labour; (l) while Bolpur and Asansol spend sizable amount on raw material and Tantipara too has to spend some amount, Dhaniakhali gets it from the cooperative; (m) the old units of Tantipara and, especially Dhaniakhali have to spend significant amount against wear and tear and for design (through frequent change of loom accessories); Bolpur and Asansol too spend substantial amount as depreciation cost; (n) although there is a substantial use of raw material and family and/or hired labour and sizable depreciation cost, in all the locations, fresh investment is miserably low; the units seem to be more interested and/or able to maintain the existing scale of operation, however, they are neither interested nor able to invest to expand the scale of production.

1.5 Chapter Organisation

This chapter is organised in the following way. In the next section, we map some of the historical-social-cultural and economic characteristics of the clusters, and overall firm-level production organisations, market structures and firm-level performances, in general. This section presents certain qualitative information providing additional insights necessary to understand the empirical results of the subsequent section. Empirical analysis of the succeeding section tries to understand whether the surveyed firms behave like non-capitalist producers or micro-capital, despite their apparent differences in terms of some firm-level characteristics as well as social-cultural-economic environment. Finally, in the concluding section, we briefly engage with the theoretical debate on the characterisation of the informal sector, after summarizing our empirical results.

2 Beyond Firm: A Brief Description of Our Study Areas Along with Some General Firm-Level Characteristics

Our objective is to understand the firms in terms of owner–labour relationships, patterns of market dependence and pattern of expenditure across household consumption and investment. However, to comprehend this, it is essential to know the environment of the surveyed locations that offer the bedrock for any organisation to take birth, evolve, grow or die.

2.1 Tantipara: A Tale of Surplus Extraction Using Feudal Institutions

Tantipara (*locale of weavers*[4]) is a village just 18 km away from the district headquarter, Suri. The occupational pattern of the weavers has been mostly shaped through their interactions with the global traders of the Mughal period, feudalistic exploitations by the Mahajans during the British period and also the struggles of modern times.

[4]Who consider themselves as '*shilpi*', i.e. artists.

The production organisation of handloom tassar silk is a unique socio-cultural-economic practice. The weavers use manual looms, locally crafted instruments with the active participation of family, especially women and community involvement and little presence of hired labour. This process is skill-intensive, prolonged, laborious and collaborative.

The weavers work for more than 9 hours a day, putting around 16 hours of family work effort daily, to produce 4–5 *thaan* for a meagre monthly family income (approximately Rs. 7000). A family of three, work all day to earn their daily needs. The weavers' aspiration for earning more by adding value to their products with new designs is constrained by resource and time shortages.

The input–output markets are controlled by the *Aratdars*, contractors, Mahajans and traders resulting in price distortions. The input-price is artificially pushed up by a few *Aratdars*, from whom the weavers are compelled to buy even low-quality cocoons, as they are either denied access to the tribal cocoon-traders or constrained by limited cash-in-hand.

Despite moderate transport and communications, the weavers cannot sell their output in national-international markets, because of lack of access to trusted networks and small scale of production. This forward linkage is skewed in favour of the contractors, Mahajans and traders who suffocate the weavers by appropriating the benefit of high price of tassar silk in urban/global market. This is happening even in the era of globalisation of capital and perhaps, through a tacit understanding between globally linked capital and these intermediaries.

Moreover, government failure in the context of backward and forward linkages and inadequate technical support have led to discontent among the weavers.

Despite owning the loom and the workspace and despite a very rich experience and understanding about design, the weavers of Tantipara have not been able to expand their business and more strikingly, have not been able to produce innovative mechanisms of weaving and designs, most probably, because of the presence of long-term sub-contracting, putting-out and a variety of hierarchical (feudalistic) arrangements with the Mahajans, Aratdars, traders and contractors.

All these problems push the weavers to intensify self-exploitation and the younger generation towards menial factory jobs (having no/little access to agricultural land and Mahatma Gandhi National Rural Employment Guarantee (MGNREG) programme). Alarmingly, the young men are rejected by brides' family for this demanding but low-income lifestyle.

Details of Synthetic Thread Used in the Production of Cheap Tassar Product

Pure tassar-silk products are relatively high priced than other silk products, and nowadays, it is rare in supply but has high international demand. Thus, to meet this high demand, the weavers, being directed by the traders and Mahajans, use this synthetic (tassar) thread to weave cheap products which are sold at urban markets to reap additional surplus per unit by the intermediaries.

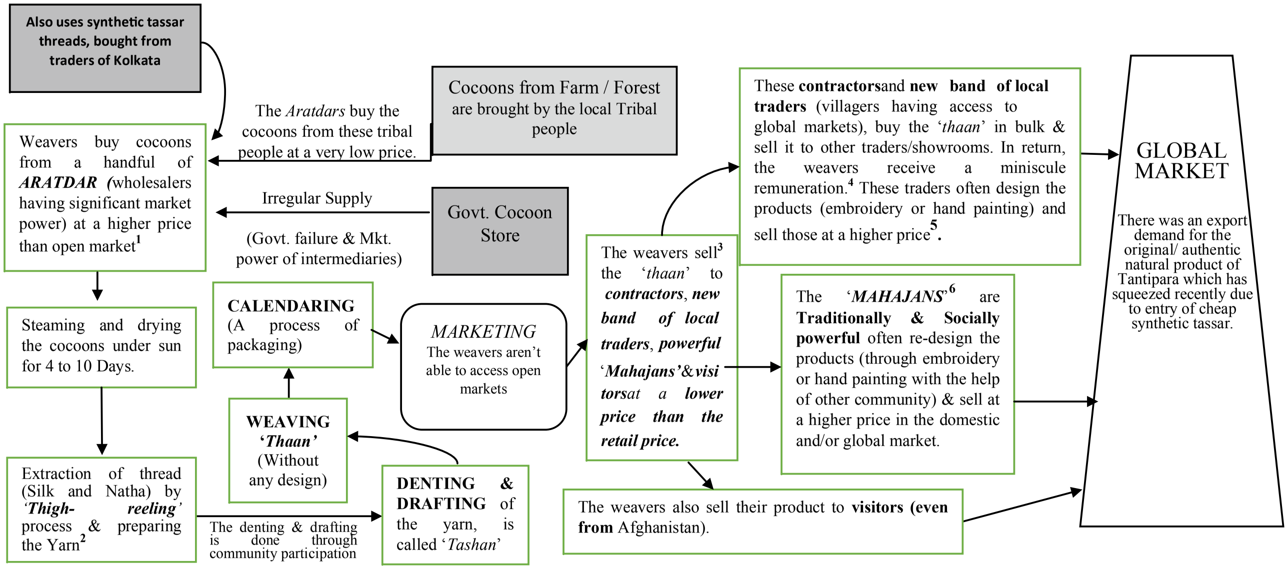
The weavers of Tantipara get this synthetic tassar thread from few traders who come from Kolkata (a metropolitan city nearby), for a price between Rs. 5000 and 5500 a kilogram. The production process for blended product and pure tassar product is no different.

Details of quantity and price of silk threads used in production

Items	Pure tassar silk required	Price (Rs.)	Synthetic tassar required	Price (Rs.)	% of weaver weave specific products (from sample)	Average remuneration/surplus of a weaver (per thaan)	Monthly income of a weaver (approx.)
Pure tassar product (per thaan)	950 g to 1 kg (approx.)	4200 (as in 2018)	–	–	64% (47 out of 74 weavers)	Rs. 1500–1800	Rs. 7700
Blended tassar product (per thaan)	250 g (approx.)	1200 (approx.) (as in 2018)	600–700 g (approx.)	3000 (as in 2018)	36% (27 out of 74 weavers)	Rs. 1000–1200	Rs. 6400

While calculating values, in Sect. 3 below, we have incorporated these issues appropriately
Source Authors' calculation from field survey data

See Chart 2.



1. The cocoons are generally sold at the end of Autumn (Oct.–Nov.). Price of one 'KAHAN' (1280 cocoons) increased from Rs.3000 to Rs.4200 over 2017-18. Government store generally sells cocoons during Jan.–Feb. instead of Autumn; but in 2019 no cocoons were available.
2. From a '*kahan*' of cocoon (depending on the quality), generally 1 kilogram of thread is extracted and one '*Thaan'* (2 Sarees) can be weaved. The process of '***Thigh reeling***' is done by the women possessing sex specific traits. Apart from the silk, a coarse by-product is also obtained locally known as '*NATHA*', which is sold at around Rs.600 a kilogram.
3. Since currently, final-product price-range is Rs.5200-5800 and cost of a Kahan of cocoon is Rs.4200, income-range is just Rs.1000-1800 weekly. Due to this laborious and less remunerative work, young people are going for other occupations.
4. Some local traders practice putting-out system with weavers against remuneration. Moreover, weavers lacking access to markets buy cocoons independently and sell the final product to the local traders.
5. Although a part of tassar 'thaan' is used after re-designing, the main value addition is done at the stage of 'thaan' production.
6. The weaver sells their output to Traders (who buy the product on spot, without any kind of pre-contract), sub-contractors (who pre-contract with the weaver for certain stipulated number of products but donot provide any financial assistance to buy raw materials) and Mahajans (with whom the weavers are engaged through neo-putting out system)

Chart 2 Production process and market structure of Tantipara

2.2 Dhaniakhali: A Tale of Unequal Competition and Lust for Life

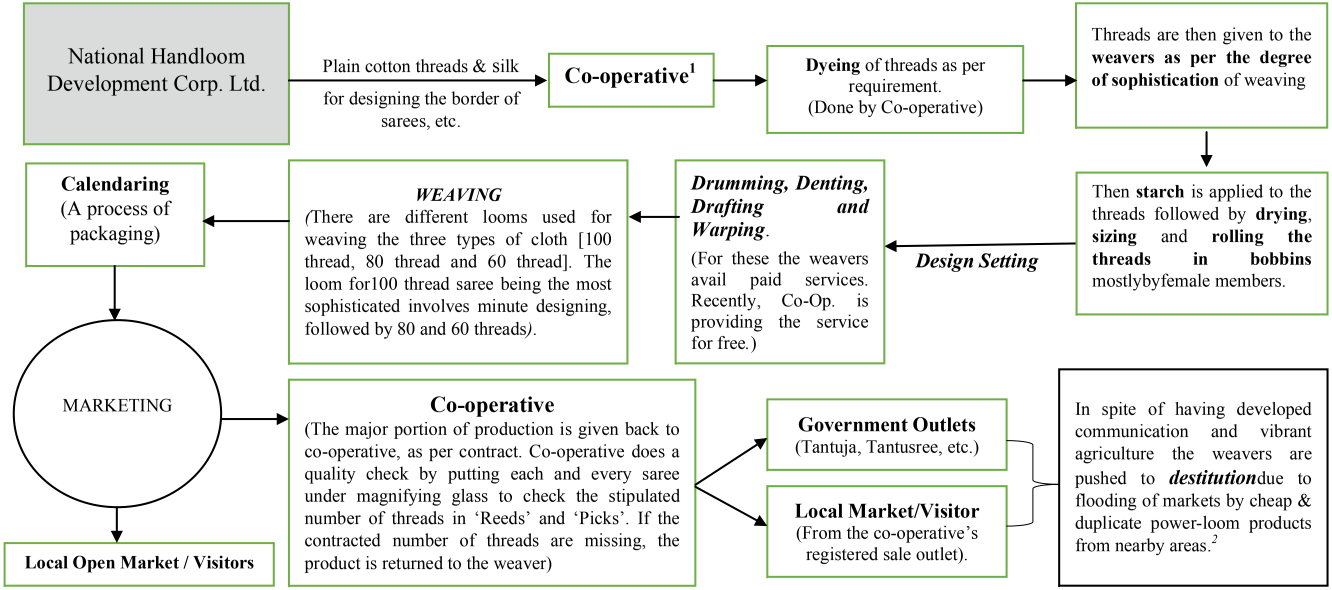
Dhaniakhali is an agriculturally prosperous village having strong urban linkages (61 km from Kolkata), where generations of weavers are engaged in handloom. The entire process of production is primarily organised by cooperatives against ‘remuneration’ to weaving-members. The production process is majorly household labour based with significant contribution of women. However, some stages of production require locally sourced skilled hired labour. The looms are heavy and require intense physical labour, as these weavers (unlike Tantipara) weave designed sarees with the help of additional accessories and intricate manoeuvering of their looms. Still, the weavers have to work for long hours to earn miniscule ‘allowance’ from cooperative. Lack of dependable government support in input–output markets, absence of direct/independent access to trusted urban market-networks, inadequate product sophistication and entry of cheap, duplicate and power-loom products from nearby locations are making the situation vulnerable for the cooperative and hence for the weavers. Therefore, despite the possibility of a variety of demand–supply support (from urban market and local agriculture), income from weaving is meagre. This situation is forcing the weavers to take part in MGNREGP along with other menial jobs like construction work. The cooperative struggles to keep the youth in this profession. The youth, however, fight to adapt to newer, distant, risky occupations for their survival, and similar to Tantipara, the bachelors are refused by the brides’ family. However, despite all these odds, weavers are not ready to compromise with the quality and distinctive nature of Dhaniakhali brand.

Not only the affinity to stick to the tradition, rigidity against shifting towards a new/different raw material or design and financial constraints restricting innovations but also the question of pride and more importantly, a scope for catering to a niche high-value national-international market are inducing them to stick to their present profession. See Chart 3 for further details.

2.3 Asansol and Bolpur: Stagnant Market and Intense Competition

Asansol is a famous industrial metropolitan city and Bolpur a municipal town reaping the benefits of tourism and real estate (of the twin town of Santiniketan); both having relatively advanced infrastructure.

Unorganised iron forging and fabrication is a major earning avenue for a large number of people in both locations. The production process is mainly based on skilled labour, machines, and electricity, using raw materials (iron rods, sheets, angles, welding-rod, chemicals, etc.). Hired labour is prominent in this process over and above family labour. In this work, skill is percolating through generations via (religion-specific) community interactions. Many of the firm owners begin as appren-



1. In 1948, the co-operative was formed by the weavers for safeguarding themselves from intermediaries. Currently, the cooperative provides the weavers coloured threads to weave 26 sarees (at least) in one '*CHARAN*' (one specific loom-design setting) estimating market demand. To avail of the benefits (bonus, allowances, paid leave, etc) of co-operative, the weavers are required to produce at least 48 sarees in a year. In return, the weavers get 'remuneration' based on degree of sophistication. One 100 thread saree takes 3 days of rigorous work for 3 persons working simultaneously. Against this, they get average remuneration of Rs.250 per saree; out of this, there is pre-weaving service cost of Rs.100 (approx.). So Rs.150 (net income per saree) is divided among 3 members working for three days. Thus, per-head per-day income is Rs.16.66 (approx.). For 80 thread sarees, this is Rs. 12 (approx.) and for 60 thread it is Rs. 8 to 10 (approx.).

Chart 3 Production process and market structure of Dhaniakhali tice in others' workshops and gradually start their own business. Thus, entry is not very costly and in fact, acquiring skill is perhaps the determining factor. Conversely, this owner acts as a skilled worker in his own firm. The owner–worker dichotomy gets blurred.

The outputs are locally sold to individual buyers or through contracts. Although those firms which are able to work through contracts with large-buyers/big-showrooms/real estate contractors/agents are better off, there are instances of contract failures. Moreover, intense competition in the output markets of Asansol and Bolpur is putting pressure on price. Increased competition from new micro-units opened by the distressed skilled workers (as elaborated afterwards) as well as from nearby villages is making the condition worse.

Conversely, the raw materials are locally sourced. Although in Bolpur, raw material supply is not reported as a serious bottleneck, in Asansol, increasing cost of raw material (due to market imperfection, introduction of GST, etc.) poses a serious problem. Additionally, in both the locations, there exists a shortage of skilled labour pushing up wage for the following reasons: (i) since the work is trying, not many young people are eager to enter into this profession; (ii) as it takes around 10 years for an unskilled labourer to become skilled, there is this long gestation period; and (iii) despite some rise in wage due to labour shortage, this increased remuneration is not adequate and hence, workers are shifting to start their own local workshops and even to other states and countries (mainly Middle East).

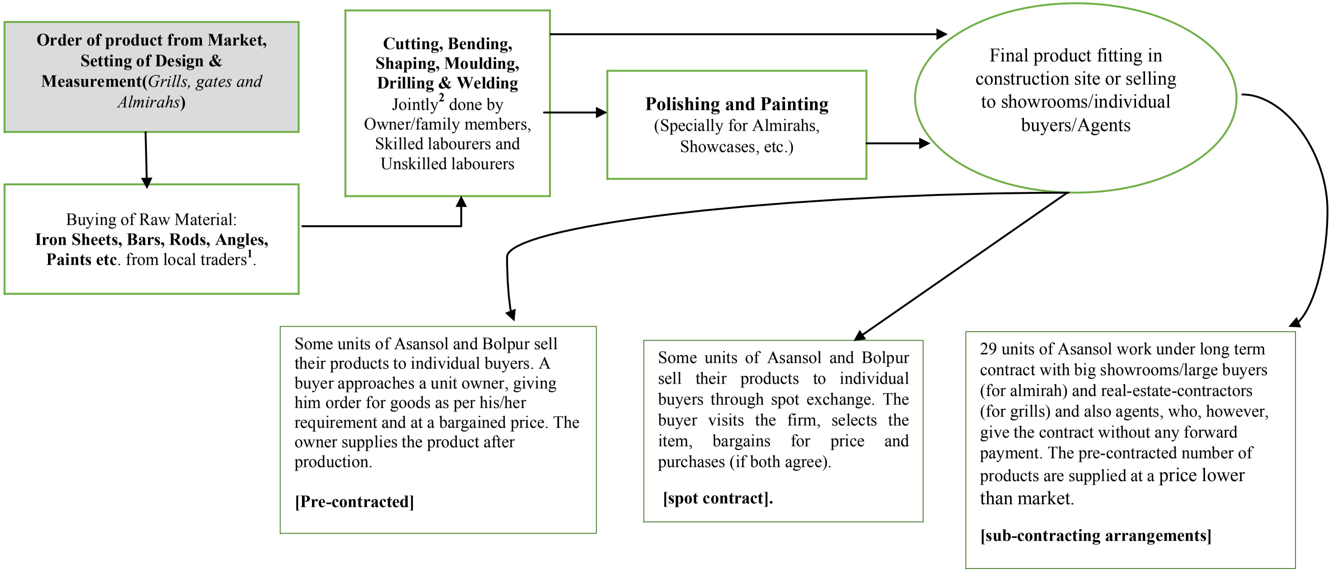
Thus, currently, deviating from the traditional owner–worker separation, sheer competition, on the one hand, and increasing costs of raw material and labour, on the other, are compelling the unit owners to physically take part in production along with managerial activities and thereby intensifying self-exploitation.[5] This self-exploitation is more intense in Asansol than Bolpur.

Thus, size of the output market, source of raw material and access to these are crucial for Bolpur and Asansol too (like Tantipara and Dhaniakhali). The micro-firms cannot have any control over the markets, rather the volume and structure of market determine their conditions. These market problems plague the firms of Bolpur and Asansol, especially those of the latter; the firms are left with very little surplus for re-investment and accumulation. Maintaining the current standard of living becomes an overriding concern. See Chart 4 for further details.

After this broad overview of the four locations presenting majorly firm-level production organisation, and market structure income-expenditure pattern along with

5

(a) "Now cannot keep worker. Even cannot manage own expenses; how can I pay them?"—An interviewee of Asansol. Thus, to reduce cost, the owner intensifies his work effort considerably.
(b) "The wages given to the skilled hired workers are not enough. To earn a bit more than the wage received, they open their own workshop which intensify the existing competition. This increased competition forces us now to work in the factory as worker."—An interviewee of Bolpur. Thus, the pressure of competition restricts the price and size of market; on the other hand, costs are rising. To maintain a minimum standard of living, the owner has to work.



1. As there are few raw material sellers, they control the price of raw material by exerting market power (in Asansol). In recent times, the price of raw material has increased along with GST application. However, the situation is not reported to be that difficult in Bolpur.
2. There are sizable number of skilled and unskilled child labourers, especially in Asansol, which has been considered in calculations of section 3.

Chart 4 Production process and market structure of iron forging and fabrication units some local level socio-economic-cultural traits, we now enter into our third section on primary data analysis.

3 Empirical Analysis

From our available cross-sectional data, we intend to locate the presence/absence of some specific traits of the firms to understand whether these firms behave like micro-capital or having mostly non-capitalistic features, in terms of labour use, market dependence and pattern of expenditure (as has already been discussed and hypothesised in Sect. 1.2 above).

We select the following specific traits: (a) use of family, community and hired labour and their relationship with the owner of the firm; (b) input–output market dependence and the associated problems; and finally (c) re-investment of surplus.

3.1 Role of Labour and Owner–Worker Relationship

We have found that (Table 2a and 2b) Dhaniakhali and Tantipara depend mostly on family labour (with some community participation),[6] whereas, in Asansol and Bolpur, the sample units use a combination of family and hired labour.

Strikingly for handloom, as we move along the revenue category, in both the locations, not only the number of household worker per firm is increasing but also there is a sharp increase in the average monthly hour of work per firm and per unit of family labour too; this perhaps indicates at a considerable increase in self-exploitation, which might be a major source of increased income in this product category.

On the other hand, while in Asansol, participation of family labour decreases as we move along the revenue categories, in Bolpur, this participation slightly rises. This is also supported by a fall (rise) in average hour of work (by family labour) per firm per month in Asansol (Bolpur). However, if we look at the labour hour devoted by each family member, this, on an average, is increasing across revenue categories in both the locations; this intensity being higher in Asansol, although the rise is faster in Bolpur; this perhaps indicates at a considerable increase in self-exploitation across revenue categories, especially in Asansol.[7]

---

[6]In Dhaniakhali, for all the sample firms, a specific part of the production process (drumming) is done by a few community members against a payment which has been incorporated within costs of production subsequently in Sect. 3.3.

[7]Our repeated visits to Asansol revealed that, within a span of just one year, establishments hiring labour have turned into self-employed units with the owner being compelled to undertake every part of the work; owner–managers are turning into owner–workers. Similar tendencies have been noted in Bolpur too.

Table 2 (a) Number of workers and monthly hours of work by family across revenue categories in Dhaniakhali and Tantipara. (b) Number of workers, monthly hours of work by family and hired labour and ratio of hired labour hour to total labour hour across revenue categories in Asansol and Bolpur

(a) Handloom

Dhaniakhali						Tantipara					
Revenue category	Family lab					Revenue category	Family lab				
	1	2	3	4	5		1	2	3	4	5
Rs. 199–1500 (Cat 1)	20	37	6194	309.7	167.41	Up to Rs. 7000 (Cat 1)	20	49	8385	419.25	171
Rs. 1501–2500 (Cat 2)	20	40	6923	346.15	173	Rs. 7001–18,000 (Cat 2)	22	48	9607	436.68	200.15
Rs. 2501–3500 (Cat 3)	22	52	8872	403.27	170.62	Rs. 18,001–25,000 (Cat 3)	19	43	8935	470.26	207.79
Rs. 3501 and above (Cat 4)	17	45	10,014	589.05	222.53	Rs. 25,001 and above (Cat 4)	14	43	7848	560.57	182.51

(b) Fabrication

Revenue Category	Asansol										Bolpur									
	Family lab					Hired lab				6	Family lab					Hired lab				6
	1	2	3	4	5	1	2	3	4		1	2	3	4	5	1	2	3	4	
Up to Rs.50,000 (Cat 1)	13	18	3129	240.69	173.83	13	21	3163	243.31	0.4	7	8	1143	163.28	143	7	8	1065	152.14	0.4
Rs.50,001–100,000 (Cat 2)	25	27	5620	224.8	208.15	25	72	12,326	493.04	0.7	25	33	5863	234.52	177.67	25	49	9344	373.76	0.6
100,001 and above (Cat 3)	35	39	8346	238.45	214	35	179	33,056	944.48	0.8	43	51	10,842	252.139	214.59	43	210	38,798	902.28	0.8

Source Authors' compilation from field survey data
1 Total number of units, 2 Total number of labourers 3 Total hour of work in a month. 4 Average hour of work per unit (monthly), 5 Average (monthly) hours of work of a family labourer involved in firm. 6 Ratio of hired labour hour to total labour hour (average)
Note: The Revenue Categories of this Table 2 (as designed above, across locations) are repeated in all subsequent tables. Hence, we just mention: Cat 1, Cat 2, so on, in the subsequent tables at the appropriate places.

Thus, despite distinctly different product categories, we find this similarity across locations: increase in self-exploitation of family labour with rising revenue. As the firms become larger, there is, in fact, greater participation of family labour (hour), even in those units which hire labour.

From Table 2b, if we look at the ratio of hired labour hour to total labour hour, we observe that a sizeable number of firms (almost 50%, mostly belonging to the lower revenue categories) use at least 30% of total labour time from family pool; even for the highest revenue category, this is 20%.

Furthermore, Table 3a, 3b and 3c show across different revenue categories:

(a) Despite similar intensity of work in terms of labour hour across family and skilled hired labour, for lower revenue categories, the difference between the incomes (per month) of family labour and hired labour is not considerable; this is a striking result, so far as owner–hired worker relationship is concerned. First of all, the owner has to put similar effort vis-a-vis his hired worker; further, the owner has to undertake direct physical work in production process along with managerial activities. Despite all these, he is able to earn an amount which is not much higher than his waged worker.

In Dhaniakhali and Tantipara, there is no hired worker and in Asansol and Bolpur, there is a little separation/difference between the owner and the hired worker in terms of skill, pattern of participation in production, intensity of work and to some extent, income per capita. Thus, our sample firms cannot be considered as capitalistic ones where capital–labour dichotomy is distinct.

Contrarily, this dichotomy is not only blurred but also the hired worker, after acquiring sufficient skill, can shift to open his own firm to earn a slightly better income (obviously, with some more freedom). On the other hand, in handloom sector, it is just petty commodity production. This proposition of our firms being mostly non-capitalistic is further strengthened by the following observation.

(b) 11% of the hired labourers of Asansol work for more than 8 hours a day, but only 17% work permanently on verbal contract; on the other hand, almost all the hired labourers of Bolpur work not more than 8 hours a day, although fully under temporary contract; despite long hours of work and long association between the owner and the hired worker, there is no formal/written permanent contract with any social security.
(c) Wages and duration of work for the hired labourers of Asansol are determined either by the firm owners or by the workers, whereas, the type of work is mainly directed by the firm owners. Contrarily, in Bolpur, wages and duration of work for the hired labourers are determined by market, whereas type of work is determined by the firm owners. Thus, wage and duration of work are not fully under the control of the firm owner.

Next, we consider some of the probable determinants of labour hiring in Asansol and Bolpur.

We set an OLS regression model (Regression model-1) to find out the probable determinants of labour hiring by the firms.

Table 3 (a) Intensity of work, income and wages for the owner and the skilled hired worker across revenue categories in Asansol and Bolpur. (b) Distribution and nature of contract of hired labour according to hours of work per day across revenue categories in Asansol and Bolpur. (c) Major determinants of wage, duration of work and type of work of the hired labourer across revenue categories in Asansol and Bolpur

(a)

Asansol							Bolpur					
Revenue Category	Total no. of firms	Total no. of family labourers	Average Per capita income of family labour (Rs./Month)	Average wage of skilled hired labour (Rs./Month)	Owners' duration of work (h/day)	Skilled hired worker's duration of work (h/day)	Total no. of firms	Total no. of family labourers	Average Per capita income of family labour (Rs./Month)	Average wage of skilled hired labour (Rs./Month)	Owners' duration of work (h/day)	Skilled hired worker's duration of work (h/day)
(Category 1)	13	18	5461.9	4521.4	8.07	7	7	8	7133.139	5750	8.14	7.5
(Category 2)	25	27	15,260.71	6598.8	8	7.7	25	33	8825.091	6977.4	7.88	8
(Category 3)	35	39	35,622.71	7891	8.14	7.5	43	51	32,890.41	8706.6	8.18	7.6

(b)

Locations	Asansol						Bolpur					
Revenue category	No. of units	No. of hired labourers	Up to 8 h/day	More than 8 h/day	Permanent contract	Temporary contract	No. of units	No. of hired labourers	Up to 8 h/day	More than 8 h/day	Permanent contract	Temporary contract
Category 1	13	21	21	–	8	13	7	8	8	–	0	8
Category 2	25	72	61	11	10	62	25	49	49	–	0	49
Category 3	35	179	161	18	28	151	43	210	208	2	3	207

(c)

Locations	Asansol	Bolpur
Determinants	Revenue	Revenue

(continued)

Table 3 (continued)

(c)

Locations	Asansol									Bolpur								
	Category 1			Category 2			Category 3			Category 1			Category 2			Category 3		
	Wage	Hrs of work	Type of work	Wage	Hrs of work	Type of work	Wage	Hrs of work	Type of work	Wage	Hrs of work	Type of work	Wage	Hrs of work	Type of work	Wage	Hrs of work	Type of work
By employer	8	14	15	31	44	61	70	95	120	0	5	8	3	6	41	56	60	177
By employee	3	5	4	9	17	11	53	52	55	0	0	0	1	1	4	0	18	3
By open market	0	2	2	13	4	1	28	32	0	3	2	0	42	42	0	132	108	0

Source Authors' compilation from field survey data
Note Minor determinants are excluded from this table. Corresponding total number of workers in each category could be found from Table 3b

Regression Model-1

(Total hired labour hour)$_i$ = $\alpha + \beta_1$ (firm's net income in rupees, representing firm's ability) + β_2 (value of asset in rupees, representing requirement of hired labour) + β_3 (total family labour hour, as a probable substitute for hired labour) + β_4 (cost of raw material in rupees, representing requirement of hired labour) + β_5 (Age of unit in month) + β_6 (rent including imputed rent in rupees – proxy for workspace) + β_7 (experience of owner in years) + μ_i

We test for normality, multicollinearity and heteroskedasticity

See Table 4.

The above regression (Table 4: 1-a on Asansol) shows that, while, strikingly, net income, volume of assets, age of the units and experience of the owner the use of hired labour, rent (representing workspace), volume of raw material and participation of family labour play significant roles. Thus, hired labour is used to assist family labour to supplement (instead of substituting) the work effort of the owner and it is used when the workspace is larger and there is higher amount of raw material to work with; but it is not guided by the size of the firm in terms of income and asset.

However, for Bolpur (regression 1-b), it is the income and asset along with rent and raw material, which are found to be significant explanatory variables; family labour along with age of unit and experience of owner is not significant.

Table 4 Result for regression-1

Regression No.: 1		
Independent Variables	(a) Dependent variable: **Total hired labour hour** Number of obs.—69 Prob > F = 0.0000 Adjusted-R^2 = 0.59 **Asansol**	(b) Dependent variable: **Total hired labour hour** Number of obs.—73 Prob > F = 0.0000 Adjusted-R^2 = 0.70 **Bolpur**
Net income	−1.92 (2.39)	4.02** (1.24)
Assets (Machineries)	0.602 (0.44)	1.34* (0.57)
Total family labour hour	0.58* (0.30)	−0.41 (0.35)
Monthly cost of raw materials	0.003*** (0.00)	0.003*** (0.00)
Age of unit	−0.128 (0.284)	0.24 (0.266)
Rent	17.73** (7.37)	25.18** (10.80)
Experience of owner	−0.318 (3.94)	2.87 (0.36)
_cons	179.80* (97.53)	62.09 (99.94)

Source Authors' calculation from field survey data
* 10% significant level. ** 5% significant level. *** 1% significant level

Thus, in Bolpur, the larger firms are able to hire labour, whereas, in Asansol, perhaps, the volume of activity is inducing the firms to use hired labour supplementing family labour.

3.2 Market Dependence

Where and how our sample firms operate for input acquiring and sale of output are essential for them to acquire and retain surplus (if any). Hence, we record firm participation in a variety of market arrangements and map its probable influences on certain traits of these firms (volume of activity and net income).

In order to map the participation of firms across a variety of market arrangements for inputs and outputs, we prepare Table 5. We find that all the firms of Asansol purchase inputs from open market, but only 44 units sell their final outputs to individual buyers and the rest 29 units are involved in sub-contracting. The contractors of Asansol only deal with final outputs without providing any financial and/or raw-material support. The sub-contracted firm owners arrange for raw materials on their own and sell the final output to their respective contractors. The contracts, although verbal, restrict the firms from selling their products elsewhere. Contrarily, almost all the units of Bolpur operate in open (input-output) markets, where inputs are bought from wholesalers and outputs are sold to individual buyers.

On the other hand, the handloom units of Dhaniakhali operate majorly under a cooperative. In Tantipara, 22 units fully operate under Mahajans[8] and the rest purchase raw material from *Aratdars*. Weavers sell or supply the final product to local traders (who buy the product on spot without any pre-contracting) and to contractors

Table 5 Destination for sale of output and source of inputs for the sample firms

Locations	Input source	Output destination		
		Individual buyers and local traders (Tantipara)	Contractor (Asansol and Tantipara), cooperative (Dhaniakhali) and local traders (Tantipara)	Total
Asansol	Open market	44	29	73
Bolpur	Open market	73	2	75
Dhaniakhali	Cooperative	0	79	79
Tantipara	Aratdars	34	19	75
	Mahajans	0	22	

Source Authors' compilation from field survey data

[8]Long-term contractual relationship between the contractors/Mahajans and the weavers over generations having hierarchical power relations (biased in favour of contractors) with both beneficial and exploitative elements.

(who enter into a contract with the weavers before production, but do not provide any financial support to buy raw materials, as in Asansol).

From Table 6, we see that, in Asansol, the units who are capable of getting engaged in a sub-contracting arrangement with large-buyers/showrooms/real estate business, are able to have access to larger market and hence, generate sizeable revenue (compared to the other group selling output to individual buyers). However, the difference (between these two groups of firms) reduces noticeably, if we consider net income. This may be because of the fact that although the sub-contractors are able to get assured market through this contracting arrangement, a large part of their surplus is appropriated by the contractors. While the ratio of revenue to total cost (raw material plus labour) is 1.27 for the sub-contracted sample units, for the other group, it is 1.33; clearly showing siphoning of surplus through sub-contracting.

In Bolpur, because of tourism and real estate potentials, the average market size (per firm) is moderate and the firms are able to appropriate a revenue greater than that of the comparable group of Asansol (44 units—who sell their output to individual buyers, as in Bolpur). But raw material and labour costs are also considerably high, which are reducing the net income by a large amount. Despite spending much higher amount on raw material and labour, the firms of Bolpur are unable to generate even equal amount of net income vis-à-vis that of Asansol. The average price (of output)[9] received by the firms of Bolpur is found to be less than that of Asansol and raw material as well as labour costs are much higher, as we have already seen in Tables 1 and 3a and also could be calculated from Table 6. However, this requires further probe, as a variety of inputs are used and a variety of products are produced, which are not always comparable across locations.

Our above comparisons within Asansol and across Bolpur and Asansol show that there is not much difference so far as income is concerned across different market arrangements; open market access rather than sub-contracting arrangement is not changing the outcome. This observation is strengthened by the following two regressions (regression model-2, Table 7), where labour and raw material are found to be significant explanatory variables for net income, but variation in market structure does not play a significant role.

We set an OLS regression model (Regression model-2)

Regression Model-2

Log (Net Income in rupees)$_i = \alpha + \beta_1$ Log (total labour hour) $+ \beta_2$ Log (value of asset in rupees) $+ \beta_3$ Log (cost of raw material in rupees) $+ \beta_4$ Log (rent including imputed rent in rupees) $+ \beta_5$ (Output market dummy) $+ \mu_i$

Output Market Dummy $= 0$; if sold to contractor; and

[9]Grill—$4' \times 4'$: Asansol (Rs. 60/kg × 30 kg) Rs. 1800; Bolpur (Rs. 58/kg × 27.5) Rs. 1595
Sheet Gate—$6' \times 3'$: Asansol (Rs. 65/kg × 70 kg) Rs. 5000; Bolpur (Rs. 62/kg × 60) Rs. 3720
Almirah—36 gauge: Asansol (Rs. 7500–8000 per piece); Bolpur (Rs. 6500–7000 per piece)

Table 6 Mean values for certain variables capturing firm-traits across various combinations of input source and output destination

Locations	Market relations (Input–Output)	No. of sample units	Total revenue (Rs./Month)	Total raw material cost (Rs./Month)	Total hired labour cost (Rs./Month)	Net income (Rs./Month)
Asansol	Open Market (Input)–Individual buyers (Output)	44	120,364.4	78,478.44	16,307.95	23,949.1
	Open Market (Input)–sub-contracted (Output)	29	166,038.1	110,190.2	26,211	26,912.8
Bolpur	Open Market (Input)–Individual buyers (Output)	73[a]	148,384	95,395.25	23,415.75	23,669.18
Dhaniakhali	Cooperative–Cooperative	79	2956.68	213.70	–	2647.36
Tantipara	Artadars (Input)–Local Traders (Output)	34	22,296.76	15,438	–	6607.31
	Aratdars (Input)–sub-contracted (Output)	19	16,262.61	7443	–	8576.94
	Mahajans–Mahajans	22	5736.46	–	–	5472.07

Source Authors' compilation from field survey data
[a]Two firms participating in sub-contracting relationship are not included

Table 7 Result for regression-2

Regression No.: 1		
Independent variables	(a) Dependent variable**: Log (Net income)** Number of obs.—140 Prob > $F = 0.0000$ Adjusted-$R^2 = 0.53$ **All fabrication units**	(b) Dependent variable**: Log (Net income)** Number of obs.—70 Prob > $F = 0.0000$ Adjusted-$R^2 = 0.47$ **Asansol only**
Log (total labour hour)	0.279** (0.116)	0.164 (0.173)
Log [value of assets (machineries)]	−0.165 (0.073)	0.004 (0.092)
Log (monthly cost of raw materials)	0.57*** (0.079)	0.558*** (0.107)
Log (rent)	−0.027 (0.060)	−0.0621 (0.087)
Output market dummy	−0.011 (0.124)	0.055 (0.161)
_cons	−5.10*** (0.665)	−4.28*** (0.971)

Source Authors' calculation from field survey data
** 5% significant level. *** 1% significant level

= 1; if sold to individual buyer (open market)

We test for normality, multicollinearity and heteroskedasticity

In Tantipara, the units operating under Mahajans earn the least; those who purchase inputs from *Aratdars* and sell to contractors (output-dealers), earn the maximum; in between, lie those who buy from *Aratdars* and sell to local traders. If we compare between the last two groups, the latter puts more than double in terms of raw material (Rs. 15,438 vs. Rs. 7443) but curiously earns less than the former. Thus, the traders are appropriating a large amount of surplus. This latter group of weavers cannot get engaged with the contractors (because of lack of network) and hence are compelled to sell to the traders. Strikingly, a kind of tied-production is in fact giving higher return to the weavers. However, bonding with Mahajan is least remunerative.

In Dhaniakhali, despite support of cooperative the condition of weavers is miserable. But, essentially, this is because of sudden breach of contract from the government and government failure in protecting the Dhaniakhali brand from fake products. Further, despite various efforts, the cooperative is unable to expand its market size because of competition from machine-made products and duplicate items sold with the brand name of Dhaniakhali.

Summarising, for most of our sample firms, a variety of hierarchical market relations and many other types of market problems may be acting as serious bottlenecks for firm-progression. Most of the firms are unable to negotiate with the market on equal terms and are compelled to behave non-capitalistically, as we see below.

3.3 Accumulation or Survival?

Now, we try to see, whether our sample firm owners have any tendency to invest in their respective businesses or they mostly consume and save their income. For this, we tabulate net firm-income, income from other source/s and distribution of expenditure of the owner households (Table 8a and 8b). We also introduce two figures (Fig. 1) for yearly investment for Bolpur and Dhaniakhali.

The striking observations are:

(a) a major part of firm-income and income from other source/s is spent on food, in all the locations;
(b) another major part goes for savings;
(c) as revenue rises, share of expenditure on food falls, but share of savings rises almost proportionately;
(d) however, intriguingly, there is almost no fresh investment in Asansol and Tantipara with only some fresh investment in Bolpur and very little in Dhaniakhali;
(e) in Dhaniakhali and Bolpur, investment rises slightly with rise in revenue;
(f) in Dhaniakhali, it is mainly an expenditure on fresh designs and hence, on loom accessories;
(g) although, investment is very low or even zero in all the locations, there is some expenditure as depreciation cost;
(h) in Bolpur, majority of the firms (45) do not invest at all; there are only few firms with sizeable investment;
(i) in Dhaniakhali, almost all spend small amount on loom accessories related to design.

As we have mentioned above, in all the four locations, share of expenditure on food and non-food decreases, but share of savings increases (although that of investment does not), as we move upward across revenue categories. Strikingly, a firm is behaving like a household. This rising savings propensity may be a mark of a desire of the firm owners to grow and lead a better life in future. This is also supported by the following opinion of the firm owners.

The firms of Bolpur and Asansol were asked about their intentions, if they are hypothetically handed over a lump-sum amount. In this context, many of the owners, especially of Bolpur, opined that they could use this fund for stocking of raw material. However, in Asansol, many responded in favour of health expenses, as the production process and overall environment are hazardous.

The monthly net income of the sample units of Dhaniakhali and Tantipara is too low to survive. Over and above consuming the earnings from firm, the firm owners participate in other petty activities; hence, almost no question of fresh investment arises (barring an essential expenditure related to design, in Dhaniakhali). In the context of hypothetical provisioning of lump-sum amount as mentioned above, most of the weavers of the two locations responded to save the full amount for future contingencies.

Table 8 (a) Monthly incomes and distribution of expenditure for the owner households across expenditure categories over revenue class and locations. (b) Monthly net firm-income and distribution of expenditure for the owner households across expenditure categories over revenue category and locations

	Revenue category	Avg. net firm-income (Rs./Month)	Income from other source/s (Rs./Month)	Food (%)	Transport (%)	Education (%)	Medical (%)	Savings (%)	Depreciation (%)	Investment (%)
(a)										
Asansol										
Revenue category	Category 1 (13 firms)	6530.35	6471 (8 firms)	77.9	2.5	11.4	6.5	0.0	1.7	0.0
	Category 2 (25 firms)	15,785.93	5307 (8 firms)	50.44	5.16	6.30	3.27	32.14	2.66	0.03
	Category 3 (35 firms)	38,750.98	36,429 (7 firms)	25.1	3.2	5.0	3.7	62.4	0.5	0.0
Bolpur										
Revenue category	Category 1 (6 firms)	7976	8667 (3 firms)	45.48	3.44	8.73	4.91	35.93	0.41	1.09
	Category 2 (25 firms)	10,661.25	10,525 (13 firms)	42.31	5.29	4.56	5.13	39.38	1.57	1.75
	Category 3 (43 firms)	35,134.98	20,800 (12 firms)	25.0	4.0	5.0	3.6	60.3	1.1	1.0
(b)										
Dhaniakhali										
Revenue category	Category 1 (20 Firms)	855.66	3877 (20 firms)	67.9	0.0	4.1	6.5	19.7	1.3	0.5

(continued)

Table 8 (continued)

	Revenue category	Avg. net firm-income (Rs./Month)	Income from other source/s (Rs./Month)	Food (%)	Transport (%)	Education (%)	Medical (%)	Savings (%)	Depreciation (%)	Investment (%)
	Category 2 (20 firms)	1582.18	5306 (18 firms)	52.4	0.0	2.0	8.5	35.2	1.2	0.6
	Category 3 (22 Firms)	2670.39	2557 (21 firms)	67.06	0.00	4.31	8.32	18.23	1.26	0.83
	Category 4 (17 Firms)	5978.59	1839 (15 firms)	53.87	0.07	6.32	5.30	32.10	1.01	1.34
Tantipara										
Revenue category	Category 1 (20 firms)	4481.35	3638 (12 firms)	82.6	0.4	4.7	11.3	0.0	1.8	0.0
	Category 2 (22 firms)	6441.58	3583 (9 firms)	58.8	0.3	5.3	5.0	28.5	2.2	0.0
	Category 3 (19 firms)	6758.88	4604 (6 firms)	74.5	0.0	6.1	7.4	11.2	0.7	0.0
	Category 4 (14 firms)	10,588.23	6833 (4 firms)	67.6	0.0	6.0	6.6	18.2	1.5	0.0

Source Authors' compilation from field survey data

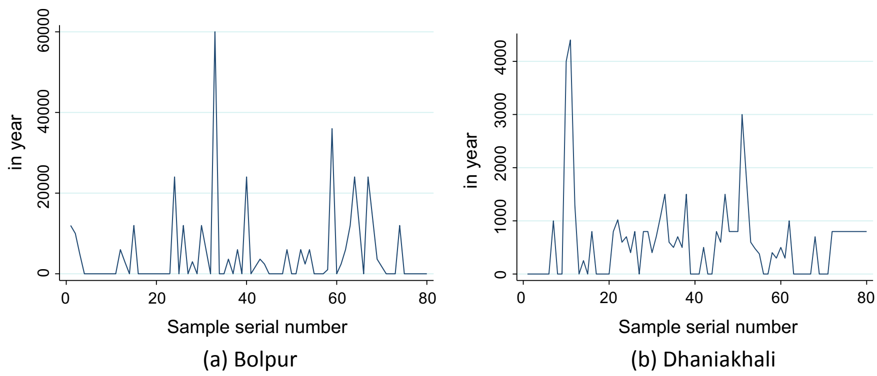


Fig. 1 Yearly investment (Rs.) across firms of Bolpur and Dhaniakhali

4 Conclusion

In this paper, we have tried to understand, using quantitative as also qualitative information on a variety of unorganised/informal manufacturing units engaged in iron forging and fabrication (in urban locations of West Bengal: Asansol and Bolpur) and in handloom weaving (in rural locations of West Bengal: Dhaniakhali and Tantipara), whether the firms have the potential to survive and grow as micro-capital or they are just subsisting and even decaying as non-capitalistic petty firms. We have tried to understand these tendencies focusing on various aspects like production organisation, especially owner–labourer relation, market structure, especially hierarchies and bottlenecks and on abilities/intentions of the firms to survive/grow. We have also considered certain firm-characteristics as well as socio-economic-cultural features of the surveyed locations to analyse how firm behaviours (as micro-capital or non-capital) are influenced by these firm-traits and environment.

We have arrived at the following interesting observations:

(a) Many of the firms do not use hired labour and more importantly, those who use hired labour, do not exhibit strict owner–labourer separation/dichotomy in terms of pattern and intensity of work, skill and even income (to some extent). Along with hierarchy of owner in organising production, there is considerable participation of worker too. Thus, although many firms hire labour, there is absence of strict capital–labour hierarchy and hence, most of these firms could be considered as non-capitalist in nature.

(b) Most of the firms work under intermediaries (of one form or the other) and a part of surplus is appropriated by these intermediaries. Some firms are free from direct intermediary intervention, but they suffer from a variety of market problems like inadequate market size and price, increasing market competition, flooding of markets by machine-made and fake products and increasing price of raw material and labour. Thus, the firms suffer from their acute market dependence; they have to submit to the inadequacies and fluctuations of the market.

(c) Even if some firms hire labour, participate in somewhat open market, generate some surplus, spend large amount on raw material and also some amount against depreciation, overwhelming majority do not invest by large amount. This is a fundamental observation in the sense that, despite showing some apparent potential for progression (as micro-capital) in terms of hiring of labour and participation in open market, the firms are either not able or are not willing to invest—they are more interested in savings for future contingencies; they are mostly behaving like petty non-capital, more concerned about their persistence than progression.

Thus, invoking the literature, we could say that although a few firms may be behaving like dynamic micro-capital having symptoms of progression and some control over markets, overwhelming majority, without much of owner–worker separation and almost no control over markets, is able to and/or interested in maintaining their existence only.

Acknowledgments Pratip Kumar Datta: The author takes this opportunity to express his gratitude to Professor Sarmila Banerjee who has not only been a great teacher but also a great source of inspiration.
Saumya Chakrabarti: The author is thankful to the editors for giving him a precious scope to express his deep gratitude to Professor Sarmila Banerjee who not only taught him the fundamentals of firm and market and helped him to enter into the world of political economy of labour but also guided him to get a humble place in academia. The author thanks his colleagues of the Department of Economics and Politics and of the administration at Visva-Bharati, Santiniketan, for allowing him Sabbatical Leave to work on the issue.
The authors are indebted to Rishav Mukherjee, Arindam Saha, Sandip Mahato, Mithun Singha Roy, Debangee Karmakar, Jhinuk Banerjee, Jayati Chatterjee, Durba Ahmed, Sonali Das and Rajkumar Kundu for a variety of research support. Rishav Mukherjee helped the authors tremendously, not only in administrative works, but also with his academic inputs. Comments from especially Amit Bhaduri, and also from Tirthankar Roy, Malabika Roy, Simantini Das, Bratati Chattopadhyay, Santadas Ghosh, Achiranshu Acharyya, Biswajit Mandal and Suman Ray have been beneficial. Financial help from ICSSR (GoI; F.No. 02/210/2016-17/RP, dated 30.03.2017) and UGC-SAP (Department of Economics and Politics, Visva-Bharati, Santiniketan) are gratefully acknowledged. The authors are thankful to the following persons for logistic support: Amit Mukherjee, Rathin Nag, Ansari, SentuSk, RentuSk, Bharat Das, Malay Ray, Saju Bagdi, Gadadhar Hazra, Dayamoy Dutta, Pritikana Char and Subrata Mukherjee. They are also grateful to the following government and non-government offices: Directorate of handloom & handicrafts—Birbhum, Tantipara Panchayat, Dhaniakhali Tant Cooperative Society, etc. Authors acknowledge Rathin Nag, Ganesh Shaw, Ansari, SentuSk, RentuSk, Bharat Das, Mohan Das, Tarak Das, Malay Ray, Gadadhar Hazra and Dayamoy Dutta for participation in the FGDs and personal interviews, and are grateful to all the respondents of the primary survey. The authors are also indebted to the referee for substantive comments, which have induced them to have a deeper look at the issue. However, the usual disclaimer applies.

References

Bardhan, P. (2009). Notes on the political economy of India's tortuous transition. *Economic and Political Weekly, 44*(49), 31–36. Retrieved from http://www.jstor.org/stable/25663859.

Basole, A., Basu, D., & Bhattacharya, R. (2015, January 19). Determinants and Impacts of Subcontracting: Evidence from India's unorganised manufacturing sector. *International Review of Applied Economics*.

Bhattacharya, R., Bhattacharya, S., & Sanyal, K. (2013). Dualism in the informal economy: Exploring the Indian Informal Manufacturing Sector. In S. Banerjee & A. Chakrabarti (Eds.), *Development and sustainability* (pp. 339–362). Berlin: Springer.

Bhattacharya, S., & Kesar, S. (2018). Possibilities of transformation: The Informal Sector of India. *Review of Radical Political,* 1–9.

Breman, J. (2013). *At work in the Informal Economy of India*. New Delhi: Oxford University Press.

Chakrabarti, S. (2016). *Inclusive growth and social change: Formal-informal-Agrarian relations in India*. New Delhi: Oxford University Press.

Chatterjee, P. (2008, April 19). Democracy & economic transformation in India. *Economic & Political Weekly*, 53–62.

Chen, M. A. (2007). *Rethinking the informal economy: Linkages with the formal economy & the formal regulatory environment*. DESA Working Paper No. 46, ST/ESA/2007/DWP/46.

De Soto, H. (2000). *The mystery of capital*. New York: Basic Books.

Khasnabis, R., & Nag, P. (2001, December 24). Labour process in the informal sector—A Handloom Industry in Nadia District. *Economic & Political Weekly*, 4836–4845.

Kundu, A. (1993). Urban informal sector: An overview of macro dimensions, socio-economic correlates and policy perspectives in India. In H. Nagamine (Ed.), *Urban informal sector employment—A cross national comparative study*. Nagoya: Research Institute for Regional Planning and Development.

Kundu, A., & Chakrabarti, S. (2008, September). Some micro-economic aspects of informal sector enterprises: Field-based evidence from West Bengal. *ArthaVijana, L*(3), 209–238.

Kundu, A., Sarangi, N., & Das, B. P. (2005). Economic growth, poverty & non-farm employment: An analysis of rural-urban interlinkages. In Nayyar & Sharma (Eds.), *Rural transformation of India: The role of non-farm sector*. IHD.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives, 28*(3), 109–126.

Liebl, M., & Roy, T. (2003, December 27). Handmade in India/preliminary analysis of crafts producers and crafts production. *Economic & Political Weekly*, 5366–5376.

Marjit, S., & Kar, S. (2011). *The outsiders: Economic reform and informal labour in a developing economy*. New Delhi: Oxford University Press.

Mezzadri, A. (2010). Globalisation, informalisation and the state in the Indian Garment Industry. *International Review of Sociology, 20*(3), 491–511.

Mitra, A. (1994). *Industry,* informal sector employment and poverty. *Indian Journal of Labour Economics, 37*(3).

Mitra, A. (1998). Employment in informal sector. *Indian Journal of Labour Economics, 41*(3).

Mitra, A., & Mitra, S. (2005). *Rural non-farm sector: A state-level profile*. In Nayyar & Sharma (Eds.), Rural transformation in India: The role of non-farm sector (pp. 121–136). New Delhi: Institute for Human Development.

Moser, C. N. (1978). Informal sector or petty commodity production: Dualism or independence in urban development. *World Development, 6,* 1041–1064.

Mukherjee, D. (2004a). Informal manufacturing sector in India: Pre and post reform growth dynamics. *Indian Journal of Labour Economics, 47*(2).

Mukherjee, D. (2004b). Problems and prospects of informal manufacturing sector: A case study of Durgapur City. *Indian Journal of Regional Science, 36*(2), 2004.

NSSO. (2001). National Sample Survey Office, Government of India, Report No 55/2.0, 459.

NSSO. (2017). National Sample Survey Office, Government of India, Report No 73/2.34.

Raj, R., & Sen, K. (2016). *Out of the Shadows? The informal sector in post-reform India.* New Delhi: Oxford University Press.

Ranis, G., & Stewart, F. (1993). Rural non-agricultural activities in development: Theory and application. *Journal of Development Economics, 40*(1), 75–101.

Sanyal, K. (2007). *Rethinking capitalist development: Primitive accumulation*. New Delhi, Routledge: Governmentality & Post-Colonial Capitalism.

Group Formation and Endogenous Information Collection in Microcredit

Sukanta Bhattacharya and Shirsendu Mukherjee

Abstract This paper attempts to address the effects of different types of loan contract on a borrower's incentive for investment in information. We model the trade-off that a borrower faces when she collects information about the potential of her intended projects both under individual and joint liability loan contracts. Even under limited liability, the borrower faces a trade-off at information collection stage between the cost of signal collection, and the cost of her time and effort for project execution in case the project fails. We show that joint liability contract induces borrowers to invest more in information than individual liability for low rates of interest. However, for some high rates of interest, borrowers invest positive amount in information collection under individual liability, but do not take up the project under joint liability.

Keywords Joint liability · Group lending · Moral hazard · Peer monitoring · Group size · Social sanction

JEL Classification G21 · D82 · O16

1 Introduction

Before undertaking a project, borrowers generally collect information about the potential of their intended projects. Undertaking a project requires the borrower to put in effort and time besides the borrowed investment which are costly and are sunk in case of project failure. Even under limited liability—where an unsuccessful borrower doesn't have to repay her loan—information about the chance of project success is welcome since this allows the borrower to save time and effort cost if she finds out that the project will fail almost surely. Moreover, this also has implications about the lender's profitability and the equilibrium rate of interest. If the borrowers only

S. Bhattacharya (✉)
Department of Economics, University of Calcutta, 56A B.T. Road, Kolkata 700050, India
e-mail: bsukanta@gmail.com

S. Mukherjee
Department of Economics, St. Paul's C. M. College, Kolkata, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_8

undertake projects when they get a good signal about project success, the lender loses much less borrowed capital on failed projects. This allows the lenders to lower the rate of interest generally leading to higher efficiency in the microcredit market.

However, collecting information is costly for the borrowers and this cost rises with the quality of information. The trade-off a borrower faces at the information collection stage is between the cost of signal collection and her potential saving of time and effort cost in case the realized signal indicates strong probability of failure. This chapter attempts to understand this trade-off in terms of a simple model of credit market. More specifically, we examine how individual lending mechanism vis-a-vis the group-based lending mechanism affects the ex-ante incentive for information collection and the welfare of the agents.

The last point needs some clarification here. The group may be formed before the borrowers collect the information. In this case, the group members may cooperatively decide on the quality of information collected by the members. Any deviation from this cooperatively chosen quality is immediately detected and punished by the group. Alternatively, a group may be formed after the borrowers independently decide on the quality of information they would collect, but before the signal is realized. Notice that group formation at this stage does not leave any scope for group punishment based on the quality of information. Finally, the group may be formed after the signal realization. In the last two cases, there will be an assortative matching in the group formation stage though quality of information is chosen independently by each borrower. But the knowledge that one can form a group with a borrower with good-quality signal only if the first borrower herself has good-quality information may create extra incentive to collect good-quality information for the borrowers.

In the large body of microfinance literature, a significant amount has been devoted to explore the potential of group lending under joint liability. Most of the existing literature on group-lending with joint liability attempts to address the main obstacles behind the non-viability of the credit support system for the poor and then tries to resolve the problems. A number of theoretical models in the existing literature of microfinance have identified various mechanisms including peer screening or assortative matching (Ghatak 1999; Ghatak and Guinnane 1999; Tassel 1999; Morduch and Aghion 2004a, b), peer monitoring (Stiglitz 1990; Aghion 1999; Conning 1999), and peer pressure or enforcement (Besley and Coate 1995; Morduch and Aghion 2004a, b) through which group lending with joint liability enables a lender to make uncollateralized loans even to the marginal people (Mukherjee and Bhattacharya 2014). Actually, shifting a burden of default from a lender to a group gives correct incentives to use their local information and social ties for ensuring repayments of peers within the same group (Tsukada 2012).

Microfinance markets reveal that limited borrower liability exposes lenders to levels of adverse selection and moral hazard which rely on formal collateral. The use of joint liability contracts for those borrowers who take group loans (with joint liability) creates an intricate strategic dynamic between groups and lenders, each of whom bear some risk in the extension of loans to individual members (Janvry et al. 2010). While group lending with joint liability has become hugely popular as an instrument for overcoming difficulties of rural credit markets, there are other mechanisms like *savings-up* mechanism, *sequential lending* mechanism, *direct lender monitoring*, *dynamic incentives*, *contingent renewal* and *progressive lending* mechanisms, *joint benefit* mechanism etc. (Sinha 2005; Roy Chowdhury 2005, 2007; Ghosh and Ray 1997; Bhattacharya et al. 2008) which microlenders use in practice, often in conjunction with the joint liability mechanism.

In a framework that allows project returns to accrue over time, Chowdhury et al. (2014), in a recent paper, provide a justification for the use of frequent repayment schemes, examine the optimal choices for the MFI, demonstrate that the MFI opts for higher project sizes under group lending with limited collusion, and also provide a theory on group size. Mukherjee and Bhattacharya (2014, 2015), have dealt with that issue of optimal group size, optimal joint liability parameter, and socially efficient group size. Unfortunately, the same literature on microcredit is silent on the aspects of endogenous information collection and group formation in the microcredit market. It attempts to examine how individual lending mechanism vis-a-vis the group lending mechanism affects the ex-ante incentive for information collection and the welfare of the agents when the group is formed after the signal realization.

It is interesting to note that the existing literature on microcredit is almost silent on the issues of endogenous information collection and group formation in the microcredit market. This paper attempts to provide a theoretical framework to analyze these issues. Borrowers (even with limited liability individual lending) generally collect information about the potential of their intended projects since undertaking a project requires the borrower to put in effort, time and money (borrowed investment) which are costly and are sunk in case of project failure. Information about the chance of project success allows the borrower to save these costs if she finds out that the project will fail almost surely. This also allows the lenders to lower the rate of interest, generally leading to higher efficiency in the microcredit market.

2 The Model

We consider a simple model. Output (Y) takes two values: high (Y^{H}) and low (Y^{L}), where $Y^{\mathrm{H}} > Y^{\mathrm{L}} \geq 0$. For simplicity, we normalize Y^{L} to 0 and denote Y^{H} by Y. Projects are indivisible, and each project requires project fund of amount 1 to be viable. For each project, the prior probability of Y being realized is $\frac{1}{2}$ i.e., $\Pr\left(Y^{\mathrm{H}} = Y\right) = \Pr\left(Y^{\mathrm{L}} = 0\right) = \frac{1}{2}$.

The lending institution has the resources to lend to a number of borrowers. The borrowers face limited liability in the sense that, in the case of default, the lender cannot seize assets that a borrower has not specifically pledged as collateral. In this context, we assume that the poor has no collateral to pledge. Hence, the lender's receipt in the event of default is zero. The limited liability constraint, along with the borrowers' lack of collateral, rule out the standard instruments used by conventional lenders to overcome information and enforcement problems.

We model the information structure as follows. The borrower collects a costly signal about the potential of the project. The quality of signal is characterized by $\theta \in \left[\frac{1}{2}, 1\right]$. $\theta = \frac{1}{2}$ indicates a completely uninformative signal, while $\theta = 1$ indicates a fully informative signal. The signal, denoted by s, has two possible realizations: $s \in \{S, F\}$, where S and F indicate success and failure, respectively. We assume that $\Pr\left[S|Y^{\mathrm{H}}\right] = \Pr\left[F|Y^{\mathrm{L}}\right] = \theta$. The cost of collecting a signal of quality θ be given by $c(\theta)$. We assume that $c'(.) > 0, c''(.) > 0$ and $c\left(\frac{1}{2}\right) = 0$. We also assume that $c'(1) > Y$ and $c'\left(\frac{1}{2}\right) < \frac{e}{2}$ to ensure that an interior solution for signal quality (θ) always exists.

Notice that if a borrower collects a signal of quality θ, then the probability of getting a Success (S) signal is

$$\Pr[S] = \Pr\left(S|Y^{\mathrm{H}}\right).\Pr\left(Y^{\mathrm{H}}\right) + \Pr\left(S|Y^{\mathrm{L}}\right).\Pr\left(Y^{\mathrm{L}}\right) = \theta.\frac{1}{2} + (1-\theta).\frac{1}{2} = \frac{1}{2}$$

Similarly, $\Pr[F] = \frac{1}{2}$. The posterior beliefs about the success of the project and realizing output Y, if the signal is S, is given by

$$\Pr(Y \mid S) = \frac{\Pr(S \mid Y).\Pr(Y)}{\Pr(S \mid Y).\Pr(Y) + \Pr(S \mid 0).\Pr(0)} = \frac{\theta.\frac{1}{2}}{\theta.\frac{1}{2} + (1-\theta)\frac{1}{2}} = \theta$$

and

$$\Pr(Y \mid F) = \frac{\Pr(F \mid Y).\Pr(Y)}{\Pr(F \mid Y).\Pr(Y) + \Pr(F \mid 0).\Pr(0)} = \frac{(1-\theta).\frac{1}{2}}{(1-\theta).\frac{1}{2} + \theta.\frac{1}{2}} = (1-\theta)$$

Finally, we assume that if the borrower implements a project, she has to incur a fixed effort cost, $e > 0$. In our model, e plays a very vital role. This is what the borrowers hope to save by collecting information. If $e = 0$, then each borrower would choose $\theta = \frac{1}{2}$. Higher the value of e, the more incentive the borrowers have to collect better-quality information (i.e., higher θ). Throughout, we assume that $(Y > 2e)$. This assumption is necessary to ensure a positive expected value of the project ex-ante.

We consider individual lending as well as group lending. For the sake of analytical simplicity, we limit the group size to 2, i.e., $n = 2$. In individual lending (IL) as well as group lending (GL), a borrower's decision about taking up the project is taken only after the realization of the signal. We assume that the borrowers do not default voluntarily, i.e., they take up the loan only if they decide to take up the project. In the case of group lending, the members of the group get access to the loan only if both members decide to take up the project.

We assume that group lending entails joint liability. Our focus in this paper is on group formation and how group formation affects the incentive for signal collection. We consider the information structure where the community members can observe each other's effort in information collection (θ or $c(\theta)$) as well as the signal realizations (s), i.e., θ and s are publicly observable within the community.[1]

The timing of the game is as follows:

1. *Stage* 1: The lender announces the gross rate of interest ($r > 1$) that is to be charged on a loan of amount 1.
2. *Stage* 2: Each borrower chooses her quality of signal/information θ at cost $c(\theta)$.
3. *Stage* 3: The signal (s) is realized, i.e. the borrower observes either S or F.
4. *Stage* 4: After observing her signal, the borrower decides whether to take up the project or not. If she decides to take up the project, she applies for loan. In case of group lending, both the partners must decide to take up the loan.[2]
5. *Stage* 5: If the borrower takes the loan, she incurs the effort cost e and the project outcome is realized.

In joint liability loan contract, we consider groups of two people ($n = 2$), with each group formed voluntarily. Individuals invest independently, but the loan contract involves joint liability. Under the joint liability contract, a borrower pays her own interest (r) in the event of her success, but pays an additional interest payment in case her partner fails. We assume that $\left(r < \frac{Y}{2}\right)$, i.e., a successful borrower has sufficiently large pay-off to pay even for her unsuccessful partner.

2.1 Limited Liability Individual Lending

We first consider individual liability lending. The borrower, as we know, takes the decision about the project after the signal is realized.

If an individual borrower takes up the project after realization of the success signal (S), her expected utility will be,

$$\mathrm{EU}^1(\theta|S) = \theta(Y - r) - e - c(\theta)$$

If she doesn't, then her expected utility will be,

---

[1]It is possible to consider other possible information structures. For example, we may consider the case when s is not publicly observable, but the quality of signal (θ) is. In a richer model, we may take up the case when neither s nor θ are publicly observable. In each of these cases, we can examine how the process of group formation is affected by the information structure. It will be further interesting to investigate the incentive for investment in quality of signal in all these three cases, and to observe how this incentive stands in comparison with individual liability.

[2]Notice that a borrower's decision about taking up the project depends on the signal realization. So when the group is formed and signals are not observable, even if a borrower receives an S signal, she may not have access to loan if her partner receives an F and decides against loan in a two member group. This is an inefficiency that arises when only group-based contracts are offered. A richer model should take this into account and allow the lender to offer both type of contracts at the same time. However, in this chapter, we analyze individual lending and group lending separately. That means that we do not allow the lender to offer both individual contract and group-based contract at a time.

$$\mathrm{EU}^0(\theta|S) = -c(\theta)$$

Hence, she takes up the project after S, iff

$$\mathrm{EU}^1(\theta|S) \geq \mathrm{EU}^0(\theta|S)$$

$$\Leftrightarrow \theta(Y-r) - e - c(\theta) \geq -c(\theta) \Leftrightarrow \theta \geq \left(\frac{e}{Y-r}\right) \tag{1}$$

Similarly, she takes up the project after F, iff

$$\mathrm{EU}^1(\theta|F) \geq \mathrm{EU}^0(\theta|F)$$

$$\Leftrightarrow (1-\theta)(Y-r) - e - c(\theta) \geq -c(\theta) \Leftrightarrow (1-\theta) \geq \left(\frac{e}{Y-r}\right) \tag{2}$$

Notice that for

$$r > (Y-e)$$

neither (1) nor (2) is satisfied for any $\theta \in \left[\frac{1}{2}, 1\right]$.[3] Hence, for $r > (Y-e)$, an individual borrower would not take up the project even after a success signal. Since the project is never taken up, the borrower would not invest any amount in collecting signals and thus $\theta^* = \frac{1}{2}$.

Now consider $r \in (Y-2e, Y-e)$. In this case

$$\left(\frac{e}{Y-r}\right) \in (\frac{1}{2}, 1].$$

Since for any $\theta \in \left[\frac{1}{2}, 1\right]$, $1-\theta \leq \frac{1}{2} < \frac{e}{Y-r}$, (2) can never be satisfied for these values of r. Thus, a borrower would never take a loan after F if $r \in (Y-2e, Y-e]$. However, the borrower may decide to take up the project if the realized signal of quality $\theta \geq \left(\frac{e}{Y-r}\right)$ is S . We know that ex-ante the probability of a Success signal (S) is $\frac{1}{2}$. Hence, for $r \in (Y-2e, Y-e]$, the borrower's ex-ante expected utility is

$$\mathrm{EU}(\theta) = \begin{cases} -c(\theta) & \text{if } \theta < \frac{e}{Y-r} \\ \frac{1}{2}[\theta(Y-r) - e] - c(\theta) & \text{if } \theta \geq \frac{e}{Y-r} \end{cases} \tag{3}$$

The borrower chooses θ to maximize (3). Notice that if the borrower chooses $\theta^* = \frac{1}{2}$, her expected utility is 0. Suppose

[3] Since,

$$r > (Y-e) \Rightarrow e > (Y-r) \Rightarrow \left(\frac{e}{Y-r}\right) > 1.$$

$$\hat{\theta} = \arg\max \left\{ \frac{1}{2} \left[\theta \left(Y - r \right) - e \right] - c \left(\theta \right) \right\}$$

Given our assumptions on $c(.)$, $\hat{\theta} \in \left(\frac{1}{2}, 1\right)$ and is unique. $\hat{\theta}$ is determined by the first order condition (FOC) of optimization

$$\frac{1}{2} \left(Y - r \right) - c' \left(\hat{\theta} \right) = 0 \Rightarrow \left(Y - r \right) = 2c' \left(\hat{\theta} \right) \tag{4}$$

The optimal choice of θ now depends on whether $\hat{\theta}$ satisfies (1), and if it does whether it generates a positive ex-ante expected utility for the borrower. Essentially, the borrower's optimal ex-ante expected utility, $\mathrm{EU}(\theta)$, is given by,

$$\mathrm{EU}\left(\theta \right) = \max \left\{ 0, \frac{1}{2} \left[\hat{\theta} \left(Y - r \right) - e \right] - c \left(\hat{\theta} \right) \right\}$$

provided $\hat{\theta}$ satisfies (1) for the corresponding rate of interest (r). To economize on notation, we denote

$$g_1^I \left(r \right) = \frac{1}{2} \left[\hat{\theta} \left(r \right) \left(Y - r \right) - e \right] - c \left(\hat{\theta} \left(r \right) \right)$$

where $\hat{\theta}(r)$ is determined from (4). By envelope theorem, we know that

$$\left(g_1^I \left(r \right) \right)' = -\frac{1}{2} \hat{\theta} \left(r \right) < 0$$

A little algebra establishes[4] that there exists $r_c^I \in \left(\left(Y - 2e \right), \left(Y - e \right) \right)$, such that $g_1^I(r) \geq 0$ for $r \leq r_c^I$. Moreover, whenever$g_1^I(r) \geq 0$, the corresponding $\hat{\theta}(r)$ satisfies (1).

Now consider $r < (Y - 2e)$. Notice that for these values of r, (1) is satisfied for all values of θ. Hence, a borrower will take up the project for all values of θ after receiving S signal. However, (2) is satisfied only if θ is sufficiently low. Thus the borrower will take up the project after F signal only if the chosen signal quality (θ) is sufficiently low.

The ex-ante expected utility of the borrower thus can be written as

$$\mathrm{EU}\left(\theta \right) = \begin{cases} \frac{1}{2} \left[\theta \left(Y - r \right) - e \right] - c \left(\theta \right) & \text{if } \theta > 1 - \frac{e}{Y - r} \\ \frac{1}{2} \left(Y - r \right) - e - c \left(\theta \right) & \text{if } \theta \leq 1 - \frac{e}{Y - r} \end{cases} \tag{5}$$

Notice that if the borrower decides to take up the project even after signal F, ex-ante she loses all incentives to invest in signal quality and hence would choose $\theta^* = \frac{1}{2}$.

[4]The result follows from continuity of $g_1^I(r)$ and $g_1^I(Y - e) < 0$ while $g_1^I(Y - 2e) > 0$. We have omitted the proof which is available from the authors on request.

Her ex-ante expected utility from doing so is denoted by

$$g_2^I(r) = \left(\frac{Y-r}{2}\right) - e$$

On the other hand, she is better-off by taking up the project only after S if her chosen signal quality (θ) exceeds $\left(1 - \frac{e}{Y-r}\right)$. In that case, her ex-ante expected utility is given by $g_1^I(r)$.

Essentially, the borrower's optimal ex-ante expected utility, EU (θ), is given by

$$\mathrm{EU}(\theta) = \max\left\{g_1^I(r), g_2^I(r)\right\}$$

Let us now define

$$g^I(r) = g_1^I(r) - g_2^I(r).$$

Notice that

$$\left(g^I(r)\right)' = \left(g_1^I(r)\right)' - \left(g_2^I(r)\right)' = \left[\left(-\frac{1}{2}\hat{\theta}(r)\right) - \left(-\frac{1}{2}\right)\right] > 0$$

Moreover at $r = (Y - 2e)$,

$$g^I(r) = g_1^I(r) > 0$$

since $g_2^I(r) = 0$ at $r = Y - 2e$, while at $r = 0$

$$\begin{aligned} g^I(0) &= \left[\frac{1}{2}\left[\hat{\theta}(0).Y - e\right] - c\left(\hat{\theta}(0)\right)\right] - \left[\frac{Y}{2} - e\right] \\ &= \frac{e}{2} - c\left(\hat{\theta}(0)\right) - \left(1 - \hat{\theta}(0)\right)\frac{Y}{2} \end{aligned}$$

The sign of $g^I(0)$ depends on the level of effort cost e. If e is high, $g^I(0)$ is positive. If e is low, $g^I(0)$ is negative.

Depending on the value of e, we have two possibilities. Remember that in our relevant zone $g(r)$ is monotonic. As r falls, $g^I(r)$ falls. If $g^I(0) \geq 0$, then for every $r \leq (Y - 2e)$, the borrower collects $\theta^* = \hat{\theta}(r)$ and takes up the project only if a success signal (S) is realized. However, if $g^I(0) < 0$, then there exists $r_d^I \in (0, Y - 2e)$ such that the borrower collects $\theta^* = \hat{\theta}(r)$ if and only if $r > r_d^I$. Otherwise, she collects $\theta^* = \frac{1}{2}$.

We now summarize the above results in the following Lemma.

Lemma 1 *Suppose e is sufficiently high such that $g^I(0) \geq 0$ and the loan contract is individual liability contract. The borrower's optimal signal quality is:*

$$\theta^* = \begin{cases} \frac{1}{2} & \mathit{if}\ r > r_c^I \\ \hat{\theta}(r) & \mathit{if}\ r \leq r_c^I \end{cases}$$

where $r_c^I \in (Y - 2e, Y - e)$. *On the other hand, if e is such that* $g^I(0) < 0$, *the borrower"s optimal signal quality is:*

$$\theta^* = \begin{cases} \frac{1}{2} & \text{if } r > r_c^I \\ \hat{\theta}(r) & \text{if } r_c^I \geq r > r_d^I \\ \frac{1}{2} & \text{if } r \leq r_d^I \end{cases}$$

where $r_d^I \in (0, Y - 2e)$ *and* $\hat{\theta}(r)$ *is determined from*

$$(Y - r) = 2c'\left(\hat{\theta}\right).$$

The results above are intuitive. First consider the case when e is high. Notice that the information is relevant only in saving the effort cost e. For high effort cost, the information is valuable because the borrower wouldn't dare to risk losing the effort cost when she receives F signal. Therefore, the information is decision-relevant for rates of interest in which she has the incentive to take up the project. When e is low, the loss in case of project failure is low and therefore, for low values of r, when the gain from project success is high the borrower may be inclined to take up the project even after an F signal. However, in that case, the signal no longer remains decision-relevant, and hence the borrower stops investing in signal quality.

The Lemma stated above enables us to arrive at our first proposition. If r is very high, the borrower does not take up the project irrespective of the signal realization ($s \in \{S, F\}$), and hence does not invest in signal quality (θ). However, as r falls below a critical value $\left(r_c^I\right)$, the borrower decides to take up the project under good signal and invests a positive amount in signal quality. The borrower's investment in signal quality rises as r falls. If the private effort cost (e) of the project is very high, this goes on for all values of r below the critical level. Notice that in this case, the borrower

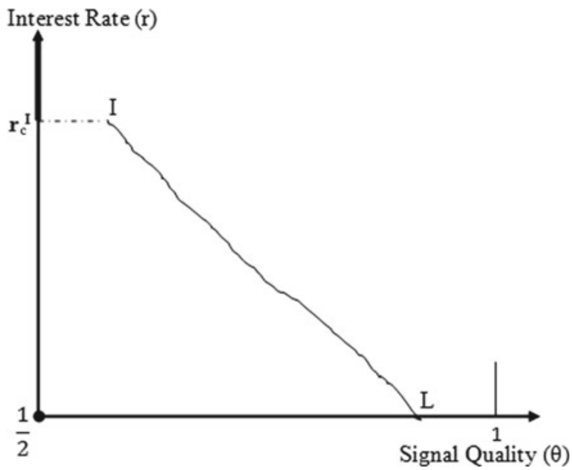


Fig. 1 Borrower's choice of θ under IL Lending with $g^I(0) \geq 0$

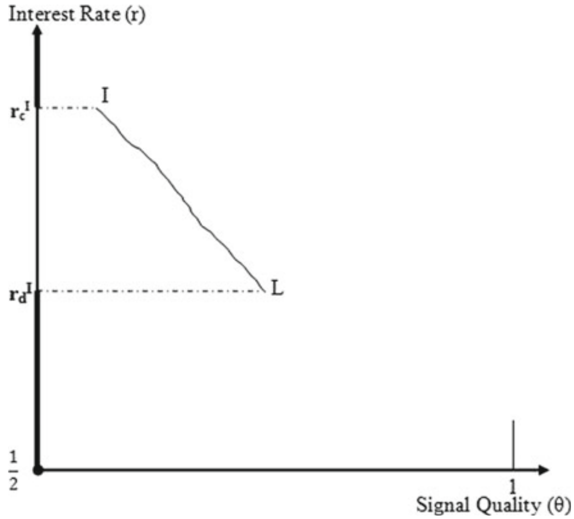


Fig. 2 Borrower's choice of θ under IL Lending with $g^I(0) < 0$

takes up the project only if the realized signal is a success signal ($s = S$) (Fig. 1). On the other hand, if the cost of effort (e) is not so high, then there is a second critical level of r $\left(= r_d^I\right)$ below which the borrower once again stops investing in signal quality although for a completely different reason. In this case, for low values of r the borrower takes up the project irrespective of the signal ($s \in \{S, F\}$) ; and hence the signal has no value to the borrower, i.e., $\theta = \frac{1}{2}$ (Fig. 2).

We state these results in the proposition below. For the ease of exposition, we denote the borrower's project decision by $d(r, s) \in \{0, 1\}$ where r is the gross interest and $s \in \{S, F\}$ is the realized signal. $d = 1$ indicates that the borrower takes up the project and $d = 0$ indicates that she doesn't.

Proposition 1 *Consider individual liability loan contracts. If e is sufficiently high such that $g^I(0) \geq 0$, the borrower's optimal signal quality and project decisions are given by*

$$\theta^* = \begin{cases} \frac{1}{2} & \text{if } r > r_c^I \\ \hat{\theta}(r) & \text{if } r \leq r_c^I \end{cases}$$

and

$$d(r, s) = \begin{cases} 0 \text{ if } r > r_c^I \text{ and } s = S, F \\ 0 \text{ if } r \leq r_c^I \text{ and } s = F \\ 1 \text{ if } r \leq r_c^I \text{ and } s = S \end{cases}$$

If e is such that $g^I(0) < 0$, the borrower's optimal signal quality and project decisions are given by

$$\theta^* = \begin{cases} \frac{1}{2} & \textit{if } r > r_c^I \\ \hat{\theta}(r) & \textit{if } r_d^I < r \leq r_c^I \\ \frac{1}{2} & \textit{if } r \leq r_d^I \end{cases}$$

and

$$d(r, s) = \begin{cases} 0 \textit{ if } r > r_c^I \textit{ and } s = S, F \\ 0 \textit{ if } r_d^I < r \leq r_c^I \textit{ and } s = F \\ 1 \textit{ if } r_d^I < r \leq r_c^I \textit{ and } s = S \\ 1 \textit{ if } r \leq r_d^I \textit{ and } s = S, F \end{cases}.$$

2.2 Joint Liability Group Lending

We now consider group lending under joint liability. We restrict the size of each group to two people ($n = 2$) with groups formed voluntarily. Individuals invest independently, but the loan contract involves joint liability. Since at the time of group formation signals, $s \in \{S, F\}$, are publicly observable, each borrower not only knows her own signal but also observes her partner's signal.

Given this set-up, notice that if borrower i with signal quality θ_i and signal realization S partners a borrower with signal quality θ_j with signal S, her expected pay-off (in case they end up taking loans) would be

$$\begin{aligned} \mathrm{EU}_i\left(\theta_i, \theta_j; s_i = S, s_j = S\right) &= \left[\theta_i\theta_j(Y - r) + \theta_i(1 - \theta_j)(Y - 2r)\right] - e - c(\theta) \\ &= \theta_i\left[(Y - r) - (1 - \theta_j)r\right] - e - c(\theta) \end{aligned} \quad (6)$$

If i's partner receives F signal her pay-off is

$$\begin{aligned} \mathrm{EU}_i\left(\theta_i, \theta_j; s_i = S, s_j = F\right) &= \left[\theta_i\left(1 - \theta_j\right)(Y - r) + \theta_i\theta_j(Y - 2r)\right] - e - c(\theta) \\ &= \theta_i\left[(Y - r) - \theta_j r\right] - e - c(\theta) \end{aligned} \quad (7)$$

Since $\theta_j \geq \frac{1}{2}$,

$$\mathrm{EU}_i\left(\theta_i, \theta_j; s_i = S, s_j = S\right) \geq \mathrm{EU}_i\left(\theta_i, \theta_j; s_i = S, s_j = F\right)$$

Hence a borrower with signal S will always seek another borrower with signal S. Similarly, a person with F signal will also have incentive to seek a partner with S signal, but wouldn't find one since every S signal holder will seek another S signal holder. Hence, when groups are formed voluntarily, group members will have similar signals. There would be two types of groups—groups of S signal holders and groups of F signal holders.

Notice also that EU of the ith borrower increases with the signal quality of her partner in both types of groups. Hence, each borrower would try to be matched with an individual with highest signal quality. If there are sufficient potential borrowers, this will lead to assortative matching—each borrower will be partnered by an individual exactly like herself—with same signal quality (θ) and same signal realization. This is stated in our next proposition.

Proposition 2 *Under joint liability loan contract with self selection of members, groups consist of homogeneous borrowers, i.e., a borrower with signal quality* $\theta \in \left[\frac{1}{2}, 1\right]$ *and signal realization* $s \in \{S, F\}$ *will be matched with a borrower with same signal quality and same signal realization.*

Once the group is formed, a borrower with a success (S) signal applies for loan only if

$$\theta\left[(Y-r)-(1-\theta)r\right]-e \geq 0 \tag{8}$$

since $c(\theta)$, the signal collection cost is already sunk. Similarly, a borrower with a failure (F) signal takes up the project only if

$$(1-\theta)\left[(Y-r)-\theta r\right]-e \geq 0 \tag{9}$$

Notice that since $\theta \geq \frac{1}{2}$, if for any r and θ (9) holds, (8) will definitely hold at the same r and θ.

For $r > (Y-e), \theta\left[(Y-r)-(1-\theta)r\right] < 0, \forall\, \theta \in \left[\frac{1}{2}, 1\right]$. Hence, the borrowers will not take up projects after either signal. For these values of r, since the project is never taken up, the borrower would not invest any amount in collecting signals and thus $\theta^* = \frac{1}{2}$.

Now suppose $\frac{2}{3}(Y-2e) < r \leq (Y-e)$. Notice that the LHS of (9) is decreasing[5] in both θ and r. At $\theta = \frac{1}{2}$ and as $r \to \frac{2}{3}(Y-2e)$, the LHS of (9) becomes

$$\frac{1}{2}\left[Y-\frac{2}{3}(Y-2e)-\frac{1}{2}.\frac{2}{3}(Y-2e)\right]-e=0$$

Hence, $\forall\, r$, such that $\frac{2}{3}(Y-2e) < r \leq (Y-e)$, the LHS of (9) is negative. Therefore, members of the borrower groups with F signals would not take up the project for these values of r.

On the other hand, members of the borrower groups with signal S would take up the project only if (8) holds. Hence, in this range of r, a borrower's ex-ante expected utility (before her signal is realized) is

---

[5] It is easy to show that

$$\frac{\delta}{\delta r}(\text{LHS}) = -\left(1-\theta^2\right) \leq 0$$

for all θ and

$$\frac{\delta}{\delta \theta}(\text{LHS}) = -Y + 2\theta r \leq 0$$

since $\theta \in \left[\frac{1}{2}, 1\right]$ and $r < \frac{Y}{2}$.

$$\mathrm{EU}\left(\theta\right)=\begin{cases}-c\left(\theta\right) & \text{if } \theta\left[\left(Y-r\right)-\left(1-\theta\right)r\right]<e\\ \frac{1}{2}\left[\theta\left\{\left(Y-r\right)-\left(1-\theta\right)r\right\}-e\right]-c\left(\theta\right) & \text{if } \theta\left[\left(Y-r\right)-\left(1-\theta\right)r\right]\geq e\end{cases} \tag{10}$$

This expression needs some clarification. Suppose $\frac{2}{3}\left(Y-2e\right)<r\leq\left(Y-e\right)$ and a borrower decides to invest in a signal quality θ at a cost $c\left(\theta\right)$. Once the signal is realized and the borrower receives signal F (which happens with probability $\frac{1}{2}$), she will be compelled to form a group with another borrower with quality θ and signal realization F by Proposition 2. However, in that case she wouldn't take up the project and her investment in quality is lost. If the borrower receives the signal S, she would take up the project by incurring the effort cost only if (8) holds.

Notice that ex-ante the borrower's best choice of θ is either $\frac{1}{2}$, or is determined from the equation

$$Y-2\left(1-\theta\right)r=2c'\left(\theta\right) \tag{11}$$

Given our assumption that $c'\left(1\right)>\frac{Y}{2}$, we have an interior solution[6] to this expression which is denoted by $\hat{\theta}_J(r)$. The ex-ante optimal expected utility of the borrower

$$\mathrm{EU}\left(\theta\right)=\max\left\{0,\frac{1}{2}\left[\hat{\theta}_J\left(r\right)\left\{\left(Y-r\right)-\left(1-\hat{\theta}_J\left(r\right)\right)r\right\}-e\right]-c\left(\hat{\theta}_J\left(r\right)\right)\right\}$$

Once again, to economize on notation we denote

$$g_1^J\left(r\right)=\frac{1}{2}\left[\hat{\theta}_J\left(r\right)\left(Y-r-\left(1-\hat{\theta}_J\left(r\right)\right)r\right)-e\right]-c\left(\hat{\theta}_J\left(r\right)\right)$$

For any $r\leq\left(Y-e\right)$, $Y-2\left(1-\theta\right)r\geq e$. Hence, (11) has an interior solution.

Once again for the sake of brevity, we omit the derivation the critical rate of interest below which the group of borrowers will choose to take up the project and thus invests in signal collection. Similar to the case under individual liability lending, there exists $r_c^J\in\left(\frac{2}{3}\left(Y-2e\right),\left(Y-e\right)\right)$, such that $g_1^J\left(r\right)\geq 0$ for $r\leq r_c^J$. Moreover, whenever$g_1^J\left(r\right)\geq 0$, the corresponding $\hat{\theta}_J\left(r\right)$ satisfies (8).

Next we consider the case when $r\leq\frac{2}{3}\left(Y-2e\right)$. Once again for these values of r, a borrower would take up the project after S signal for all values of θ since (8) holds for all θ . On the other hand, the project would be taken up after F only if θ satisfies (9). Hence, the ex-ante expected utility of the borrower can be written as

$$\mathrm{EU}\left(\theta\right)=\begin{cases}\frac{1}{2}\theta\left\{Y-\left(2-\theta\right)r\right\}-e-c\left(\theta\right) & \text{if } \left(1-\theta\right)\left[\left(Y-r\right)-\theta r\right]<e\\ \frac{Y-r}{2}-\theta\left(1-\theta\right)r-e-c\left(\theta\right) & \text{if } \left(1-\theta\right)\left[\left(Y-r\right)-\theta r\right]\geq e\end{cases} \tag{12}$$

[6] For the second-order condition, we need to assume that $c\left(.\right)$ is sufficiently convex everywhere. This can be ensured if we assume that for the relavant values of r, $c''\left(\theta\right)>r$. If we assume that $c''\left(\theta\right)>\frac{Y}{2}$ for all $\theta\in\left[\frac{1}{2},1\right]$, then the SOC will always be satisfied.

Unlike in the individual liability case, even when the borrower decides to take up the project irrespective of the signal realization, she may have an incentive to invest in signal quality in joint liability lending.

Under individual liability, the signal has a single purpose. It tells the borrower the likelihood of project success under different signal realization and thus helps the borrower in her decision regarding taking up of the project. If the likelihood of project success is low, the borrower can save her own effort cost by not taking up the project. If the borrower decides to take up the project irrespective of the signal realization, there is no point in investing in signal quality under individual liability.

In the case of joint liability, the choice of signal quality also determines one's partner (see Proposition 2). Investment in signal quality thus not only improves one's own information about the likelihood of project success, but also helps in finding a more informed partner whose success or failure in turn determines the borrower's own pay-off. Thus even when the borrower decides to take up the project for both signal realizations, she may have an incentive to invest in signal quality because of this second effect. However, this incentive gets diminished as r falls since the joint liability payment falls with r. In our analysis, we assume that the cost function is such that the first effect dominates and thus if the borrower decides to take up the project irrespective of the signal realization, she does not invest in signal quality.

Let us look at a borrower's decision more closely. Suppose the borrower decides to take up the project under both signal realizations. Then she chooses θ to maximize

$$\phi(\theta) = \left(\frac{Y-r}{2}\right) - \theta(1-\theta)r - e - c(\theta)$$

subject to (9). Suppose that the solution to the above problem is denoted by $\tilde{\theta}(r)$. Since

$$\phi'\left(\frac{1}{2}\right) = -c'\left(\frac{1}{2}\right) \leq 0$$

$\tilde{\theta}(r) = \frac{1}{2}$ is one solution to the problem. Notice that $\tilde{\theta}(r) = \frac{1}{2}$ satisfies (9) for all $r \leq \frac{2}{3}(Y-2e)$. The solution is unique if $\phi(\theta)$ is concave in θ. Since $\phi''(\theta) = 2r - c''(\theta)$, to ensure concavity of $\phi(\theta)$, we need $c''(\theta) > 2r$. This definitely holds for very low values of r, but the assumption made earlier—$c''(\theta) > \frac{Y}{2}$—cannot guarantee the concavity of $\phi(\theta)$ for all $r \leq \frac{2}{3}(Y-2e)$. For the sake of avoiding analytical complication, we assume that this is indeed the case.[7]

[7] We are not losing much by making this assumption. If it does not hold, then we may have multiple local maxima for relatively higher values of r. For any given value of r, the borrower would choose θ that gives her highest expected utility among these local maximums. However, the value function (optimized expected utility function) will be continuous in r, though the optimal choice of $\tilde{\theta}(r)$ would change discontinuously as r falls. When r is low enough, $\tilde{\theta}(r) = \frac{1}{2}$. Unless we make the assumption mentioned above, there might be a range of r, in which $\tilde{\theta}(r) > \frac{1}{2}$ and falls as r falls. Once r falls below a critical level, $\tilde{\theta}(r)$ falls to $\frac{1}{2}$ and remains there for all lower values of r.

We use the notation $g_2^J(r)$ to denote the optimized value of $\phi(\theta)$. Since by virtue of the assumption mentioned above $\tilde{\theta}(r) = \frac{1}{2}$ for all $r \leq \frac{2}{3}(Y - 2e)$, we can write

$$g_2^J(r) = \frac{Y}{2} - \frac{3r}{4} - e$$

Alternatively, the borrower may decide to take up the project only if she receives a success signal and in that case her expected utility is

$$g_1^J(r) = \frac{1}{2}\left[\hat{\theta}_J(r)\left(Y - r - (1 - \hat{\theta}_J(r))r\right) - e\right] - c\left(\hat{\theta}_J(r)\right)$$

Essentially, the borrower's optimal ex-ante expected utility under joint liability lending for $r \leq \frac{2}{3}(Y - 2e)$ is given by

$$\max\left\{g_1^J(r), g_2^J(r)\right\}$$

We now define $g^J(r) = g_1^J(r) - g_2^J(r)$. Notice that

$$\left(g^J(r)\right)' = -\frac{1}{2}\hat{\theta}_J(2 - \hat{\theta}_J) + \frac{3}{4} = \frac{1}{2}\left[\frac{3}{2} - \hat{\theta}_J(2 - \hat{\theta}_J)\right] > 0$$

since $2 - \hat{\theta}_J < \frac{3}{2}$ and $\hat{\theta}_J < 1$. Hence as r falls $g_J(r)$ falls. We already know that $g_1^J(r) > 0$ at $r = \frac{2}{3}(Y - 2e)$ while $g_2^J(r) = 0$. However, at $r = 0$,

$$g^J(r) = \frac{1}{2}\left[\hat{\theta}_J(0)\,Y - e\right] - c\left(\hat{\theta}_J(0)\right) - \left(\frac{Y}{2} - e\right)$$

A comparison of (4) and (11) immediately tells us that at $r = 0$, $\hat{\theta}_J(0) = \hat{\theta}(0)$ and thus

$$g^J(0) = g^I(0) = \left(\hat{\theta}(0) - \frac{1}{2}\right)\frac{Y}{2} - c\left(\hat{\theta}(0)\right) - \frac{1}{2}\left(\frac{Y}{2} - e\right)$$

As r falls, $g^J(r)$ falls. If $g^J(0) \geq 0$, then for every $r \leq (Y - 2e)$, the borrower collects $\theta^* = \hat{\theta}_J(r)$ and takes up the project only if a success signal is realized under joint liability. However, if $g^J(0) < 0$, then there exists $r_d^J \in \left(0, \frac{2}{3}(Y - 2e)\right)$ such that the borrower collects $\theta^* = \hat{\theta}_J(r)$ if and only if $r > r_d^J$. Otherwise, she collects $\theta^* = \frac{1}{2}$. This is stated in our next Lemma.

Lemma 2 *Suppose the loan contract is joint liability contract with full liability. If e is sufficiently high such that $g^J(0) \geq 0$, the borrower's optimal signal quality is*

$$\theta^* = \begin{cases} \frac{1}{2} & \textit{if } r > r_c^J \\ \hat{\theta}_J(r) & \textit{if } r \leq r_c^I \end{cases}$$

where $r_c^J \in \left(\frac{2}{3}(Y-2e), Y-e\right)$*. On the other hand, if* e *is such that* $g^J(0) < 0$*, the borrower's optimal signal quality is:*

$$\theta^* = \begin{cases} \frac{1}{2} & \text{if } r > r_c^J \\ \hat{\theta}_J(r) & \text{if } r_c^J \geq r > r_d^J \\ \frac{1}{2} & \text{if } r \leq r_d^J \end{cases}$$

where $r_d^J \in \left(0, \frac{2}{3}(Y-2e)\right)$ *and* $\hat{\theta}_J(r)$ *is determined from*

$$Y - 2(1-\theta) r = 2c'(\theta)$$

The decision of an individual borrower regarding choice of signal quality and the project choice decision under joint liability loan contract can be summarized in the following proposition:

Proposition 3 *Consider joint liability loan contracts. If* e *is sufficiently high such that* $g^J(0) \geq 0$*, the borrower's optimal signal quality and project decisions are given by*

$$\theta^* = \begin{cases} \frac{1}{2} & \text{if } r > r_c^J \\ \hat{\theta}_J(r) & \text{if } r \leq r_c^J \end{cases}$$

and

$$d(r,s) = \begin{cases} 0 \text{ if } r > r_c^J \text{ and } s = S, F \\ 0 \text{ if } r \leq r_c^J \text{ and } s = F \\ 1 \text{ if } r \leq r_c^J \text{ and } s = S \end{cases}$$

If e *is such that* $g^J(0) < 0$*, the borrower's optimal signal quality and project decisions are given by*

$$\theta^* = \begin{cases} \frac{1}{2} & \text{if } r > r_c^J \\ \hat{\theta}_J(r) & \text{if } r_d^J < r \leq r_c^J \\ \frac{1}{2} & \text{if } r \leq r_d^J \end{cases}$$

and

$$d(r,s) = \begin{cases} 0 \text{ if } r > r_c^J \text{ and } s = S, F \\ 0 \text{ if } r_d^J < r \leq r_c^J \text{ and } s = F \\ 1 \text{ if } r_d^J < r \leq r_c^J \text{ and } s = S \\ 1 \text{ if } r \leq r_d^J \text{ and } s = S, F \end{cases}$$

The borrowers invest in costly information only when the information is decision-relevant. At very high rates of interest, the project is not worth taking up and thus information about the potential success have no value. If the rate of interest is very low, it may be worthwhile to take up the project irrespective of the signal realization when the effort cost is low as well since the gain from project success is high while the failure only costs the low effort cost given limited liability. In such a scenario, information is irrelevant. However, if the effort cost is high, then the loss from project failure is also high and hence the information remains decision-relevant even at low rates of interest.

3 Comparisons Between Individual and Joint Liability Loan Contracts

We have seen from Propositions 1 and 3 that under both types of loan contracts a borrower invests in signal quality if she decides to take up the project only under S signal. If she decides to take up the project irrespective of her signal realizations, she does not invest in information.

However, if a borrower decides to invest in signal quality under both types of loan contract, she invests more under joint liability. This is stated in our next proposition.

Proposition 4 *If at any r, both $\hat{\theta}(r)$ and $\hat{\theta}_J(r)$ are greater than $\frac{1}{2}$, then $\hat{\theta}_J(r) > \hat{\theta}(r)$.*

Proof Notice that $\hat{\theta}(r)$ and $\hat{\theta}_J(r)$ are determined from

$$Y - r = 2c'(\theta)$$

and

$$Y - 2(1-\theta)r = 2c'(\theta)$$

respectively. Since $[Y - 2(1-\theta)r] > (Y - r)\ \forall\, r, \theta \in (\frac{1}{2}, 1]$, and $c''(\theta) > 0$, the result holds. ■

This is intuitive. The signal collection decisions are taken ex-ante before the group is formed. Under joint liability lending, a borrower, by collecting a better quality signal, not only gets better information about project potential but also enables herself to form a group with a better informed partner through assortative matching. This gives her incentive to collect better quality signal than under individual liability contract.

Notice that at the interest rates $\left(r_c^I, r_c^J\right)$ and $\left(r_d^I, r_d^J\right)$ are critical levels at which the borrowers switch from *no-investment* to *investment* in signal quality, or from *investment* to *no-investment* ,respectively under two types of loan contract. We now compare these switching rates of interest under the two regimes. We call these interest rates as the switching rates of interest.

Proposition 5 *The switching rates of interest are higher under individual liability lending than under joint liability lending; i.e. $r_c^I > r_c^J$, and when $g(0) < 0$, then $r_d^I > r_d^J$.*

Proof Please see Appendix. ■

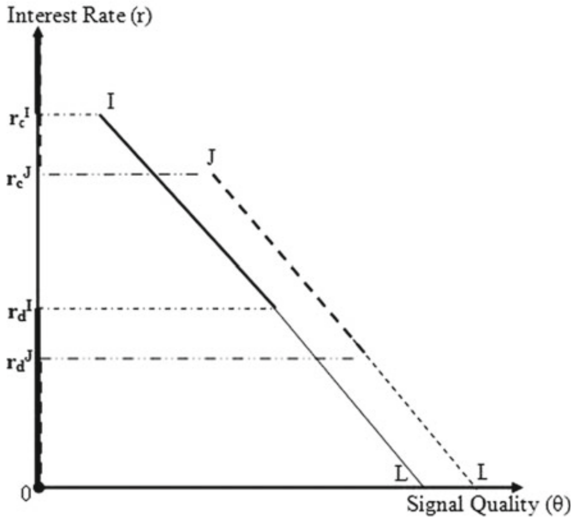


Fig. 3 Borrower's choice of θ under IL Lending and JL Lending

The interest rate at which the borrower starts investing in signal quality is higher under individual liability (Fig. 3). For very high rates of interest, the gain from project success is low, but under joint liability the cost of partner's failure is high. This may induce a borrower to abandon a project under joint liability contract which she may take up under individual liability. On the other hand for low rates of interest, a borrower may stop investing in information under individual liability contract, while she still may have incentive for costly information collection for finding a better partner.

Remark 1 If the rate of interest is high, the individual liability loan contract may be welfare improving for the borrowers.

This is obvious from Proposition 5. For rates of interest in between r_d^J and r_c^I, a borrower would not take up the project under joint liability contract, but would do so under individual liability. The fact that she has incentive to take up the project indicates that the borrower's utility from doing so is positive. If she doesn't take it up, her utility is zero. Thus, the borrower does better under individual liability than under joint liability.

4 Concluding Remarks

This paper attempts to provide a theoretical framework to analyze the issues of endogenous information collection and group formation in the microcredit market. Borrowers generally collect information about the potential of their intended projects since undertaking a project requires the borrower to put in effort, time and money which are costly and are sunk in case of project failure. Information about the chance of project success allows the borrower to save these costs if she finds out that the project will fail almost surely. This also allows the lenders to lower the rate of interest generally leading to higher efficiency in the microcredit market. However, cost of collecting information rises with the quality of information. The borrower faces a trade-off at this stage of information collection: whether to bear the cost of signal collection, or to bear the cost of time and effort in case the realized signal indicates strong probability of failure. This paper attempts to understand this trade-off in terms of a simple model of credit market. Further, we have examined how individual lending mechanism vis-a-vis the group lending mechanism affects the ex-ante incentive for information collection and the welfare of the agents when the group is formed after the signal realization.

Here, both under limited liability individual lending and joint liability group lending, we find that if interest rate is very high, the borrower(s) does not take up the project irrespective of the signal realization and hence does not invest in signal quality. However, as interest rate falls below a critical value, the borrower(s) decides to take up the project under good signal and invests a positive amount in signal quality. The investment in signal quality rises as interest rate falls. If the private effort cost of the project is very high, this goes on for all values of interest rate below the critical level. In this case, the borrower(s) takes up the project only if the realized signal is a success signal. On the other hand, if effort cost is not so high, then there is a second critical level of interest rate below which the borrower(s) once again stops investing in signal quality. In this case, for low values of interest rate, the borrower(s) takes up the project irrespective of the signal and hence the signal has no value to the borrower(s)—she does not invest in information. However, we interestingly notice that if a borrower decides to invest in signal quality under both types of loan contract, she invests more under joint liability.

Comparing the two types of lending contracts, we further observe that the switching rates of interest are higher under individual liability lending than under joint liability lending. We further observe that if the lender charges a high rate of interest, the borrower may have the incentive to take up the project under individual liability, but not under joint liability. Our model, in conventional line, finds that there occurs assortative matching in the group formation stage under joint liability contract (though quality of information is chosen independently by each borrower) and this incentivizes better signal collection under joint liability lending at low and medium rates of interest.

We intend to extend the model to examine the effect of lender motivation on information collection. We may also examine how the ex-ante incentive for information collection and the welfare of the agents is affected by group formation in different stages under group lending. We have not discussed the cases where the group members may cooperatively decide on the quality of information collected by the members; or, a group that is formed after the borrowers independently decide on the quality of information they would collect, but before the signal realization. Interestingly, group formation at this stage does not leave any scope for group punishment based on quality of information. We would explore these possibilities in our future research.

Appendix

Proof of Proposition 5 Notice that at $r = r_c^I$,

$$g_1(r) = \frac{1}{2}\left[\hat{\theta}(r)(Y - r) - e\right] - c\left(\hat{\theta}(r)\right) = 0$$

and

$$\begin{aligned} g_1^J(r) &= \frac{1}{2}\left[\hat{\theta}_J(r)\left(Y - r - (1 - \hat{\theta}_J(r))r\right) - e\right] - c\left(\hat{\theta}_J(r)\right) \\ &= \frac{1}{2}\left[\hat{\theta}_J(r)(Y - r) - e\right] - c\left(\hat{\theta}_J(r)\right) - \hat{\theta}_J(r)(1 - \hat{\theta}_J(r))r \\ &< \frac{1}{2}\left[\hat{\theta}_J(r)(Y - r) - e\right] - c\left(\hat{\theta}_J(r)\right) \leq \frac{1}{2}\left[\hat{\theta}(r)(Y - r) - e\right] - c\left(\hat{\theta}(r)\right) \\ &= 0 \end{aligned}$$

where the last inequality follows from the fact that $\hat{\theta}(r)$ maximizes $\left[\frac{1}{2}\left[\theta(Y - r) - e\right]\right.$ $\left.-c(\theta)\right]$. Since $g_1^J\left(r_c^I\right) < 0$, $\left(g_1^J(r)\right)' < 0$ and $g_1^J\left(r_c^J\right) = 0$, it must be the case that $r_c^I > r_c^J$ (please see, Fig. 3).

Suppose $g(0) < 0$. Notice that

$$g^J(r) = \frac{1}{2}\left[\hat{\theta}_J(r)\left(Y - r - (1 - \hat{\theta}_J(r))r\right) - e\right] - c\left(\hat{\theta}_J(r)\right) - \left(\frac{Y}{2} - \frac{3r}{4} - e\right)$$

and

$$g^I(r) = \frac{1}{2}\left[\hat{\theta}(r)(Y - r) - e\right] - c\left(\hat{\theta}(r)\right) - \left(\frac{Y}{2} - \frac{r}{2} - e\right)$$

Hence, for any r, $g^J(r) > g(r)$ if and only if

$$\begin{aligned} &\frac{1}{2}\left[\hat{\theta}_J(r)\left(Y - r - (1 - \hat{\theta}_J(r))r\right) - e\right] - c\left(\hat{\theta}_J(r)\right) - \left(\frac{Y}{2} - \frac{3r}{4} - e\right) \\ &> \frac{1}{2}\left[\hat{\theta}(r)(Y - r) - e\right] - c\left(\hat{\theta}(r)\right) - \left(\frac{Y}{2} - \frac{r}{2} - e\right) \end{aligned}$$

$$\Leftrightarrow \frac{r}{4} - \frac{1}{2}\hat{\theta}_J(r)(1 - \hat{\theta}_J(r))r > c\left(\hat{\theta}_J(r)\right) - c\left(\hat{\theta}(r)\right) - \left(\hat{\theta}_J(r) - \hat{\theta}(r)\right)\frac{Y - r}{2}$$

Now, the RHS of the above expression

$$
\begin{aligned}
& c\left(\hat{\theta}_J(r)\right) - c\left(\hat{\theta}(r)\right) - \left(\hat{\theta}_J(r) - \hat{\theta}(r)\right)\frac{Y-r}{2} \\
&= \int_{\hat{\theta}}^{\hat{\theta}_J}\left[c'(\theta) - \frac{Y-r}{2}\right]d\theta \\
&< \int_{\hat{\theta}}^{\hat{\theta}_J}\left[c'\left(\hat{\theta}_J\right) - \frac{Y-r}{2}\right]d\theta
\end{aligned}
$$

since $c''(\theta) > 0$. But from (11)

$$
\begin{aligned}
\int_{\hat{\theta}}^{\hat{\theta}_J}\left[c'\left(\hat{\theta}_J\right) - \frac{Y-r}{2}\right]d\theta &= \int_{\hat{\theta}}^{\hat{\theta}_J}\left[\frac{Y-2\left(1-\hat{\theta}_J\right)r}{2} - \frac{Y-r}{2}\right]d\theta \\
&= \int_{\hat{\theta}}^{\hat{\theta}_J}\frac{\left(2\hat{\theta}_J - 1\right)r}{2}d\theta \\
&= \left(\hat{\theta}_J - \frac{1}{2}\right)\left(\hat{\theta}_J - \hat{\theta}\right)r
\end{aligned}
$$

We now show that for any $r > 0$,

$$
\frac{r}{4} - \frac{1}{2}\hat{\theta}_J(1-\hat{\theta}_J)r > \left(\hat{\theta}_J - \frac{1}{2}\right)\left(\hat{\theta}_J - \hat{\theta}\right)r \Leftrightarrow \frac{1}{2} > \hat{\theta}_J^2 - 2\hat{\theta}_J\hat{\theta} + \hat{\theta}
$$

which holds since $\hat{\theta}_J, \hat{\theta} \in \left(\frac{1}{2}, 1\right)$. Hence, we can conclude that for any $r > 0$, $g^J(r) > g^I(r)$. Since, at $r = r_d^I$, $g^I(r) = 0$, we can conclude that $g^J\left(r_d^I\right) > 0$. Since $g^{J\prime}(r) > 0$ and $g^J\left(r_d^J\right) = 0$, it must be the case that $r_d^J < r_d^I$. The interest rate at which the borrower starts investing in signal quality is higher under individual liability.

References

Besley, T., & Coate, S. (1995). Group lending, repayment incentives and social collateral. *Journal of Development Economics*, *46*, 1–18.

Bhattacharya, S., Banerjee, S., & Mukherjee, S. (2008, March). Group lending and self help groups: Joint benefit as an alternative governance mechanism. *The Journal of International Trade and Development*, *17*(1).

Chowdhury, S., Roy Chowdhury P., & Sengupta, K. (2014). Sequential lending with dynamic joint liability in micro-finance, Discussion Papers in Economics, Discussion Paper 14-07. Indian Statistical Institute, Delhi, India.

Conning, J. (1999). Outreach, sustainability and leverage in monitored and peer-monitored lending. *Journal of Development Economics*, *60*, 51–77.

de Aghion, B. A. (1999). On the design of a credit agreement with peer monitoring. *Journal of Development Economics*, *60*, 79–104.

Ghatak, M. (1999). Group lending, local information and peer selection. *Journal of Development Economics*, *60*, 27–50.

Ghatak, M., & Guinnane, T. W. (1999). The economics of lending with joint liability: Theory and practice. *Journal of Development Economics*, *60*, 195–228.

Ghosh, P., & Ray, D. (1997). *Information and repeated interaction: Application to informal credit markets*. Draft: Texas A & M and Boston University.

Janvry, A., McIntosh, C., & Sadoulet, E. (2010). The supply and demand side impacts of credit market information. *Journal of Development Economics, 93*(2), 173–188.

Morduch, J., & de Aghion, B. A. (2004a). Microfinance: Where do we stand. In C. Goodhart (Ed.), *Financial development & economic growth: Explaining the links*. London: Palgrave McMillan.

Morduch, J., & de Aghion, B. A. (2004b). Microfinance beyond group lending. *The Economics of Transition*, *8*, 401–420.

Mukherjee, S., & Bhattacharya, S. (2014). Optimal joint liability and group size in microcredit. *Arthaniti-Journal of Economic Theory and Practice*, *13*(1–2), (19–47).

Mukherjee, S., & Bhattacharya, S. (2015). Optimal group size under group lending with joint liability and social sanction. *Indian Growth and Development Review*, *8*(1).

Roy Chowdhury, P. (2005). Group lending: Sequential financing, lender monitoring and joint liability. *Journal of Development Economics*, *77*, 415–439.

Roy Chowdhury, P. (2007). Group-lending with sequential financing, contingent renewal and social capital. *Journal of Development Economics*, *84*, 487–507.

Sinha, F. (2005, April 23). Access, use and contribution of microfinance in India: Findings from a national study. *Economic and Political Weekly, XL*(17).

Stiglitz, J. E. (1990). Peer monitoring and credit markets. *The World Bank Economic Review*, *4*, 351–366.

Tsukada, K. (2012). Microfinance revisited: Towards a more flexible lending contracts. In A. Shonchoy (Ed.), *Seasonality adjusted flexible micro-credit: An randomized experiment in Bangladesh*, Interim Report. Chosakenkeu Hokokusho, IDE-JETRO 2012.

Van Tassel, E. (1999). Group lending under asymmetric information. *Journal of Development Economics*, *60*, 3–25.

Formal Contract Enforcement and Entrepreneurial Success of the Marginalized

Tanika Chakraborty, Reema Kumari, Anirban Mukherjee and Sarani Saha

Abstract Effective contract enforcement is the key for the formation and expansion of business enterprises. Contract enforcement which comes under the purview of the judicial system forms an important constituent of the formal institutional framework. In this paper, we aim to address the question whether better functioning formal judiciary institutions facilitate business in a less-developed country where entrepreneurs often take recourse to informal networks rather than approach formal courts for enforcing contracts. More specifically, we are empirically estimating this relationship between efficiency of formal institutions and various indicators of business performance by utilizing a unique administrative data collected from the judicial authority. We merge this data with the survey of medium- and small-scale industries to find the effect of several court efficiency measures such as duration rate and congestion rate on different business performance indicators. We find that improved court efficiency help entrepreneurs across the board, but the impacts are stronger for disadvantaged groups such as scheduled castes and tribes.

Keywords Contract enforcement · Court quality · Entrepreneurship · Caste

1 Introduction

Judicial efficiency plays a critical role for business proliferation by facilitating contracts. In this paper, we examine the effect of judicial efficiency on different types of firm-level decisions which have important implications for a firm's efficiency. More

T. Chakraborty
Economics, Indian Institute of Management Calcutta, Kolkata, India

R. Kumari
Gartner India Research and Advisory Services Pvt. Ltd, Bengaluru, India

A. Mukherjee (✉)
Department of Economics, University of Calcutta, Kolkata, India
e-mail: ameco@caluniv.ac.in

S. Saha
Department of Economic Sciences, Indian Institute of Technology Kanpur, Kanpur, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_9

importantly, we check whether courts help the marginalized groups such as Scheduled Castes and Tribes more than their general caste counterpart. Our working hypothesis suggests that in districts with more efficient judiciary system, firms will find it more convenient to operate at the optimal scale. The argument that courts help business proliferation is far from trivial in the context of less-developed countries where entrepreneurs avoid approaching courts even when they are cheated. The bias against moving to court was confirmed in our interviews with the entrepreneurs in Kanpur and Kolkata. Such tendency to avoid court is widespread in many less-developed countries and possibly the reason why Acemoglu and Johnson (2003) found no significant impact of contract enforcement institutions on economic development in their cross-country study.

Besides invoking scholarly interest, the issue of slow judicial process in India has started to attract the attention of media and government administration alike. In January 2013, there were around 30 million cases pending in different courts in India. At present on an average, it takes around 15 years for a civil case to get resolved. If the situation does not improve, there will be an estimated 150 million pending cases by 2040 (Times of India, January 17, 2013). There is one strand of literature computing the degree of pendency in Indian courts and another strand of literature examining the impact of court quality on business performance. Our paper stands at the crossroad of these two avenues of research.

In the first strand of the literature, many papers looked at the situation of case pendency in India. In one such paper, Sinha (2018) compared pendency rates in Supreme Court, High Court, and subordinate courts between the years 2006 and 2017. She found substantial increase in the caseload of High Courts and Supreme Courts (17% and 35%, respectively), whereas the number remained more or less stable for the lower courts. They found that there is a rise in the disposal rate, but it was not sufficient to deal with increasing admission rate. They also looked at the vacancy rate as a possible reason behind this large pendency rate and found that overall, there has been a rise in vacancy from 23% in 2006 to 35% in April 2018 in High Court and the Supreme Court. Examining in the same line, Hazra and Micevska (2004) found that there is some correlation between vacancy rate and court congestion.

In another related article, Rabiyath and Rupakula (2010) examined whether variables such as population, net state domestic product (NSDP), number of judges, and judicial expenditure have any impact on case pendency in the states of Kerala and Andhra Pradesh between the years 1980–81 and 2000–01. They also test whether higher pendency discourages litigants to file cases. They did not find any relation between NSDP and pendency but find that higher pendency discourages litigants to file new cases.

While the papers mentioned above discuss the possible reasons behind pendency, our paper is more closely related to the section of the literature that looks at the effect of contracting institutions on trade and technology adoption and explains cross-country difference in development through these channels. Nunn (2007), for example, shows that countries with high quality of contract enforcement export goods which use more inputs which require relationship-specific investments. Acemoglu et al.

(2007), on the other hand, show that greater contractual incompleteness—which could result from weak institutions—leads to the adoption of less advanced technologies. This effect enhances when there is greater complementarity among the intermediate inputs. In a similar line of enquiry on the relationship between contract enforcement institutions and trade, Levchenko (2007) models institutional differences among the countries as the source of comparative advantage and shows that the less-developed countries may not gain from trade.

Besides the cross-country studies mentioned above, the current paper is also related to the section of the literature that examines the impact of court efficiency on business performance in India. In one of the most important studies in this area, Chemin (2010) looked at the effect of Civil Procedure (Amendment) Act—a judicial reform in India that facilitated speedy disposal of civil lawsuits—on business practices. He found that this reform led to fewer breaches, encouraged investment, and facilitated access to finance. In another related study, Chemin (2009a) used some temporal and spatial reforms in Indian judiciary system to find the causal impact of judicial efficiency on credit markets, agricultural development, and manufacturing performance. The results are consistent with the theoretical prediction that lengthy judicial trials lead to worse outcomes in credit, agricultural, or manufacturing sector. Working in the same area, Ahsan (2013) looked at the complementarity between trade liberalization and judicial efficiency. He found that enterprises situated in states with better judicial efficiency gain more from trade liberalization than the enterprises located in states with weaker judiciary. Our paper is also related to Amirapu (2015) who builds on Nunn (2007) to see how judicial efficiency in India affects industries differentially depending on different contract intensity of the industry. He found that faster courts have a strong effect on more contract-intensive industries.

In this paper, we use a novel data set on district-level caseloads and combine that with firm-level data of micro, small, and medium enterprises from West Bengal to study how court efficiency affects business performance. We find that firms are more likely to register their business sooner than later when courts are more efficient. We also extend our analysis to see if improvements in courts differentially affect marginalized groups such as Scheduled Castes and Tribes (SC/ST). An important section of the literature examines how institutions such as market or state mitigate inter-caste inequality (Chin and Prakash 2011; Hnatkovska et al. 2012, 2013). Our paper also contributes in this area by examining how another state run institution, viz. court, affects this mechanism helping SC/ST entrepreneurs.

In both our baseline and extended analysis, we account for several firm-level characteristics in our empirical specification. However, given the cross-sectional nature of the data, we cannot rule out the possibility that the observed relationship is driven by district-level characteristics that correlate both with court efficiency and the decision of firms to register sooner. In the case of the baseline findings for instance, it is a possibility that districts that have better governance enable quicker registration and also have better functioning courts. However, these correlates are unlikely to explain the more beneficial effects of court quality on marginalized groups. In the latter case, the potential endogeneity could be driven by the possibility that the strength of institutions is themselves determined by the caste compositions. Thus, areas with higher fraction of marginalized sections might have developed weaker institutions. In light of these possibilities, caution should be exercised in interpreting our results causally.

We propose a simple theoretical structure explaining why formal courts may help entrepreneurs from underprivileged social groups. In our structure, business owners buy inputs from sellers who can potentially cheat the entrepreneurs by supplying bad quality inputs. In the absence of an effective formal court, the only way an entrepreneur can punish an input supplier is by community sanction—if an input seller cheats an entrepreneur from community A, no one from community A does business with him. Note that the effectiveness of such punishment to a dishonest input supplier is limited by the size of the entrepreneur's community. Historically, Scheduled Castes and Tribes (SC/ST) are not the traditional entrepreneurial social groups in India. This feature is also reflected in our sample that we used for regression. The proportion of SC/ST entrepreneurs is 13%. Hence, given this historical regularity, our theoretical structure predicts that an SC/ST entrepreneur cannot effectively discipline a potential cheater using his network input suppliers. Hence, the development of formal courts helps the groups which are socially marginalized.

The paper is structured as follows. In Sect. 2, we discuss the model which draws on Greif (1993); in Sect. 3, we detail the court data that we have collected from different courts; in Sect. 4, we present the regression results that we have obtained by combining our court data with enterprise data from Ministry of Small and Medium Scale Enterprises (MSME) where we test the effect of court quality on entrepreneurial potential of the marginalized groups and various other organizational structures of the firms. In the last section, we conclude.

2 Model

Suppose there is a commodity Y that is produced by entrepreneurs using an input X. The entrepreneurs can be divided into two social groups which could be caste, religion, or gender. Let us call them high and low. The entrepreneurs hailing from the high group are the traditional business class, and therefore, high entrepreneurs are in higher proportion among the entrepreneurs than their low community counterpart. The number of high and low entrepreneurs is M_H and M_L, respectively. The input suppliers, however, are not classified into communities (or put differently, and their identities are not important in any of the decision-making by the suppliers in our model). Total number of suppliers is A and $A < M$. There is a one sided moral hazard problem meaning that only the input supplier may cheat the entrepreneur by supplying bad inputs. Each of the entrepreneurs is different in terms of their ability and so is their outside options. The opportunity income of the jth entrepreneur is given by $\bar{\pi}_j$.

At the beginning of each period, an entrepreneur can choose a supplier from the pool of unemployed suppliers and the supplier cannot supply input to more than one producer. Now, an input supplier who has been employed can decide whether to be honest or to cheat. If the supplier is honest, her payoff is P_x and the producer's payoff is $\pi_y - P_x$. Hence, gross gain from cooperation is π_y. However, if the supplier cheats, her payoff is λ and producer's payoff is 0. The entrepreneur gets $\bar{\pi}$ if he does not hire an input supplier. A supplier's reservation utility is $\bar{u} > 0$. At the end of each period, allocations are realized and the entrepreneur decides whether to terminate the supplier or keep him in job. However, the contract may be also terminated for some exogenous reason which is beyond the control of the entrepreneur or the supplier. The probability that a contract can get terminated due to an exogenous reason is τ. It is assumed that $\pi_y > \bar{\pi}_j + \bar{u}$ for all j. This means cooperation is efficient—total social return to doing business (π_y) is more than *no business* where both entrepreneurs and suppliers get their reservation incomes—$\bar{\pi}_j$ and $\bar{u}$, respectively. We also assume $\pi_y > \lambda > \bar{u}$. This condition means that cheating entails a loss to the society and suppliers prefer cheating to receiving their reservation utility.

In the presence of one-sided moral hazard and the absence of any third-party formal enforcement, the employer can use some variant of efficiency wage as an instrument to solve a typical principal-agent problem. In this structure, the only punishment that an employer can enforce is firing the dishonest agent where the cost of the punishment for the agent is her foregone wage. Therefore, promising her a higher wage would increase the cost of punishment and make cheating less likely. The underlying mechanism of this devise is detailed in Shapiro and Stiglitz (1984). This mechanism gets further effective if the the cheated principal belongs to a community which imposes a social sanction against the cheater agent. Under this mechanism if an agent (input supplier in our case) cheats a principal (entrepreneur in our case) from a certain community, no other principal from that community ever employs the agent again. Hence, the cost of cheating for the agent is not only the forgone wage of his current employment but wages from all possible future employment as well. This mechanism is detailed in Greif (1993) in the context of the agency relation of a medieval trading group known as the Maghribi traders.

The equilibrium strategy is labeled as multilateral punishment strategy (MPS). According to this strategy, an entrepreneur offers a supplier the input price of P_x^*, rehires the same supplier if he was honest (unless forced termination has occurred), fires the supplier if he has cheated, and never hires a supplier who has cheated any entrepreneur from his own community, and in the event of forced separation, he hires a supplier with no cheating history. A supplier's strategy would be to remain honest if paid at least P_x^* and cheat if paid less than P_x^*. Before showing that this strategy constitutes a sub-game perfect equilibrium (SGPE), we need to determine P_x^*. Let us also assume that δ be the time discount factor of each supplier, $\delta \in (0, 1)$.

Present value of lifetime expected utility of an employed supplier, who has always been honest, is denoted by v_h

$$v_h = P_x + \delta(1 - \tau)v_h + \tau v_h^u \tag{1}$$

This equation implies that if someone is honest in the present period, she gets P_x. Next period, she may lose her job with exogenous probability τ or keep her job with probability $(1 - \tau)$. If she keeps her job, she gets the lifetime utility of an honest supplier. If she loses her job, she from the next period onwards, she gets the lifetime utility of an honest, unemployed supplier.

If someone cheats, she gets a one-time payoff of λ but then gets fired. Hence, from the next period gets the lifetime utility of a cheater supplier v_c^u. Hence,

$$v_c = \lambda + v_c^u \tag{2}$$

Note that the discount factor does not appear explicitly in Eqs. 1 and 2 because they are included in the definitions of v_h^u and v_c^u. Present value of lifetime expected utility of an unemployed supplier who has been honest and an unemployed supplier with a cheating history is given by

$$v_h^u = \delta q_h v_h + \delta(1 - q_h)(\bar{u} + v_h^u) \tag{3}$$

$$v_c^u = \delta q_c v_h + \delta(1 - q_c)(\bar{u} + v_c^u) \tag{4}$$

where q_h is the probability of hiring an honest supplier and q_c is the probability of hiring a cheater supplier. Equation (3) describes the lifetime utility of an honest supplier who lost her job due to some accidental reason. In the next period, she can be hired with probability q_h, remain honest, and enjoy the lifetime expected utility of an honest supplier. On the other hand, with probability $(1 - q_h)$ she may not be hired in the next period. In this case, he gets reservation payoff in this period, and from this period onwards, she enjoy the lifetime utility of a honest, unemployed supplier (v_h^u). In Eq. 4, we describe the lifetime utility of an unemployed, cheater agent who lost her job after cheating. In the next period, she can be hired with probability q_c, remain honest and v_h. On the other hand, with probability $(1 - q_c)$ she may not be hired in the next period. In this case, he gets reservation payoff in that period, and from this period onwards, she enjoy the lifetime utility of a cheater, unemployed supplier (v_h^u).

Note that we theoretically do not preclude the possibility that someone who gets fired for cheating can get a job in the future. However, in some special cases, the probability of hiring a supplier with cheating history (q_c) can be zero.

So a supplier will not cheat if

$$v_h \geq \lambda + v_c^u \tag{5}$$

This yields the condition

$$P_x \leq (T - \delta\tau Q_h)\Big[\frac{\lambda}{1 - \delta Q_c} + \delta\bar{u}\Big(\frac{E_c}{(1 - \delta Q_c)} - \tau E_h\Big)\Big] = P_x^* \tag{6}$$

where
$T = 1 - \delta(1 - \tau)$
$Q_i = \frac{q_i}{[1-\delta(1-q_i)]}$

$E_i = \frac{1-q_i}{1-\delta(1-q_i)}, i = h, c.$

For the MPS to be SGPE, entrepreneurs must find it in their interest to hire suppliers. On the equilibrium path, this condition can be stated as follows—the input price is low enough so that the entrepreneur finds it optimal to hire a supplier

$$P_x^* \leq \pi_y - \bar{\pi} \tag{7}$$

From Eq. (6), we get our first proposition:

Proposition 2.1 *The honesty inducing price for the input supplier is rising in the probability of hiring a cheater supplier.*

From this point, we make departure from Greif (1993), where only one social group was hiring from one pool of unemployed agents and sharing information about cheating incidence within that group. We introduced two groups of entrepreneurs in our model—H and L. Now, we see the implications of this formulation. Both the groups are hiring from the same pool of unemployed suppliers but only sharing the information about cheating incidence within their own groups. None of the group members hire a supplier who had cheated her own group members in the past. But an entrepreneur does not know if a supplier has cheated an entrepreneur from the other group. Alternatively, we can justify the formulation by assuming that the existing cultural norm makes the entrepreneurs from group $i (i = H, L)$ to sanction suppliers who cheated entrepreneurs from group i only and not suppliers who cheated entrepreneurs from group j.

Suppose there are M number of entrepreneurs and by construction at the end of each period τ fraction of suppliers lose their job even if they are honest. Among the entrepreneurs, M_H are from H group and M_L from L group and $M_H + M_L = M$. A is the total number of supplier in the economy, and out of them, τM is the number of people who lost their jobs because of accidental reasons. Hence, $(1 - \tau)M$ are the number of suppliers who are still in job and $[A - (1 - \tau)M]$ are the total number of job seekers in any period. For a cheater who cheated an entrepreneur from group i (we call him i, cheater $(i = H, L)$), he can find a job next period only with an entrepreneur from the j community $(j = H, L, i \neq j)$. So, now we have three relevant parameters q_c^H (probability of getting a job for a H cheater), q_c^L (probability of getting a job for a L cheater), and q_h (probability of getting a job for an honest supplier). Now, for H cheaters, they can only be hired by L entrepreneurs in jobs which got accidentally vacant. Hence,

$$q_c^H = \frac{\tau M_L}{[A - (1 - \tau)M]} \tag{8}$$

$$q_c^L = \frac{\tau M_H}{[A - (1 - \tau)M]} \tag{9}$$

$$q_h = \frac{\tau M}{[A - (1 - \tau)M]} \tag{10}$$

Note that the relevant probabilities are the ratios of jobs available for a certain type (H cheater, L cheater, or honest) and the total number of job seekers. This formulation embeds the assumption that if a supplier cheats an entrepreneur from group H, then in the next period he can only be hired by an entrepreneur from group L (and vice versa). Hence, for a H cheater, the only people who can employ him are the L entrepreneurs who lost their existing suppliers for accidental reasons. The number of such vacancies is τM_L. For a similar reason, the number of vacancies for an L cheater is τM_H. An unemployed, honest supplier, however, can be hired by entrepreneur from any group. Given that $M_H > M_L$, $q_c^H < q_c^L$. This implies that the honestly inducing price to be paid by an H entrepreneur is less than that to be paid by an L entrepreneur, i.e., $P_x^H < P_x^L$. But why do two prices exist in the market for inputs? In other words, why would anyone supply to the higher caste if they pay lower price? We close this gap by assuming that the number of entrepreneurs in the low caste groups is too small to employ all suppliers in the market. Now, the condition for entering business for the jth individual from the ith group is

$$\pi_y - P_x^i > \pi_j \tag{11}$$

If the distribution of the opportunity income is the same for both groups and represented by the distribution function Φ, then the probability that a H group entrepreneur will enter the business is $\Phi(\pi_y - P_x^H)$ while that for the L group is $\Phi(\pi_y - P_x^L)$. Because $P_x^L > P_x^H$, it must be the case that $\Phi(\pi_y - P_x^H) > \Phi(\pi_y - P_x^L)$. This implies that with only the informal contract enforcement in place, people from the H group (any group which are traditionally in business, e.g., upper caste) are more likely to enter the pool of entrepreneurs than their L group counterpart. From this, we get the following proposition.

Proposition 2.2 *Entrepreneurs from the minority group pay higher input price. This restricts the further entry of entrepreneurs from that group.*

Now, let us consider the introduction of formal institutions for contract enforcement. With the introduction of formal institutions for contract enforcement if an input supplier cheats the entrepreneur takes her to the court of law and the court finds her guilty with a probability. If she does not cheat, she gets P_x. If she cheats, with probability μ (which can also be seen as the quality of formal institutions), she gets caught and bears the punishment cost of κ. However, with probability $(1 - \mu)$ she does not get caught and enjoys λ. She decides to be honest if

$$P_x \geq \mu(-\kappa) + (1 - \mu)(\lambda) \tag{12}$$

This yields

$$P_x \geq \lambda - \mu(\kappa + \lambda) = P_x^F \tag{13}$$

The entrepreneur will offer the minimum price which means that the equilibrium input price under the formal contract enforcement is P_x^F. We find that price is falling in the quality of institutional quality. Now if the institutional quality is too high ($P_x^F < P_x^H < P_x^L$), all entrepreneurs will move to formal court in case of cheating and identity of the entrepreneur will not be a binding constraint in entry decision. If the institutional quality is too low ($P_x^F > P_x^H > P_x^L$) formal institutions will not matter. However, it is possible to have an intermediate case where the following condition holds:

$$P_x^H < P_x^F < P_x^L \tag{14}$$

Under this condition, L group entrepreneurs will sign formal contract to avail lower input prices than otherwise available to them but not the H group entrepreneurs. In this situation, any improvement in the formal institution will drive the price of input down, and as long as it remains higher than P_x^H, it will only benefit the L group entrepreneurs—with the rise in formal institutions, the number and profitability of L group entrepreneurs will increase.

3 Data

3.1 Data on Court Quality

One of the major contributions of the paper is to present district-level panel data for around 16 districts in West Bengal for the period 2000–2013 collected from Calcutta high court.[1] This is an important value addition as the existing official sources only publish state-wise data for such a long period. As we will see later, that even within West Bengal, there is a substantial variation in various court efficiency measures and this district-level variation in court quality can be exploited to find its effect on several outcome variables. The database has information on year-wise information regarding the number of civil cases pending at the beginning and end of the year, number of civil cases instituted during a year, and the number of civil cases disposed during a year for several year and districts. However, we do not know the nature of the cases. The data also provides information on the age breakup of pending cases, such as cases pending for 0–1, 1–5, 5–10, and more than 10 years. Table 1 gives a summary of these variables. Using this information, we construct various measures of court inefficiency.

[1]Out of a total of 19 districts, we obtained information only on 16 districts. More specifically, the data that we obtained from Kolkata High Court did not provide us information for three districts—Dakshin Dinajpur, Murshidabad, and Nadia.

Table 1 Summary: court statistics

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Pending in start of the year	224	30,513	34,609	1736	149,808
Filed during year	224	7301	6839	627	31,505
Solved during year	224	7035	6924	474	44,444
Pending in end of the year	224	30,779	34,936	1782	149,808
Pending for <year	224	6590	6221	613	26,119
Pending for 1–5 years	224	13,315	16,991	509	68,918
Pending for 5–10 years	224	7546	9044	215	40,498
Pending for >10 years	224	3323	4153	12.83	22,954

3.2 Measures of Court Quality

1. **Duration Index** Our main measure of court quality is the duration index that has been used in Chemin (2009b). This index measures the number of years that judiciary would take to address the existing caseload (sum of pending cases at the beginning and newly instituted cases). It is defined as the ratio of number of cases pending (plus the number of cases filed within the year) to the number of cases solved during a year. The measure is defined as λ_t

$$\lambda_t = \frac{p_t + f_t}{d_t} \tag{15}$$

where p_t is the number of cases pending at the beginning of the year t, f_t is the number of cases filed in year t, and d_t is the number of cases disposed within year t. We find considerable variation in terms of the duration index across districts. We normalize the duration index using the following formula and depict it on Indian map for the years 2000 and 2013 in Fig. 1:

$$\lambda_{td}^{\text{norm}} = \frac{\lambda_{td} - \lambda_t^{\text{Min}}}{\lambda_t^{\text{Max}} - \lambda_t^{\text{Min}}} \tag{16}$$

where λ_{td} represents the duration index for district d and year t, λ_t^{Min} represents the minimum value of λ across all districts in year t, and λ_t^{Max} represents the maximum value of λ across all districts in year t.

2. **Pendency clearance time** Our second measure-pendency clearance time [used in LCI (2014)] captures the time that will be taken to clear the existing backlog of cases in a particular period, if no more cases are accepted in court and the cases are disposed at the same rate. The measure is defined as ρ_t

$$\rho_t = \frac{p_t}{d_t}$$

Like duration, the second index is also an inefficiency index.

3. **Backlog creation rate** Our third indicator is backlog creation rate [also taken from LCI (2014)] which is the ratio of cases filed to cases disposed in any given year. It is given by γ_t where

$$\gamma_t = \frac{p_t}{d_t}$$

This is also an inefficiency index. A value greater (less) than 1 indicates more cases are being instituted (solved) than are being solved (instituted). The summary statistics of these three indices are presented in Table 2. It takes on average five years to clear the existing caseload consisting of the pending cases from earlier and the newly instituted cases and around 4 years to just clear the backlog of the pending cases.

Figure 1 shows how the courts are performing across the districts in West Bengal for two years, namely 2000 and 2013. Both the maps show that there is a considerable variation in this efficiency measure across the districts. On an average, we observe districts situated in the northern region to be consistently performing better than their counterparts.

More specifically, we noted North and South 24 Parganas had highest court inefficiency throughout the period. Also, on average the proportion of districts with inefficient courts have increased overtime. To be specific, Nadia, Purulia, Bardhaman, and Malda have added to the list of worst-performing districts over the years.

In Fig. 2, we see how different districts are faring in terms of court efficiency. For that, we plot district-wise difference of duration from the state average of duration for each year. The positive points indicate better performance when compared to the state while points below the zero scale indicate that the performance was below the state average in that district and year. To elaborate, we see that on average, Cooch Behar, Darjeeling, Jalpaiguri, Paschim Medinipur, and Uttar Dinajpur were performing better than the average for most of the years. Birbhum, Burdwan, Malda, and Purba mostly depicted an average performance. Finally, Bankura, Hooghly, Howrah, Kolkata, North and South 24 Parganas and Purulia were the least efficient districts in terms of court performance.

Table 2 Summary: court inefficiency

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Duration in the year of operations	224	5.021	1.605	1.772	12.32
Pendency clearance time	224	4.021	1.605	0.772	11.32
Backlog creation rate	224	1.073	0.211	0.385	2.158

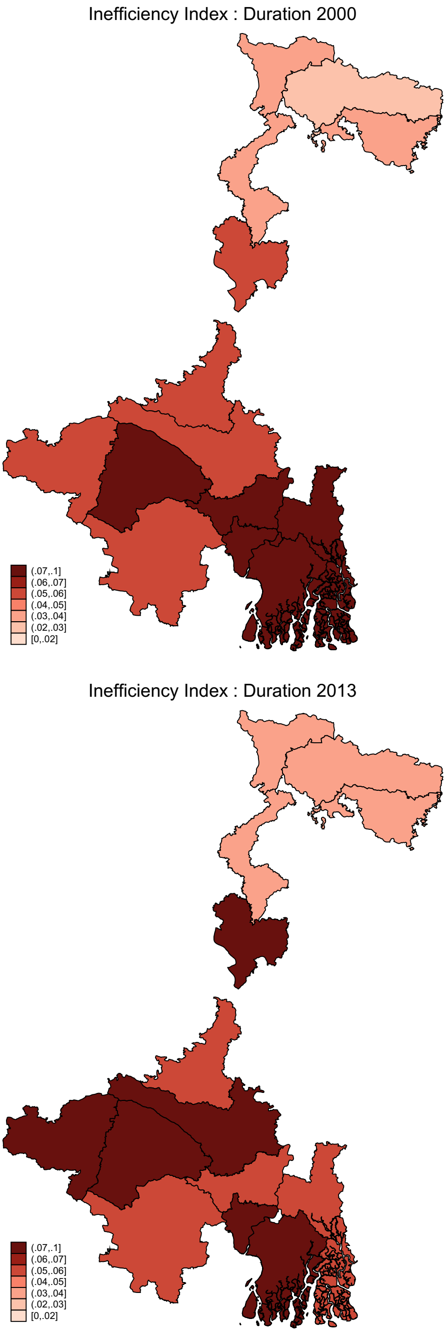


Fig. 1 Duration based inefficiency index for year 2000 and 2013

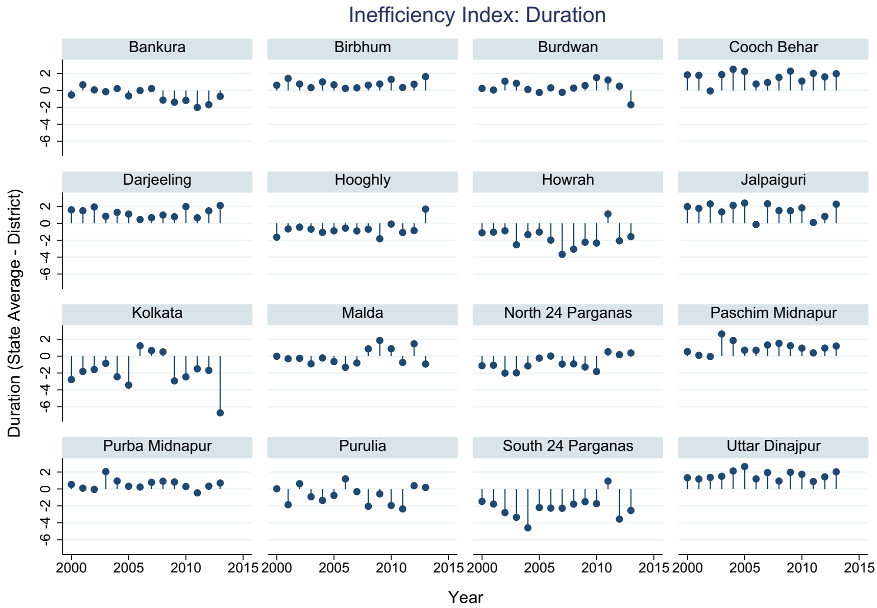


Fig. 2 Relative position of districts with relation to the state average of duration index

3.3 Industry Performance Data

We use industry data from Fourth All India Census for Medium, Small scale and Micro Enterprises (MSME data), for the year 2006–07 for our study. We have used data for only 16 districts of West Bengal for which we received data on court performance. The data set contains information on all those registered enterprises whose year of establishment ranges from 2000 to 2007. The survey round of firms is 2007. This implies that in 2007, we observe firms that came into operation in various preceding years including 2007. There are a total of 10,321 such firms in our sample. The MSME database consists of information on a number of firm parameters like year in which the firm starts production, year in which it gets registered, total number of employees, gender, religion, and social group of the firm owner. On the basis of this information, we construct three outcome variables reflecting various measures of business efficiency, namely delay in registration, same year registration, and firm size. We also look at the proportion of businesses operated by the disadvantaged castes and females. We do not have data for year of registration of some firms, and some firms have year of registration before they started production. Leaving out those firms, we have 7867 firms for which we have delay in registration and same year registration data.

From Table 3, we see that 38% of the enterprises which have started production in the period 2000–2007 chose to get registered in the same year in which they started production. Average number of employees in micro, small and medium enterprises

Table 3 Summary of business efficiency measures

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Delay in registration	7867	1.892	2.110	0	7
Same year registration	7867	0.380	0.485	0	1
Firm size	7867	7.117	15.16	1	648
SC or ST	7867	0.139	0.346	0	1
Female	7867	0.140	0.347	0	1

Table 4 Delay in registration

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Bankura	335	1.065672	1.763362	0	7
Birbhum	274	2.007299	2.086048	0	7
Burdwan	569	1.463972	1.966941	0	7
Cooch Behar	218	1.330275	2.01174	0	7
Darjeeling	345	2.576812	2.256911	0	7
Hooghly	400	2.3425	2.310528	0	7
Howrah	587	0.979557	1.44037	0	6
Jalpaiguri	567	2.063492	2.257189	0	7
Kolkata	878	2.173121	2.080393	0	7
Malda	1218	2.117406	2.089653	0	7
North 24 Pargana	1043	1.864813	2.146715	0	7
Paschim Medinipur	305	1.239344	1.687344	0	7
Purba Medinipur	303	2.485149	2.195482	0	7
Purulia	168	1.577381	1.946891	0	7
South 24 Pargana	437	2.199085	2.077445	0	7
Uttar Dinajpur	220	2.240909	2.506823	0	7
Total	7867	1.891827	2.109841	0	7

in West Bengal is 7. Also, on an average an enterprise gets registered within 2 years of initial production. Around 14% of the enterprises are owned by SC/ST and women entrepreneurs.

Next, we show that for each outcome variable, there is a significant variation across the districts. In Table 4, we see that the average delay in registration for the 2000–2007 period ranges from less than a year in Howrah to around 2.5 years in Darjeeling. The range is even wider for the same year registration variable in Table 5. For 2000–2007 period, in Purba Medinipur, on average around 24% of the firms got registered in the same year in which it started its production while this proportion is as high as around 61% in the district of Bankura. Mean total number of employees vary from

Table 5 Same year registration

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Bankura	335	0.61194	0.488037	0	1
Birbhum	274	0.335766	0.473122	0	1
Burdwan	569	0.448155	0.497742	0	1
Cooch Behar	218	0.536697	0.499799	0	1
Darjeeling	345	0.243478	0.429805	0	1
Hooghly	400	0.285	0.45198	0	1
Howrah	587	0.563884	0.496325	0	1
Jalpaiguri	567	0.372134	0.483801	0	1
Kolkata	878	0.284738	0.451547	0	1
Malda	1218	0.353038	0.478111	0	1
North 24 Pargana	1043	0.390221	0.488034	0	1
Paschim Medinipur	305	0.488525	0.50069	0	1
Purba Medinipur	303	0.240924	0.428352	0	1
Purulia	168	0.39881	0.491117	0	1
South 24 Pargana	437	0.254005	0.435799	0	1
Uttar Dinajpur	220	0.413636	0.493608	0	1
Total	7867	0.379687	0.48534	0	1

as little as 3 in Malda to as high as 17 in Birbhum as can be seen from Table 6. In Table 7, we observe that Jalpaiguri has the highest incidence of SC/ST ownership with 31% belonging to this group while Purulia with 2% ownership has the lowest incidence of minority ownership.

We have used ownership gender, religion, and social category as our controls. Whether a firm is managed by a male or a female is also controlled in the regression. The summary statistics of various control variables are presented in Table 8. As we can see, only 14% of the enterprises are owned by a female and only 9% are managed by a female. Thus, the industry is male dominated in terms of ownership and management. In case of social category, around 13% of the owners belong to Scheduled Caste and Scheduled Tribes, whereas only 5% belong to other backward castes. When we see owner religion, the industry is dominated by Hindu owners with more than 80% belonging to Hindu religion. It is followed by 15% Muslims, and other religions are in minority.

Table 6 Firm size

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Bankura	335	8.549254	31.58872	1	500
Birbhum	274	17.5146	28.0715	1	248
Burdwan	569	9.414763	15.20159	1	250
Cooch Behar	218	7.307339	13.56201	1	150
Darjeeling	345	6.530435	34.94497	1	648
Hooghly	400	8.4625	14.20098	1	200
Howrah	587	6.310051	7.953531	1	89
Jalpaiguri	567	5.839506	6.990958	1	85
Kolkata	878	8.013667	13.0138	1	247
Malda	1218	2.655993	3.128954	1	60
North 24 Pargana	1043	7.244487	11.50226	1	200
Paschim Medinipur	305	5.118033	6.597386	1	58
Purba Medinipur	303	6.914191	9.194328	1	59
Purulia	168	7.577381	7.843909	1	56
South 24 Pargana	437	10.64073	14.39312	1	110
Uttar Dinajpur	220	6	9.09664	1	70
Total	7867	0.379687	0.48534	0	1

4 Empirical Model

We estimate the following specification to test the extent to which the efficiency of formal legal institutions enhances the performance of small and medium enterprises.

$$\theta_{idt} = \alpha + \beta_1 X_{dt} + \beta_2 Z_{idt} + \beta_3 D_d + u_{idt} \tag{17}$$

θ_{idt} is our dependent variable. As mentioned before, from the MSME data, we have constructed four outcome variables namely delay in registration, same year registration, total number of people employed, and probability of an owner being an SC/ST. Delay in registration is constructed by taking the time gap between the year in which the enterprise starts production and the year in which it gets registered. In other words, it is measured by the number of months it takes a firm to register from the year it started production. Same year registration is a binary variable which takes the value of 1 if a firm gets registered in the same year as it starts production and 0 if a firm gets registered after it has started production. Total number of people employed in the enterprise as of 2007 is used as a proxy for enterprise size.

X_{dt} denotes the court efficiency index. We construct various measures of court efficiency on the basis of the information in our court quality database. For the baseline index, we use the duration index which is constructed by taking the ratio of

Table 7 Probability of SC or ST ownership

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
Bankura	335	0.095522	0.294375	0	1
Birbhum	274	0.083942	0.277808	0	1
Burdwan	569	0.070299	0.255875	0	1
Cooch Behar	218	0.192661	0.395296	0	1
Darjeeling	345	0.234783	0.424479	0	1
Hooghly	400	0.075	0.263721	0	1
Howrah	587	0.044293	0.205921	0	1
Jalpaiguri	567	0.313933	0.464499	0	1
Kolkata	878	0.063781	0.244502	0	1
Malda	1218	0.273399	0.445887	0	1
North 24 Pargana	1043	0.100671	0.301037	0	1
Paschim Medinipur	305	0.04918	0.2166	0	1
Purba Medinipur	303	0.042904	0.202977	0	1
Purulia	168	0.029762	0.170438	0	1
South 24 Pargana	437	0.16476	0.371389	0	1
Uttar Dinajpur	220	0.177273	0.38277	0	1
Total	7867	0.138554	0.345502	0	1

Table 8 Ethnic identity of entrepreneurs

Variables	(1)	(2)	(3)	(4)	(5)
	N	Mean	sd	Min	Max
SC	7867	0.125	0.331	0	1
ST	7867	0.0133	0.115	0	1
OBC	7867	0.0564	0.231	0	1
Hindu	7867	0.808	0.394	0	1
Muslim	7867	0.155	0.362	0	1
Sikh	7867	0.00140	0.0374	0	1
Jain	7867	0.00178	0.0422	0	1
Christian	7867	0.00483	0.0693	0	1

number of cases that court has at disposal in a certain year including both pending and newly instituted cases, and number of cases disposed of during the year. This indicates the speed at which a court disposes of available cases. To check for the consistency of our results, we also use two alternative measures of court inefficiency; namely, pendency clearance time and backlog creation rate Z_{idt} are the controls as discussed in the data section. The MSME data allows us to control for the caste, gender, and religion of the owner.

We also investigate whether an efficient judiciary facilitates the entrepreneurs from disadvantaged sections of the society in running their businesses. This is an important question to analyze as traditionally people from these backward castes could not rely on informal network for various business deals as it is primarily dominated by the individuals from upper caste. Hence, an improvement in formal judiciary is expected to lead to a higher proportion of SC/ST ownership of businesses. Hence, we estimate the following specification.

$$Y_{idt} = \alpha + \beta_1 X_{dt} + \beta_2 Z_{idt} + \beta_3 D_d + u_{idt} \tag{18}$$

$Y_i dt$ denotes the probability of the enterprise being owned by an individual from the disadvantaged classes namely SC/ST category. It is a binary variable which takes the value 1 when the owner belongs to SC/ST category and 0 otherwise.

5 Results

5.1 Baseline

We first look at the relation between legal efficiency and the extent of delay in registering a start-up. Table 9 reports the results from an OLS regression of number of years of delay in registration on our first court efficiency index. Note that this analysis is relevant only for the registered firms. The positive coefficient in column 1 suggests that it takes a longer for a firm to get registered in a district where courts are more inefficient.

In column 2, we control for the caste of the entrepreneur. It is possible that districts that have more efficient judiciary encourage higher participation of socially disadvantaged groups in entrepreneurship. However, it has to be noted that registration might be a costly procedure both in terms of information acquisition and pure monetary costs. Hence, socially disadvantaged entrepreneurs might take more time to afford the high costs as well as to gather the right information. However, if socially disadvantaged groups have less access to information than the other entrepreneurs, then they might take a longer time to get their firms registered. Controlling for social background increases the magnitude of the effect of a weaker judiciary on the extent of delay in registration.

For a similar argument, we control for the religion of the owner in column 3. However, the inclusion of the religion indicator does not affect the correlation between legal inefficiency and business inefficiency beyond the ones captured by caste.

In column 4, we control for the gender of the owner. Since labor force participation of men in general is much higher than that of women in India, it is possible that gender of firm owner affects the efficiency of the firm it-self which is captured in registration delay. On the other hand, gender of the owner might also be linked with the efficiency of judiciary in a district. For instance, formal institutions, courts

Table 9 Baseline regression

Dependent variable: delay					
Variables	(1)	(2)	(3)	(4)	(5)
	None	+Caste	+Religion	+Owner gender	+Manager gender
Duration in the year of operations	0.07146***	0.08450***	0.08648***	0.08584***	0.08933***
	(0.01578)	(0.01580)	(0.01580)	(0.01580)	(0.01579)
Owner characteristics					
SC		0.5901***	0.5614***	0.5654***	0.5562***
		(0.07205)	(0.07267)	(0.07268)	(0.07259)
ST		0.7945***	0.8233***	0.8066***	0.7641***
		(0.2066)	(0.2068)	(0.2069)	(0.2067)
OBC		0.1088	0.08274	0.08892	0.08486
		(0.1032)	(0.1035)	(0.1035)	(0.1034)
Hindu			0.1803***	0.1751***	0.1837***
			(0.06083)	(0.06086)	(0.06079)
Male				−0.1490**	0.1522*
				(0.06816)	(0.09148)
Male manager					−0.5494***
					(0.1115)
Constant	1.5219***	1.3637***	1.2126***	1.3475***	1.5660***
	(0.08506)	(0.08709)	(0.1009)	(0.1182)	(0.1261)
Observations	7867	7867	7867	7867	7867
R-squared	0.003	0.013	0.014	0.014	0.017
District FE	NO	NO	NO	NO	NO
Time FE	NO	NO	NO	NO	NO

Robust standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

in this case, might facilitate business operations more for women entrepreneurs, who are less likely to get the benefits from informal networks dominated by men. For a similar reason, we control for the gender of the manager of a firm in column 5. We find that the magnitude of the effect increases further with a control for the gender of the manager and owner.

5.2 *Various Business Efficiency Measures*

Overall, the results in Table 9 suggest that a more efficient judiciary helps businesses to operate in the formal sector, possibly by reducing the cost of formalization. To test this further, in Table 10 we include a few more measures of business efficiency. For comparison, column 1 of Table 10 reports the results on delay in registration from the full specification. Column 2 looks at the probability of getting registered in the same year as the start of production. The negative coefficient suggests that a more efficient court system helps businesses to register immediately and operate in the formal sector.

In addition to the threshold of registration, an important decision for an enterprise is the scale of operation. Since larger organizations are more likely to face contract enforcement problems, a more efficient legal system is likely to help in the expansion of business. Note, however, that the dependence of a firm on the formal judiciary for

Table 10 Alternative business efficiency measures

Variables	(1)	(2)	(3)
	Delay in registration	Same year registration	Firm size
Duration in the year of operations	0.08933***	−0.02070***	0.01499
	(0.01579)	(0.003639)	(0.1141)
Owner characteristics			
SC	0.5562***	−0.07455***	−3.5480***
	(0.07259)	(0.01673)	(0.5244)
ST	0.7641***	−0.1414***	−2.7628*
	(0.2067)	(0.04764)	(1.4934)
OBC	0.08486	0.02316	−1.8080**
	(0.1034)	(0.02382)	(0.7467)
Hindu	0.1837***	−0.05953***	−0.3058
	(0.06079)	(0.01401)	(0.4391)
Male	0.1522*	0.008293	−0.1343
	(0.09148)	(0.02108)	(0.6608)
Male manager	−0.5494***	0.09866***	0.1186
	(0.1115)	(0.02570)	(0.8055)
Constant	1.5660***	0.4478***	7.8772***
	(0.1261)	(0.02907)	(0.9111)
Observations	7867	7867	7867
R-squared	0.017	0.014	0.007
District FE	NO	NO	NO
Time FE	NO	NO	NO

Robust standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

contract enforcement of any sort is only true for firms registered under the companies' act. To test this hypothesis, column 3 reports the link between court efficiency and the number of employees in registered firms. As expected, we find that firms employ less when the judiciary is less efficient.

5.3 Robustness

In the analysis so far, we used the inefficiency index duration which we have already explained in the previous section. Now for robustness check, we use all other measures that we have described there. They all come from Chemin (2009b) and Chemin (2010). We test the sensitivity of our results in Tables 11 and 12 to these alternative measures of efficiency. Tables 11 and 12 report the results from using our second and third index, respectively. While the second index also reflects inefficiency of the

Table 11 Robustness check: pendency clearance time

Variables	(1)	(2)	(3)
	Delay in registration	Same year registration	Firm size
Pendency clearance time	0.08933***	−0.02070***	0.01499
	(0.01579)	(0.003639)	(0.1141)
Owner characteristics			
SC	0.5562***	−0.07455***	−3.5480***
	(0.07259)	(0.01673)	(0.5244)
ST	0.7641***	−0.1414***	−2.7628*
	(0.2067)	(0.04764)	(1.4934)
OBC	0.08486	0.02316	−1.8080**
	(0.1034)	(0.02382)	(0.7467)
Hindu	0.1837***	−0.05953***	−0.3058
	(0.06079)	(0.01401)	(0.4391)
Male	0.1522*	0.008293	−0.1343
	(0.09148)	(0.02108)	(0.6608)
Male manager	−0.5494***	0.09866***	0.1186
	(0.1115)	(0.02570)	(0.8055)
Constant	1.6553***	0.4271***	7.8922***
	(0.1162)	(0.02678)	(0.8395)
Observations	7867	7867	7867
R-squared	0.017	0.014	0.007
District FE	NO	NO	NO
Time FE	NO	NO	NO

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 12 Robustness check: backlog creation rate

Variables	(1)	(2)	(3)
	Delay in registration	Same year registration	Firm size
Backlog creation rate	0.5903***	−0.08996***	−2.8731***
	(0.1013)	(0.02336)	(0.7308)
Owner characteristics			
SC	0.5080***	−0.06490***	−3.4601***
	(0.07248)	(0.01672)	(0.5231)
ST	0.7003***	−0.1266***	−2.7723*
	(0.2064)	(0.04763)	(1.4897)
OBC	0.08061	0.02662	−1.9653***
	(0.1033)	(0.02384)	(0.7456)
Hindu	0.1925***	−0.05966***	−0.4248
	(0.06087)	(0.01404)	(0.4393)
Male	0.1612*	0.008705	−0.2920
	(0.09154)	(0.02112)	(0.6607)
Male manager	−0.5591***	0.09789***	0.3082
	(0.1116)	(0.02574)	(0.8053)
Constant	1.3777***	0.4389***	11.184***
	(0.1476)	(0.03406)	(1.0655)
Observations	7867	7867	7867
R-squared	0.017	0.012	0.009
District FE	NO	NO	NO
Time FE	NO	NO	NO

Standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

judiciary, the third index captures efficiency. Overall, the results are robust to the use of these alternative court indices. Delays in firm registrations are lower, and the proportion of firms registering in the same year when it starts production is higher in districts with higher court efficiency. Moreover, the firm size is higher in districts with higher court efficiency.

5.4 *Business Ownership: Caste*

5.4.1 Caste

Given the historical occupational bindings of the caste system in India, it is well established that occupational mobility across different caste groups is typically very low. This means the historical presence of marginalized social groups such as Scheduled Caste or Scheduled Tribe and women are very low in the business. In a setup

Table 13 Probability of business ownership

Variables	(1)
	SC/ST
Duration in the year of operations	−0.01749***
	(0.002570)
Owner characteristics	
Hindu	0.09215***
	(0.009811)
Male	0.04220***
	(0.01495)
Male manager	−0.05831***
	(0.01822)
Constant	0.1715***
	(0.02044)
Observations	7867
R-squared	0.019
District FE	NO
Time FE	NO

Standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

like this, such business minority groups will find it difficult to enforce contracts using reputation mechanism. The rationale behind such argument is already discussed in the section on theoretical structure.

Therefore, formal judicial system has a critical role to play in the business success of the business minorities such as SC/ST and women. For the logic we detailed above, formal judiciary is more important for SC/ST entrepreneurs than their general caste counterpart. The anecdotes regarding experiences of successful Dalit entrepreneurs like Devanand Londhe and Ratibhai Makwana hint at such favorable role of formalization. Their transition from "job-seekers" to "job-givers" would not have been possible without improvements in court quality. Hence, we argue that a formal judicial system is likely to be more helpful for disadvantaged sections of the society who traditionally do not have a very strong informal network to bank on. Table 13 reports the results from the OLS regression of the proportion of SC/ST ownership on duration index measuring court inefficiency. Specifically, we test that in the presence of efficient judiciary whether registered entrepreneurs are more likely to be from the Scheduled Caste or Scheduled Tribe compared to the General or OBC category. The negative coefficient indicates that more efficient courts encourage a higher participation of disadvantaged groups in operating an enterprise.

6 Conclusion

Enforcing contract is critical for business proliferation. However, in less-developed countries with inefficient courts, the informal networks are ubiquitous for enforcing contracts. The prohibitively high litigation costs in India make moving to court for resolving disputes the last option for an entrepreneur. But theoretically, people choose not to move to court under two types of circumstances when courts are very good and nobody breaches a contract, and when the courts are so bad that going to court cannot provide a remedy. Therefore, the direct effects of these two different qualities of the court system are the same on the number of litigation (low in both cases). However, the shadow effects of differing court quality will be different for the number of contracts. There will be more contracts signed under a good court regime than a bad court one. In this paper, we try to estimate this shadow effect by exploiting the district-level court quality variation within the state of West Bengal. We find that the shadow indeed works districts with better court efficiency have bigger firm sizes than the districts with inefficient courts. Moreover, firm registrations are larger in districts with higher court efficiency. More importantly, we find that socially disadvantaged groups such as Scheduled Caste/Tribes benefit more than their upper caste counterpart. We find the districts with better court have more business run by SC/ST than the districts with inefficient courts. To summarize, we find significant impact of formal court system on entrepreneurship. Our findings suggest that improving formal court helps entrepreneurs in general, but the effect is stronger for socially disadvantaged groups such as Scheduled Castes and Tribes.

Acknowledgements We are grateful to International Development Research Center (IDRC), Canada, for funding our research.

References

Acemoglu, D., Antràs, P., & Helpman, E. (2007). Contracts and technology adoption. *The American Economic Review*, *97*, 916–943.

Acemoglu, D., & Johnson, S. (2003). Unbundling institutions. Technical report, National Bureau of Economic Research.

Ahsan, R. N. (2013). Input tariffs, speed of contract enforcement, and the productivity of firms in india. *Journal of International Economics*, *90*(1), 181–192.

Amirapu, A. (2015). Judicial institutions, relationship specificity and growth: Evidence from India. Technical report, working paper.

Chemin, M. (2009a). Do judiciaries matter for development? Evidence from India. *Journal of Comparative Economics*, *37*(2), 230–250.

Chemin, M. (2009b). Do judiciaries matter for development? Evidence from India. *Journal of Comparative Economics*, *37*(2), 230–250.

Chemin, M. (2010). Does court speed shape economic activity? Evidence from a court reform in India. *Journal of Law, Economics, and Organization*.

Chin, A., & Prakash, N. (2011). The redistributive effects of political reservation for minorities: Evidence from India. *Journal of Development Economics*, *96*(2), 265–277.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *The American Economic Review*, *83*(3), 525–548.
Hazra, A. K., & Micevska, M. B. (2004). The problem of court congestion: Evidence from Indian lower courts. Technical report, ZEF discussion papers on development policy.
Hnatkovska, V., Lahiri, A., & Paul, S. (2012). Castes and labor mobility. *American Economic Journal: Applied Economics*, *4*(2), 274–307.
Hnatkovska, V., Lahiri, A., & Paul, S. B. (2013). Breaking the caste barrier intergenerational mobility in India. *Journal of Human Resources*, *48*(2), 435–473.
LCI. (2014). Arrears and backlog: Creating additional judicial (wo)manpower.
Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, *74*(3), 791–819.
Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 569–600.
Rabiyath, S., & Rupakula, R. V. (2010). Disposal rates, pendency, and filing in Indian courts: An empirical study of the two states of Andhra Pradesh and Kerala. In J. Babu, P. G. T. Eger, A. V. Raja, H.-B. Schafer, & T. S. Somashekar (Eds.), *Economic analysis of law in India: Theory and application*. New Delhi: Oxford University Press.
Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 433–444.
Sinha, R. (2018). Pendency of cases in the judiciary. Technical report.

Does Social Connectivity Influence Tap Water Access? Evidence from India

Prasenjit Sarkhel and Subhalakshmi Paul

Abstract Contemporary assessment of water availability in India predicts that by 2020 close to 600 million Indians would be under water distress. The threat is more potent for the rural households as more than 80% of them are yet to have tap water within their premise. Public authorities have scaled up the rural water supply schemes and have set the target of universalizing indoor tap water in rural areas by 2024. In this background, using a panel of rural household water use data from 2005 and 2012 rounds of India Human Development Survey (IHDS), this paper attempts to empirically investigate whether the extent of social network influences the households' access to the public water supply via tap water connection. Our paper shows that even in water-scarce areas the planners might fail to tap the potential demand for tap water if community ties are weak and households are not well integrated into social network. We find that if access to public water schemes is contingent on the intensity of social ties, it might exclude asset poor and socially disadvantaged groups from its ambit. Our result, thus, suggests that strengthening networks including poor households and scaling up of information and communication activities might be effective strategies to ensure increased access to piped water.

Keywords Drinking water · India human development survey · PROBIT · Recursive bivariate PROBIT

JEL Classifications C35 · H44 · Q58

P. Sarkhel (✉)
Department of Economics, University of Kalyani, Kalyani, India
e-mail: prasenjitsarkhel@gmail.com

S. Paul
Department of Economics, Anandamohan College, Kolkata, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_10

1 Introduction

Recent assessment of water availability in India finds that close to 80% of the rural populace are yet to have access to indoor piped water. Rather there is an increasing reliance on groundwater and that is also getting depleted at an alarming rate.[1] In addition, poor maintenance of surface water sources and inadequate storage facilities of rainwater has further worsened the scarcity of potable drinking water. Public agencies like Niti Ayog have recently issued warning that over 100 million households would be denied access to potable water by 2020. Early signals of such catastrophe became visible when states like Tamilnadu, in Southern India, found the local aquifers dry in the month of June 2019, and the government had to resort to emergency transfer of drinking water to the citizens from the neighbouring areas. The government has responded to this crisis by forming a new ministry called *Jal Shakti* and have further announced a scheme called *Jal Jeevan Mission* that would ensure universal access to indoor tap water by 2024 to all the rural households. In an effort to scale up the water supply initiatives, the government has allocated INR 10,000 crore to National Rural Drinking Water Programme (NRDWP) in the union budget of 2019–20, which is almost double the previous year's allocation.[2] The NRDWP was launched in 2009 by subsuming the earlier schemes of drinking water supply.[3] Two major features distinguish NRDWP from its predecessors: first, the unit of coverage has shifted from habitation to household under NRDWP and secondly, by design, it is a demand-driven scheme and requires coordinated action among various tiers of local institutions. As per guidelines of the NRDWP, water-scarce habitations are listed by the Ground Water Survey and Development Authority (GSDA) and this is passed on to the district collector. The habitations themselves can also register their demand through the Gram Sabha in the Gram Panchayat (Prasad et al. 2014). In short, the revised emphasis of NRDWP is more towards decentralized delivery of water by greater involvement of the community in planning, operation and management of in-village water schemes through formation of institutions like Village Water and Sanitation Committee (VWSC) (MDWS 2013). In this backdrop, our paper provides evidence that even in water-scarce areas the planners might fail to tap the potential demand for tap water if community ties are weak and households are not well integrated into social network. In other words, the availability of public water schemes would not automatically translate into access to indoor tap water.

---

[1] Globally, India is the highest extractor of groundwater and uses 250 cubic kilolitres of water per annum and has around 33 million bore wells. This amounts to almost one-fourth of the total global use of groundwater.

[2] https://economictimes.indiatimes.com/news/economy/policy/piped-water-to-rural-households-by-2024-finance-minister-nirmala-sitharaman/articleshow/70092345.cms.

[3] The initiative to provide water in rural areas was first spelt out by the Bhor Committee in 1949 and only a year later the constitution of India specified water as a state subject. Over time, National Rural Drinking Water Mission (NRDM) was launched in 1986 to accelerate the provision of drinking water in rural areas the first National Water Policy was drafted in 1987. NRDM was subsequently renamed as Rajiv Gandhi National Drinking Water Mission (RGNDWM) in 1991.

In fact, the records for NRDWP in terms of indoor tap water connection are far from satisfactory. For instance, the Comptroller and Auditor General (CAG) of India reviewed the performance of NRDWP during the period 2012–2017 and found that against the targeted connection of 35% of rural households only 17% were provided tap water connection within the stipulated time frame. During this period, the coverage of the rural habitations increased by only 5.5% at 55 litres per capita per day (lpcpd) when the target was to serve 50% of the rural population at this rate. Such underprovision occured despite the fact that 90% of the money allocated for the program has been spent during the period.[4] India is one of those countries where under-five mortality from diarrheal disease is high. The paucity of potable water, resulting from such underperformance of the public water schemes, one thus poses a potential threat to the public health, especially for the rural households.[5] There are other associated costs if, in the absence of piped water, households have to meet their water requirements from distant source. Allocating higher time or resource to fetch quality water from a distant source would involve tightening of time-income budget constraint for the household and can lead to offsetting changes that may outweigh health benefits from improved source. Studies have shown that the burden of water collection falls disproportionately on the women and children, and this distorts their time allocation for the labour market and school going, respectively (Blackden and Wodon 2006; Ilah 2000). In India, 63% of the rural households are still to have *any* water source within their premise and one-fifth of the rural households have to travel more than half a kilometre to fetch water. Further, one-fourth of the rural households spend at least an hour to reach and collect water from the chosen source (Tewari & Bapat 2019).[6] Evidently, understanding the barriers to piped water connection is a much-needed policy input for the effective implementation of programmes like NRWDP.

Extant investigations often ascribe supply-side issues of the inadequate access to improved water sources via piped water connection. This involves leakages in pipelines and poor quantity of water supplied, especially in the summer season, as well as poor maintenance. Interestingly, a study covering 600 rural water schemes in 10 states also find that in addition to inadequate water supply, there are wide-ranging disparities in allocation of water supply schemes (World Bank 2008). In states like Tamilnadu, Uttar Pradesh and Karnataka, there were over-provisioning of water sources and on an average ten households are using the stand posts in these states as against the norm of 50 households per water source. Such overprovision reduces the opportunity cost of water collection, but at the same time escalates the per capita expenditure and might dominate the economic impact of time savings from the former. Using tap water access data from census 2001 Balasubramaniam et al. (2014) shows that over-provision of water schemes might be chosen by the government in response to the religious fragmentation and the potential strife and conflict in those

[4]Out of the total allocation of INR 89,956 crores the total expenditure was INR 81168 crores.

[5]Diarrhoea is the third leading cause of death in India and is responsible for 13% of all under-five mortality per annum (Lakshminaryanan and Jayalakshmy 2015).

[6]https://scroll.in/article/810363/the-great-indian-water-walkathon.

areas. In such a case, the government might transfer the disproportionate amount of public amenities to offset welfare loss from social unrest. Our study contributes to the growing interest in water supply in rural India by providing evidence that controlling for religious and caste identities households' informal contact in government network significantly influences the likelihood of tap water connection. We use a panel of over 20,000 rural households from two rounds of nationally representative India Human Development Survey (IHDS) to estimate the determinants of tap water connection. In addition to social connectivity, our study also finds that the degree of exposure to media might be a significant predictor of piped water connection. Our result thus supports the government initiatives to promote information communication and education (IEC) through mass media and print media allocating 60% of the funds to village-level IEC activities. Besides addressing the issue of water supply in particular our research is also a part of the broad literature that examines the interlinkage between politics and public good allocation in rural India. Studies in the Indian context have found that public goods might be subjected to elite capture (Bardhan and Mookherjee 2000) and the supply of public goods can be severely constrained by patronage and clientelism (Anderson et al. 2015). Our results suggest that weak social contacts can be a significant factor in explaining poor level of access to potable water among economically backward and socially disadvantageous class.

The rest of the paper is organized in the following schema: Sect. 2 discusses the pattern of piped water connection in India and locates the issue in the existing literature on access to drinking water. The data and the estimation strategy are discussed in Sect. 3. We discuss the results in Sect. 4 and finally offer our concluding observations in Sect. 5.

2 Access to Piped Water

2.1 *Tap Water Coverage in India*

Access to piped water in rural India has lagged behind urban agglomerates in the last decades. The distribution of tap water connection across rural and urban India shows that more than 50% of the household in urban areas could access water within their premise while the corresponding figure in rural areas is 14%. In fact over the census period 2001 and 2011, there has been a marked increase in tap water connection in urban areas but improvements in rural areas were scant. In both rural and urban areas, tap water has increased both within premise and near premises but the rate of increase is much higher for urban areas (Fig. 4). However, there are considerable within state variation in tap water connection as well.

Figure 5 in Appendix indicates that tap water coverage for any of the state did not exceed 20% across both the census. The scenario is even worse for the percentage share of the in-house tap water connection, starting from 0.05 for Manipur to 17% for Maharashtra, respectively.

2.2 Determinants of Tap Water Use

In areas like rural India where there are a large number of households that are still to be connected with the public water network, an important issue is their decision regarding the choice of water source. In particular, whether or not the household decides to use public water sources is an important datum for the policy makers (Nauges and Whittington 2009). Past research indicates that educational status and income of the households have played an important role (Mu et al. 1990; Madanat and Humplick 1993; Larson et al. 2006; Nauges and Strand 2007; Nauges and van den Berg 2009) in having access to tap water. High education level raises the probability of using tap water for everyday routine activities reflecting a higher consciousness about the health and hygiene. Further, households with higher income tend to favour piped water and motor pumps over hand pumps as they might have higher ability to pay to afford the installation cost of the water systems (Madanat and Humplick 1993, Persson 2002).

However, one problem of using observational data on tap water connection might arise from unobservable that determine the allocation of water supply networks. For instance, as Jessoe (2013) notes that public taps and tube wells might be intentionally placed in areas that are weak in infrastructure and health amenities. Another strand of literature considers public good that is delivered by the political actor with discretion (Besley et al. 2004). As water is mostly supplied and regulated by public authorities in India, these concerns apply to tap connection as well. Betancourt and Gleason (2000) find that districts with higher proportion of Muslims and disadvantaged class received lower health and educational inputs in rural India. They also report that rural areas of more urbanized districts receive comparatively higher number of medical professionals probably due to the lower cost of provision. Similarly, Banerjee and Somanathan (2007) also find that ethnic heterogeneity influenced provision of public good in India; in particular, they found high-caste areas receive better educational facilities and post offices compared to areas that are dominated by disadvantaged population. These studies have examined the allocation of public goods but one important area of concern are the constraints that might prevent households from getting access to the public good even when it is available. Recent experiences with rationing of wage employment programmes like Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS) suggest that the extent of households' social connectivity can significantly influence the share of public goods.[7] In fact, the social ties and group memberships might provide a community institution that facilitates information sharing and could reduce transaction costs associated with water connection. This might involve getting updated about government schemes and minimizing queuing time at public offices.

Admitting the influence of social ties in tap water access would also mean that there could be a possibility that economically affluent and socially connected individuals

[7] Huge body of evidence exists for wage employment programmes like National Rural Employment Guarantee Programme (NREGA) that having political connection enhances the chance of securing jobs, especially in the aftermath of shock events (Sarkhel 2013; Dutta et.al. 2014).

might self-select themselves in water supply network. As beneficiaries of the water scheme must pay 10% of the total capital investment in NRDWP decision to connect to piped water is made at the household level and is not automatically guaranteed if the water supply scheme is allotted to the village (Vanaja 2018).[8] Thus, it also needs to be investigated if the access to public water supply is skewed against economically poor and socially disadvantaged class who might also have a lower endowment of social contact due to his poor asset base.

3 Data and Methods

3.1 Tap Water Use in IHDS

The information regarding the choice of water source for the households is extracted from two waves of India Human Development Survey Data (IHDS). The IHDS-I (2004–05) covered 33 states and territories within which it interviewed 41,554 households in 1503 villages and 971 urban neighbourhoods across India. For the next round (2011–12), 42,152 households were surveyed within which two thousand households are newly added. The re-contact rate for these two IHDS round is 83% overall, whereas it is 90% for the rural zone and 72% for the urban areas. In our analysis, we use the information of those households that are present in both the rounds.

The IHDS data records the principal source of drinking water chosen by the households in both the rounds. The survey recorded 11 principal sources that include piped water, groundwater sources like tube well and hand pump as well as open wells. Across the states, less than 50% of the households consider piped water as their principal source of drinking water followed by hand pump and tube well (Fig. 1). For households fetching drinking water from outside sources, groundwater remained the dominant source in the rural areas in both the rounds. On the contrary, urban households mostly collected water from piped water indicating greater proliferation of tap water in the urban community. As far as the inside premise water source is concerned, groundwater is the dominant source in rural areas (Fig. 2). However, the share of households using piped water indoors registered much higher increase across the two rounds in the rural areas compared to the urban areas. In 2005, 30% of the sampled households had piped water within their premise when the corresponding figure for urban areas was close to 70%. In 2011, 39% of rural households had tap water connection while in urban areas the figure increased to 72% of the households.

In this study, our aim is to analyse indoor tap water connection as well as its change over the years. In doing so, we concentrate only on the rural subsample of the households. We first take all the rural households in each round and identify the factors that determine tap water access. Second, we follow Coffey et al. (2017) and

[8] https://ageconsearch.umn.edu/record/275954/files/2393.pdf.

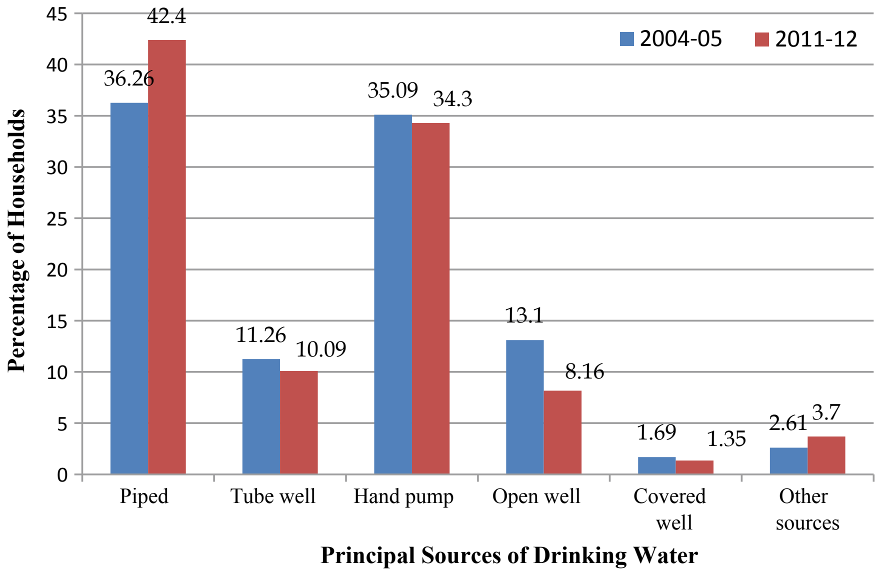


Fig. 1 Principal drinking water source. *Source* IHDS I & II

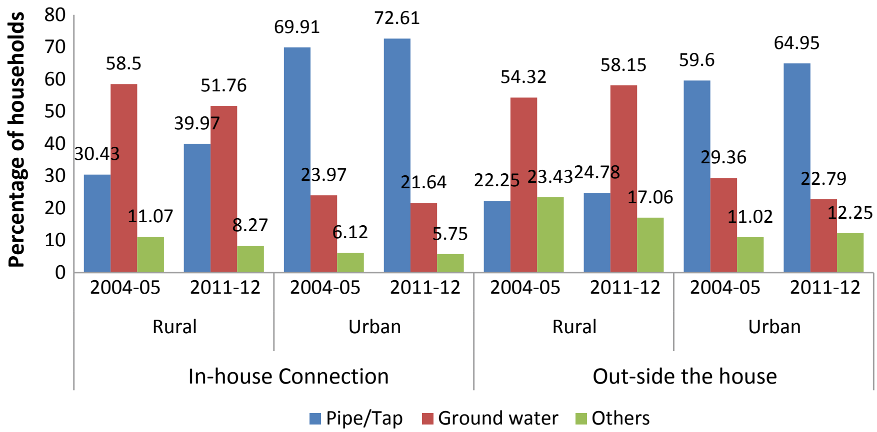


Fig. 2 Distribution of principal drinking water sources across rural and urban areas. *Source* IHDS I & II

take only those rural households that did not have tap water connection in round 1 and analyse their switching probabilities in round 2 of the IHDS survey. For the PROBIT estimation we have 28,227 and 27,308 rural households for round I and II respectively. After accounting for the missing observations the final estimation sample got reduced to 19,549 in round I and 22,434 in round II. In the switching model, we consider 23,594 rural households that did not have piped water connection in the round I. The final estimation sample for the switching households accounting for the missing values of the explanatory variables came to 14,490 households.

3.2 Estimation Strategy

We thus estimate two sets of variables: first is a binary one that takes on a value of 1 if households own a tap water inside premise and 0 otherwise. We estimate tap water connection separately for both the rounds of the IHDS. We expect that measures of social connectivity and exposure to IEC programmes would positively influence the tap water connection. The IHDS captures social integration of the households in terms of memberships in nine types of organizations[9] and network contacts with educational, medical and government institutions. As argued earlier, water supply is a public good in rural India so we take network contacts in government institutions as the relevant measure of social connectivity. In other words, we chose to investigate the extent to which informal network with government officials rather than formal memberships in groups or organizations influence the chances of availing piped water connection. Our choice is guided by the fact that membership in formal organizations might be contingent on their availability in the neighbourhood, whereas informal networks of the households are likely to be more widespread.[10] In fact, network contacts are assumed to interact more intensely with local contexts to influence the nature of social relationships and consequent access to public goods. For IEC impact, we consider households exposure to three mass media: radio, television and newspaper. In addition, we take controls for household wealth and education.

To account for water availability in the district, we include average depth of aquifer (in metres) in each district obtained from central groundwater commission. Average aquifer depth should denote the depth that must be reached to construct a tube well (Jessoe 2013). We expect this to characterize the trade-off between the decisions to opt for tap water vis-à-vis groundwater sources. If groundwater levels are lower, we expect that the cost of extracting that would be higher and people would have more incentive to switch towards piped water connection. We use groundwater-level data fetched from the India WRIS portal[11] for district level. WRIS provides information

[9]This includes trade unions, caste- and religious-based organizations, non-governmental organizations (NGOs) and local government.

[10]https://ihds.umd.edu/sites/ihds.umd.edu/files/publications/pdf-files/11HDinIndia.pdf Desai et. al. (2010).

[11]http://www.india-wris.nrsc.gov.in/wrpinfo/index.php?title=Main_Page.

for aquifer depth across districts for four seasons: Monsoon, Post-monsoon-Kharif, Post-monsoon-Rabi and Pre-monsoon. We use the minimum aquifer depth out of these four recorded levels as it is likely to indicate the minimum cost people must incur to extract groundwater for domestic use or drinking purposes. The state-wise variation of mean groundwater level reflects higher exhaustion of groundwater in Uttarakhand (possibly due to huge depletion of groundwater for farming) and Gujarat and Rajasthan that are known to be water-scarce areas (See Fig. 6 in Appendix). We use a logarithmic transformation of the groundwater variable in the regression analysis.

We denote the household's propensity to avail piped water at home by the latent variable y_{it}^* that is related to social network, groundwater levels and other household characteristics as:

$$y_{it}^* = \beta_t S_{it} + \theta_t G_{jt} + \gamma_t X_{it} + u_{it} \tag{1}$$

We further define $y_{it} = \begin{cases} 1 \text{ if } y_{it}^* > 0 \\ 0 \text{ if } y_{it}^* \leq 0 \end{cases}$ where, y_{it} is household's observed access to indoor piped water. Here, y_{it} takes a value of unity if the ith household owns a tap water at time t and 0 otherwise, $t = 1, 2$ denotes first and second round of IHDS, S_{it} is the variable of social connectivity indicating contact with government network that is binary and G_{jt} represents the level of groundwater in the jth district where the households resides. The vector of household characteristics like wealth, education, caste and religion is given by X_{it}. We recognize that the PROBIT model would only identify the factors that are associated with tap water use and may not identify the causal factors behind tap water connection. This is because we do not have any information about the time when household got their tap water connection and whether observed government network or contacts precedes household's connection or otherwise. As a consistency check for the estimates of Eq. (1), we take advantage of the panel nature of the data and verify whether the hypothesized relationship between social connection and media exposure does actually initiate the change in tap water connection across the two rounds by considering only the subsample of households that did not have tap water in the first round. We analyse the newly connected household vis-à-vis those households in this subsample that remained without indoor tap water in second round. Our aim is to estimate the switching probability and investigate whether the characters that displayed significant relationships with tap water connection is also significant in this equation. In particular, we are interested in observing whether changes in the level of social (ΔS_{ids}) connectivity and media exposure explain the changes in tap water connection (Δy_{ids}) of overtime controlling for wealth and groundwater levels. The equation of interest is

$$\Delta y_{ids} = \beta \Delta S_{ids} + \theta \Delta G_{ds} + \gamma \Delta X_{ids} + \mu F_{ids}^{2005} + \varepsilon \tag{2}$$

Here, Δ denotes the difference operator and F_{ids}^{2005} denotes the baseline (2005) values of the variables of interest, i.e. social connectivity, media exposure, aquifer depth as well as education and dwelling characteristics for the ith household in the dth district and sth state.

However, government networks or contacts may not be exogenous and might be simultaneously determined with tap water connection and indoor piped water might also signal his social status.[12] Higher economic status might enable households to have more contacts, but at the same time better contacts in government could increase job opportunities and enhance income earning potential. If this holds true, then our variable of government network becomes a dummy endogenous variable. In such cases, our PROBIT estimates in (1) would be potentially biased due to possible overlap in unobserved characteristics that influences access to piped water and government network. Thus, we have a system of PROBIT equations that allows the error terms to be correlated and the binary dependent outcome variable in one equation becomes the endogenous regressor in the other equation. This is known as recursive bivariate probit model (RBP) and have found wide application in explaining labour market participation (Morris 2007), the impact of job training programme on employment (Falk et al. 2006) and health issues like choice of private or public hospitals for dialysis treatments (Gitto and Sobbrio 2006). We then specify the model for each round as

$$\begin{gathered} S_{it}^{*} = \alpha_t Z_{it} + \pi_t G_{jt} + \delta_t X_{it} + v_{it},\ S_{it} = 1 \text{ if } S_{it}^{*} > 0,\ S_{it} = 0, \text{ otherwise} \\ y_{it}^{*} = \beta_t S_{it} + \theta_t G_{jt} + \gamma_t X_{it} + u_{it},\ y_{it} = 1 \text{ if } y_{it}^{*} > 0,\ y_{it} = 0, \text{ otherwise} \\ \left[u_{it,} v_{it,}\right] \sim \Phi_2[(0,0),(1,1),\rho],\ \rho \in [-1,1] \end{gathered} \tag{3}$$

In the RBP model, two error terms u_{it} and v_{it} are described by Φ_2—bivariate standard normal distribution such that $E(u_i) = E(v_i) = 0$ and $\mathrm{Var}(u_i) = \mathrm{Var}(v_i) = 1$. This implies that $\mathrm{Cov}(u_i, v_i) = \mathrm{Corr}(u_i, v_i) = \rho$. Here, the variable Z_{it} denotes the exclusion restriction that is required for the identification of the parameters.[13] In our case, Z_{it} denotes the extent to which community cooperation is practised in the household's immediate neighbourhood. We hypothesize that if the community has relatively higher household cooperation, then that can encourage individual network participation through more information sharing.[14] We infer about the community cohesiveness in IHDS by the response to the question:

[12] We tried to account for other factors like whether the household resides in his own apartment as a correlate of indoor piped water. However, majority of the households in IHDS I and II resides in their own house and have been staying in their locality for more than 50 years and so we are unlikely to get any variations in this regard.

[13] The requirement for exclusion restriction for parameter identification is a well-debated issue in the literature. Early works like Maddala (1983) argued in favour of exclusion restriction while Wilde (2000) rules out the need for instruments if both the equations contain varying exogenous regressors. However, recent works like Mourifie and Menago (2014) and Han and Vytlacil (2017) provide evidence that exclusion restrictions via valid instruments might be necessary for parameter identification in case of binary endogenous regressors. For a detailed survey about the identification issues, see Li et al. (2019).

[14] In fact, existing evidences show that social network is associated with better information flow among households (Bandiera and Rasul 2006; Oster and Thornton 2012). Studies have also demonstrated that social networks can be an important source of social spill over via dispersing information

> In some villages or neighbourhood, when there is a community problem such as water supply problem, people bond together to solve the problem. In other communities, people take care of their own families individually. What is your community like? (Desai et al. 2005, 2012).

The response to this question is either bond together to solve the problem or act individually. We hypothesize that households from communities that bond together are more likely to find connections in the government rather than those that act individually. We code community bonding as 1 and individual action as 0.[15] Next, we obtain the mean of this community cooperation at the district level and define it as the variable for community cooperation. This is used as an instrument for the government network equation.

We have chosen our instrument drawing support from the findings that social network reduces search costs of jobs in the labour market by providing job referrals (Beaman and Magruder 2012; Galeotti and Merlino 2009). From the point of view of the public authority those with referrals also offer an advantage of screening and so they could put their faith on them (Montogomery 1991). Inter-community ties might also assist people to acquire skills and resources to participate in networks across their community (Granovetter 2005). In common property resource literature collective action has frequently seen to have emerged as a social norm and influence individual behaviour (Dasgupta 2005). In case of India, studies have found that community water management through VWSC works well in villages that are rich in social capital (Stalker et al. 2001; Prokopy 2005). Thus, the test for exogeneity of the community collective action variable would require testing the null hypothesis that ρ is not significantly different from zero. The sign of ρ in the RBP model, however, does not necessarily reflect the correlation between the endogenous variables. As Filippini et al. (2018) shows that in the RBP setting, it is perfectly possible that ρ and the coefficient of the endogenous variable in the reduced form equation have opposite signs and that ρ is unlikely to reflect the correlation of the binary variables under analysis.[16]

One important criterion for the exclusion restriction to be valid requires that it must not be correlated with piped water access. We have already argued in earlier sections that there are reasons to believe that variations in the access of piped water are not at the village/community level but at the level of the household. Even if schemes are made available at the community level, households are most likely to self-select themselves for piped water connection as this requires payment of

---

and can encourage use of cleaner technologies like liquefied petroleum gas (LPG) in cooking (Srinivasan and Carattini 2016) and solar power (Bollinger and Gillingham 2012).

[15]We recognize that the community cooperation is defined at the level of the villages but IHDS data provides geographic information that is identifiable only at the level of the districts. Hence, we aggregate this data at the district level.

[16]Filippino et al. (2018) model the RBP as a special case of bivariate probit (BP) with correlation parameter ζ. They show that if the data generating process is RBP with correlation parameter ρ and is erroneously modelled as BP, then the relation between the two correlation parameters would be given by $\zeta = \frac{\delta \mathrm{cov}(.)+\rho}{D}$ where D includes covariance and variance terms for the error terms and the endogenous variables and is positive. Also, δ indicates the coefficient of the endogenous variable. The existence of recursive structure would be still justifiable even if ζ and δ have opposite signs.

connection fee as well as maintenance expenditure.[17] For instance, Prasad et al. (2014) find that households in Raigad district, Maharashtra, often refuse to pay for the piped water scheme for dissatisfaction with the scheme, and often because women failed to persuade their male counterparts to incur the expenditure.

3.3 Variable Construction

Social Network Variable: In choosing the social network variable as mentioned earlier, we primarily concentrated on government network given the nature of drinking water supply as a public good in India. We expect that people that have access to government network are more likely to have access to piped water as well. The IHDS survey for both rounds asks the households "Among your acquaintances and relatives, are there any who are in government service", the reported answered is coded as '1' if household reported 'yes' and 'zero' for the remaining cases.

Media Exposure: The IHDS data set also has information regarding the radio, television and newspaper exposure of the household members. Thus, for each type of media, it assigns a score of 0 if the *i*th household member never uses it, 1 if it is rarely used and 2 for regular use. To account for the cumulative effect of the three types of media on the information set of the household, we propose the following index of media exposure: for each type of media, we add up the score of the male member, female member and the child in the household. The maximum score for these three members would be 6 (six) if all three of them regularly use the media and the minimum score would be 0 (zero). Thus, for each media, the index of exposure would be (Actual Score − Minimum Score)/(Maximum Score − Minimum Score). Finally, for all the three media, we attach equal weightage, i.e. one-third to each of this index and add them to arrive at the aggregate index of exposure (MEDIA) such that

$$\sum_{j=3} \frac{1}{3} \frac{\sum m_i}{6}$$

Wealth/Living Standard Indicators: To account for the ability to pay, we considered several measures of income and asset: we measure household's economic status by using per household consumption expenditure of the households (COPC) for the last 30 days at the date of interview for each of the IHDS waves.[18] We use logarithmic transformation of consumption expenditure data for both the

[17] For instance, Gadchiroli Panchayat in Maharashtra collects a nominal user fee per day for drinking water (https://www.thehindu.com/news/national/user-charge-on-water-in-rural-areas-favoured/article3456712.ece).

[18] There exists a vast difference in value of consumption expenditure between the two IHDS waves. This is because round I reported the consumption per households on a monthly basis, whereas round II reports annual per capita consumption expenditure. We take the mean value for the IHDS wave-I and multiply by 12 then divided it by the average IHDS-II **deflator**, 0.545.

rounds.[19] In the difference equation, we obtain the change in log consumption as $\left(e^{\ln(\mathrm{COPC}_{2012})} - e^{\ln(\mathrm{COPC}_{2005})}\right)^{1/7}$ following Coffey et al. (2017). This gives the average, annualized household-level changes in consumption. Our intention is to investigate whether households that were richer in the baseline and have improved their economic status have better access to piped water.

In addition, we include other controls for roof, wall and floor as an indicator for wealth. Following Panda (2014), we categorized the roof as pukka if it has a cement or concrete roof, asbestos, iron brick or stone, and "kutcha" if it is made of tile, slate, plastic, grass, wood and mud and other materials. We similarly coded the wall made of grass, mud, plastic and wood as zero and brick, metal, stone, cement or concrete as one. For floor made up of mud, wood and bamboo are considered as zero and brick, stone, cement and tiles are considered as unity. Thus, all three variables are binary coded.

Caste/Religion and State: The IHDS data provides two types of information about the caste and religion of the households. The caste group (ID13) considers five broad caste categories—Brahmin, Other Backward Classes, Scheduled Caste and Scheduled Tribe and Others. We also control for the religion of the households, and eight categories of them are recorded in IHDS. We observe that more than 80% among them belong to Hindu religion. Consequently, we consider Hindu and Muslim as separate categories and club rest of the religion into a single variable. Finally, to account for the state-specific heterogeneity, we estimate all the models with state-fixed effects.

3.4 Descriptive Statistics

We first analyze the state-wise variation in in-house piped water connection and the association of government network with piped water access. When we consider the proportion of households that switched from no-tap water connection in the first round to tap water in the second round, we find that switching increases for those with higher access to government network. The mean proportion of households shifting to new connections with and without government contact is 0.21 and 0.11 and the difference is statistically significant at less than 1%. In addition, the effect of having access to government network is relatively higher for households in rural areas compared to their urban counterparts (Fig. 3).

We report the comparison of the descriptive statistics of the variables included in our estimation sample in Table 1 across the two rounds of IHDS.

The descriptive statistics reveal that in round I 17% of the sampled households had access to tap water and it is 20% in round II. Thus, a substantial portion of the sample remains out of the public water supply networks. However, 32% of the households

[19]Exploring the histogram of monthly per capita consumption expenditure for both the rounds of IHDS, we find that log transformation alters the scale of the variable such that it approximates normal distribution more closely.

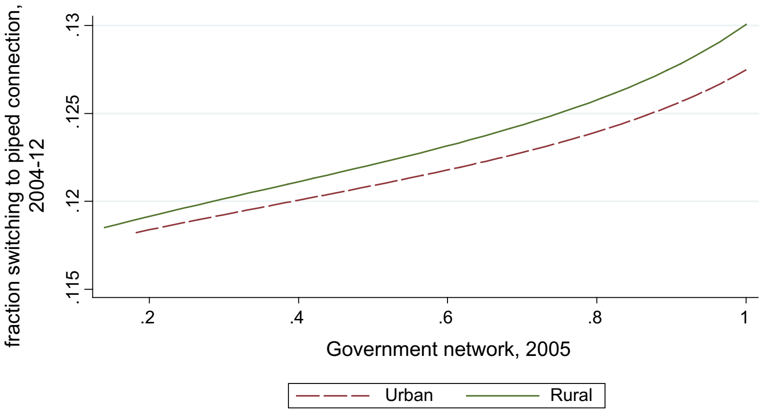


Fig. 3 Rural–urban switch to piped water and government network

in round I have access to government network while the figure is 22% in round II. In round II, mean education levels were higher for both males and females relative to round I. However, more households in round II have improved roof and dwelling characteristics though per-capita consumption expenditure remains almost identical on average. In terms of caste and religion, majority of the rural households in the estimation sample belongs to OBC followed by SC while 84% of the households in the population are from Hindu religion.

4 Results

4.1 Estimation Results

We now present the estimation results of Eqs. (1)–(3). All the estimation presented includes the state-fixed effects. We do not report the magnitude of the fixed effects for brevity of the exposition.

The marginal effects of the PROBIT model for both the rounds are reported in Tables 2 and 3. Here, Model I shows the bivariate relationship between government network and tap water connection without controlling for any covariates. In Model II, we include groundwater levels and see whether accounting for district-level hydrological characteristics affects the influence of government contact. Finally, we report the estimates for our full specification in Model III including household-level wealth, education and caste. We note that social connectivity and groundwater depth both remain positively significant across both the rounds. We also find that the impact of government network on piped water is dampened after we account for household wealth and demographics though the influence of groundwater levels on piped water use remains stable. In Models I and II, informal contact with government increases

Table 1 Variable description

Variable	IHDS wave-I	IHDS wave-II
	Mean (S.D.)	Mean (S.D.)
Dependent variable		
In-house pipe connection (% households)	0.17 (0.37)	0.20 (0.40)
Explanatory and control variables		
Government network	0.32 (0.46)	0.22 (0.41)
Level of groundwater		
Log of groundwater level (in m)	1.66 (0.71)	1.53 (0.72)
Awareness indicator		
Media index	0.25 (0.23)	0.30 (0.22)
Highest education for Female (in years)	3.72 (4.49)	4.34 (4.75)
Highest education for Male (in years)	6.67 (4.84)	6.83 (4.89)
Wealth indicator		
Log of consumption per capita per family (in Rupees)	6.37 (0.66)	6.67 (0.64)
Wall conditions of the house—Dummy “Pukka”	0.54 (0.48)	0.63 (0.48)
Roof conditions of the house—Dummy “Pukka”	0.32 (0.46)	0.45 (0.49)
Floor conditions of the house—Dummy “Pukka”	0.37 (0.48)	0.44 (0.49)
Caste		
Brahmin/Forward	0.04 (0.20)	
OBC	0.41 (0.49)	
SC	0.23 (0.42)	
ST	0.11 (0.31)	
Others	0.21 (0.41)	
Religion		
Hindu	0.84 (0.37)	
Muslim	0.09 (0.28)	
Others	0.07 (0.25)	

Note Figures in Parentheses indicates standard deviations
Source IHDS I and II

Table 2 Marginal effects of the PROBIT model for IHDS I (2004–05)

Variables	Model I	Model II	Model III
Government network	0.0570***	0.0544***	0.00922*
	(0.00478)	(0.00486)	(0.00523)
Level of groundwater			
Log of groundwater level (in m)		0.0362***	0.0327***
		(0.00368)	(0.00439)
Awareness			
Media index			0.0599***
			(0.0117)
Highest education for Female (in years)			0.00267***
			(0.000614)
Highest education for Male (in years)			0.00391***
			(0.000591)
Wealth			
Log of consumption per capita per family (in Rupees)			0.0165***
			(0.00408)
Wall conditions of the house—Dummy "Pukka"			0.0501***
			(0.00570)
Roof conditions of the house—Dummy "Pukka"			0.0163***
			(0.00581)
Floor conditions of the house—Dummy "Pukka"			0.0283***
			(0.00582)
Caste			
OBC			−0.0286***
			(0.0105)
SC			−0.0562***
			(0.00906)
ST			−0.0694***
			(0.00834)
Others			−0.0359***
			(0.00958)
Religion			
Muslim			−0.0312***
			(0.00858)
Others			−0.0107
			(0.0105)

(continued)

Table 2 (continued)

Variables	Model I	Model II	Model III
State FE	Yes	Yes	Yes
Wald chi^2	3227.87***	3050.68***	2846.62****
Observations	27784	25903	19549

Note *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Figures in parentheses indicate robust standard errors

the possibility of having tap connection by approximately more than 5% for both the rounds. However, after accounting for household wealth and education, the influence of social network on piped water, though statistically significant, is reduced in both the rounds. The partial effect of social network in round I is reduced to 1%, and it becomes 1.32% in round II. This indicates that economic status of the household might significantly influence the possession of public water supply and the asset poor households could be left out of the ambit of the scheme. Lastly, the wealth and the education variables for both male and female have the usual positive signs. This, however, indicates having in-premise tap water might be associated with home improvements.

The PROBIT results presented above for each year, however, are unable to identify the causal effects of the variables on tap water access. For instance, we cannot be sure whether increases in any independent variable say pukka house access allows the households to help to build a good infrastructure or is there an independent effect of tap water access or housing conditions. This concern applies equally for all the other variables of interest and more so as there is no information whether possession of piped water precedes the reported independent variables like dwelling quality and government connection. The difference regression in Eq. 2, reported in Table 4, attempts to describe what factors are associated with the up-take of new tap water connection. In addition to taking the differences, we also include the baselines values of government network, aquifer depth and media exposure index as here the initial values might leverage the chances of incremental gain in gaining access to tap water. For male and female education we include only the baseline figures.

The result in Table 4 shows that government network has a strong impact on the access to potable water in rural India, and the impact remains more or less stable across multiple model specifications. The significance of the network variable in the difference equation could imply that within the seven-year period between the surveys those with higher endowments of government network have actively invested in this social capital given the overall groundwater availability. Notice further that the households having social network in the base period, i.e. round I are 3% more likely to avail tap connection in the next round. Clearly, the influence of government contact is higher in the difference model compared to the cross-sectional PROBIT estimates in Tables 2 and 3. In addition, media exposure plays an important role in initiating new piped water connection. We also find that groundwater scarcity induces switching to piped water and that its effect remains statistically significant. Moreover, improvement in housing conditions is also related to tap water connection. Though

Table 3 Marginal effects of the PROBIT model for IHDS II (2011–12)

Variables	Model I	Model II	Model III
Government network	0.0595***	0.0507***	0.0132**
	(0.00603)	(0.00606)	(0.00574)
Level of groundwater			
Log of groundwater level (in m)		0.0434***	0.0376***
		(0.00424)	(0.00439)
Awareness			
Media index			0.109***
			(0.0129)
Highest education for female (in years)			0.00339***
			(0.000567)
Highest education for male (in years)			0.00164***
			(0.000566)
Wealth			
Log of consumption per capita per family (in Rupees)			0.0140***
			(0.00411)
Wall conditions of the house—Dummy "Pukka"			0.0381***
			(0.00575)
Roof conditions of the house—Dummy "Pukka"			0.0120**
			(0.00554)
Floor conditions of the house—Dummy "Pukka"			0.0261***
			(0.00546)
Caste			
OBC			−0.000806
			(0.0126)
SC			−0.00851
			(0.0127)
ST			−0.0514***
			(0.0112)
Others			−0.0451***
			(0.0110)
Religion			
Muslim			−0.0200**
			(0.00900)
Others			0.00226
			(0.0102)
State-fixed effects	Yes	Yes	Yes
Wald chi^2	3955.77	3877.26	4076.58
Observations	27178	25566	22434

Note *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Figures in parentheses indicate robust standard errors

Table 4 Estimates of switching propensity to in-house tap connection, 2004–05 to 2011–12 (Marginal effects)

Variables	Model I	Model II	Model III
Difference of government network	0.0251***	0.0201***	0.0157**
	(0.00410)	(0.00419)	(0.00534)
Government network, 2004–05	0.0638***	0.0508***	0.0313***
	(0.00699)	(0.00688)	(0.00800)
Level of groundwater			
Differences of log of groundwater level (in m) over two waves		0.176***	0.201***
		(0.0378)	(0.0480)
Log of groundwater level (in m), 2004–05		0.0297***	0.0365***
		(0.00362)	(0.00459)
Awareness Indicator			
Difference of media index			0.0310***
			(0.00524)
Media index, 2004–05			0.00677
			(0.0135)
Highest education for Female (in years), 2004–05			0.00231***
			(0.000613)
Highest education for Male (in years), 2004–05			0.00115**
			(0.000563)
Wealth Indicator			
Difference of log of consumption per capita per family (in Rupees)			0.00652
			(0.00565)
Log of consumption per capita per family (in Rupees)			0.0213***
			(0.00413)
Difference of wall conditions of the house—Dummy “Pukka”			0.00763***
			(0.00415)
Difference of roof conditions of the house—Dummy “Pukka”			0.00893*
			(0.00429)
Difference of Floor conditions of the house—Dummy “Pukka”			- 0.000141
			(0.00434)
Caste Control			
OBC			- 0.00636
			(0.0127)
SC			- 0.0104
			(0.0125)
ST			−0.0455***

(continued)

Table 4 (continued)

Variables	Model I	Model II	Model III
			(0.00978)
Others			−0.0444***
			(0.00985)
Religious Control			
Muslim			−0.0138
			(0.00871)
Others			−0.0199**
			(0.00787)
State Control	Yes	Yes	Yes
Wald chi^2	2543.51	2459.54	1842.47
Observations	22424	20932	14490

Note *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Robust SE in parentheses

religious identity does not have any significant association with new connection, we find that SC and ST households are less likely to avail new connection relative to households belonging to upper classes.

Finally, we report the BVP estimates in Table 5.

The RBVP results indicate that the proposed relationship between community influences on household access to piped water is unlikely to hold for round I. Here, the correlation coefficient ρ is not statistically significant. In addition, the government network variable, though positive, is not statistically significant in the piped water equation. We argue that this might be the case as the survey period coincides with that phase of water policy when coverage by habitation was the norm rather than individual connection. In such situation, community connection might be less effective in initiating individual connection. Note that even in case of PROBIT estimates, the coefficient of the government contact in round I was less than that of round II. For the second round, ρ is negative and statistically significant indicating that the unobservables associated with the higher likelihood of piped water access are negatively correlated with the unobservable determining the probability of getting connected to the government network. However, as Filippino et al. (2018) show that in RBP models, the correlation between binary variables under analysis is subsumed in the estimated coefficient of the endogenous variable rather than ρ. In our case, for both the rounds, the estimated coefficients for government network is positive (though statistically significant only in the second round) and thus indicates that piped water access and government contacts are positively related after accounting for community-level cooperation.

In our case for both the rounds, the estimated coefficients for government network is positive and thus indicates that piped water access and government contacts are positively related controlling for community-level cooperation. Other variables of interest like groundwater levels and media exposure retain their expected positive

Table 5 Bivariate PROBIT estimates

Variables	2004–05		2011–12	
	Dependent variable			
	Government network	In-house pipe connection	Government network	In-house pipe connection
Government network		0.195		0.533***
		(0.172)		(0.162)
Bond together to solve local problems	0.295***		0.379***	
	(0.0348)		(0.0347)	
Level of groundwater				
Log of groundwater level (in m)	0.145***	0.240***	−0.120***	0.172***
	(0.0321)	(0.0333)	(0.0314)	(0.0315)
Awareness indicator				
Media index	0.749***	0.542***	1.016***	0.435***
	(0.0836)	(0.101)	(0.0875)	(0.113)
Highest education for Female (in years)	0.0183***	0.0230***	0.0188***	0.0150***
	(0.00412)	(0.00484)	(0.00360)	(0.00423)
Highest education for Male (in years)	0.0338***	0.0129***	0.0157***	0.00453
	(0.00394)	(0.00491)	(0.00396)	(0.00503)
Wealth indicator				
Log of consumption per capita per family (in Rupees)	0.330***	0.116***	0.227***	0.0183
	(0.0282)	(0.0355)	(0.0279)	(0.0330)
Wall conditions of the house—Dummy "Pukka"	0.0393	0.218***	0.0740*	0.156***
	(0.0403)	(0.0442)	(0.0390)	(0.0446)
Roof conditions of the house—Dummy "Pukka"	0.0452	0.0646	−0.0831**	0.0343
	(0.0379)	(0.0409)	(0.0384)	(0.0429)
Floor conditions of the house—Dummy "Pukka"	0.0268	0.160***	−0.0131	0.163***
	(0.0374)	(0.0438)	(0.0371)	(0.0412)

(continued)

Table 5 (continued)

Variables	2004–05		2011–12	
	Dependent variable			
	Government network	In-house pipe connection	Government network	In-house pipe connection
Caste				
OBC	−0.338***	−0.239***	−0.151**	−0.201**
	(0.0705)	(0.0833)	(0.0681)	(0.0840)
SC	−0.345***	−0.436***	−0.230***	−0.240***
	(0.0741)	(0.0884)	(0.0716)	(0.0869)
ST	−0.429***	−0.480***	−0.0950	−0.518***
	(0.0888)	(0.112)	(0.0815)	(0.100)
Others	−0.309***	−0.275***	−0.191***	−0.453***
	(0.0743)	(0.0851)	(0.0728)	(0.0893)
Religious				
Muslim	−0.123**	−0.119	0.00273	−0.0480
	(0.0624)	(0.0785)	(0.0595)	(0.0680)
Others	0.0147	0.0912	−0.00140	0.127
	(0.0856)	(0.100)	(0.0758)	(0.0871)
State-fixed effects	Yes	Yes	Yes	Yes
Constant	−3.299***	−3.313***	−3.130***	−1.937***
	(0.244)	(0.289)	(0.232)	(0.275)
Observations	18,913		22,434	
Rho ρ (μ, €) (unconstrained)	−0.07 (0.43)		−0.24 (0.00)	
Wald test $\rho = 0$ chi^2 (p-value)	4788.25 (0.0000)		5407.95 (0.0000)	

Note Robust SE in parentheses, ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

signs in the water equation and are statistically significant after controlling for wealth and education. Here, per-capita consumption expenditure becomes significant in the government network equation indicating the possibility that those with higher economic status are likely to find government contact and are more likely to have tap water access.

4.2 Discussion

Three important results stand out from our estimations. First, when we take all the rural households from both the rounds, the choice of connecting to piped water seems to be influenced by the extent of connection with the government officials, media exposure and the availability of groundwater in both the rounds. This continues to hold even when we consider the new connections and reinforces the importance of government networks and regional hydrological characteristics via groundwater availability for tap water connection. Finally, we also find that households from communities that engage in collective action are more likely to connect to government network but the effect is significant only for round II. We argue that this reflects the fact that decentralization of the water supply network has begun from the initiation of NRDWP in 2009 and perhaps the community effect is captured in the survey responses of the household in the second round of IHDS that was conducted in 2012.

The emphasis of the government in NRDWP towards building up community institutions like Village Water and Sanitation Committee (VWSC) further substantiates our claim. However, it is also a fact that the community-led initiatives have not been developed fully in rural India. In their assessment of Rural Water Supply Scheme in Maharashtra Bassi and Kabir (2016) notes that the role VWSC has only been restricted to operation and maintenance of rural water schemes and it failed to give impetus to the demand driven process. Using observations from 20 villages in Telangana, Andhra pradesh, Ramachandrudu and Snehalata (2010) find that VWSC is largely non-functional in these areas. Studies have argued that VWSC and other community institutions need strengthening for successful decentralization of rural water supply in India (Hutchings et al. 2017). The weak correlation in the RBVP model is perhaps reflective of the weak community structure of the local institutions.

However, the fact that the importance of social network is somewhat reduced when we include households wealth in the PROBIT estimates needs to be investigated further. This could mean that wealthy households might also have better social connections and could thus effectively exclude the asset poor from public water supply network. Mean comparison of social network across poor and non-poor households across both the rounds reveal that households below poverty line have a statistically significant lower level of social contact compared to the non-poor.[20] Finally, we also note that for SC and ST household's access to piped water is significantly lower relative to Brahmin and upper-caste households. The estimation results for Eqs. (1) and (2) depict that for both the rounds households belonging to SC and ST have relatively lower odds of indoor tap water connection. Also, ST households have a significantly lower probability of switching to piped water connection relative all other castes. While SC and STs are known to have poor asset base and thus lower ability to pay we suggest that another source of their exclusion could be their lower endowments of government network. In fact for both the IHDS rounds taking SC/ST households

[20]In IHDS, poverty is indicated on the basis of the monthly per-capita consumption expenditure and the official planning commission poverty line. Here, poor is a dichotomous variable and takes a value of zero if the household is below poverty line and otherwise unity.

as one group, we find that their average government network endowment is significantly lower than other caste and social groups. The *t*-test of government network for the SC/ST versus other caste groups fails to accept the null hypothesis of equality of means at less than 1% level of significance. We also find that within SC and ST households, the government network is significantly lower for ST households. As expected in the RBP models, the sign of the coefficients of SC/ST in the government network displays negative sign and is persistently significant for SC households for both the rounds.

5 Conclusion

Drinking water in India is a public good since it is solely supplied by local government institutions. This study helps to understand the role of the influential factors which play a significant role to bring private water connection. Using two rounds of cross-section estimation and panel analysis, we see that households with government network have positively significant effects on access to pipe connection, i.e. this paper implies that access to piped water as a private connection is influenced by the extent of social connectivity of the households. We find that water scarcity in terms of falling groundwater levels does provide incentive to switch to tap water but social network in government offices might be another important factor in gaining water connection. We also find limited evidence that community cohesiveness might facilitate water access through increasing government contacts. While the exact mechanisms and the required institutional design are outside the scope of our work, the presence of community effects in the period of post-NRDWP implementation suggests that public efforts to strengthen existing institutions like VWSC might facilitate better access of tap water. Another important implication of our analysis is the possible exclusion of asset poor and SC/ST households from public water supply due to lower contacts in government offices. In fact, analysis of census data reveals that in 2011, only one-fourth of the ST households have tap water connection while the figure is two-fifth for SC households Bhagat (2013). Recent efforts to address this deprivation like SC-/ST-specific water projects like *Mukhyamantri Jan Jal Yojna* in Ranchi indicates increasing public attention to such concerns.[21] In addition, public authorities would have to make the community institutions more inclusive by ensuring adequate representation of the poor and socially disadvantaged groups. Finally, we have reasons to suggest that IEC programmes need to be scaled up in order to stimulate demand for improved water. These findings become all the more important in face of the continuous budget cuts of the Ministry of Drinking Water and Sanitation in NRDWP from 87% in 2009–10 to 31% in 2018–19.

There are several important issues that need to be re-examined in the context of our paper. First one needs to assess the influence of membership in formal orga-

[21] https://www.dailypioneer.com/2019/state-editions/drinking-water-projects-to-focus-sc--st-dominated-areas.html.

nizations like trade unions and caste-based organizations on access to tap water connection. Given the fact disadvantaged caste has limited access to tap water the interaction between caste-based network and government network would be an interesting research agenda to explore. Also, it needs to be seen how availability of secondary water sources is likely to influence the choice of public supplied water. In absence of that our estimates are likely to be on the higher side but would still suggest addressing supply-side issues in NRWDP may not fully address the problems of drinking water access in rural India.

Appendix

See Figs. 4, 5 and 6.

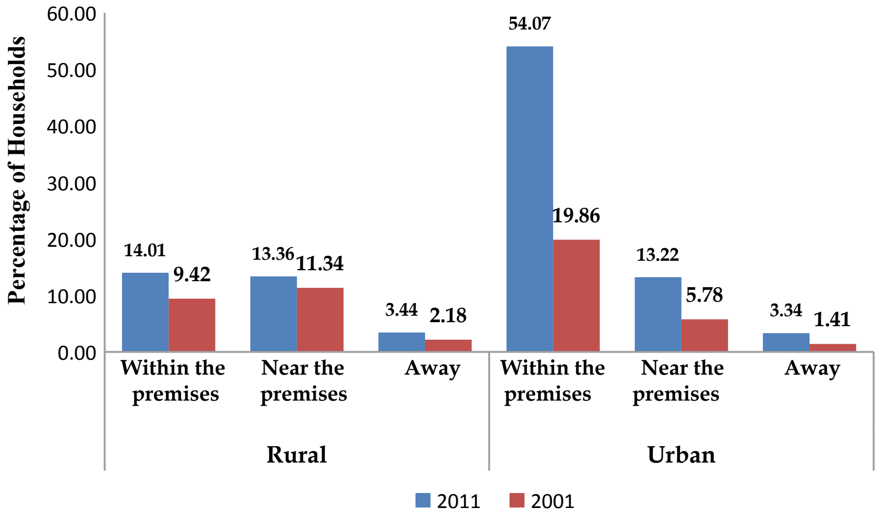


Fig. 4 Distribution of tap water in rural and urban area. *Source* Census (2001) and Census (2011) Household Amenities

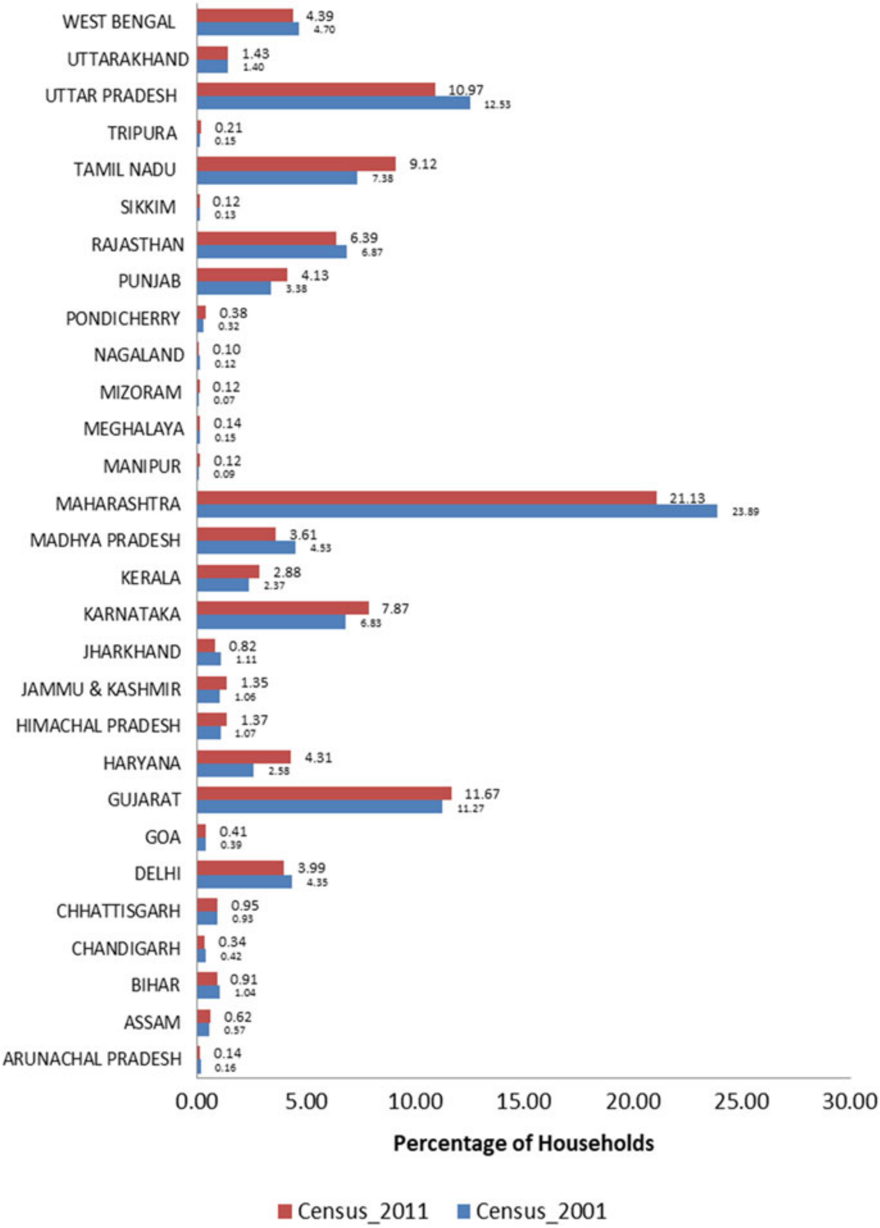


Fig. 5 Statewise variation in in-house tap water connection

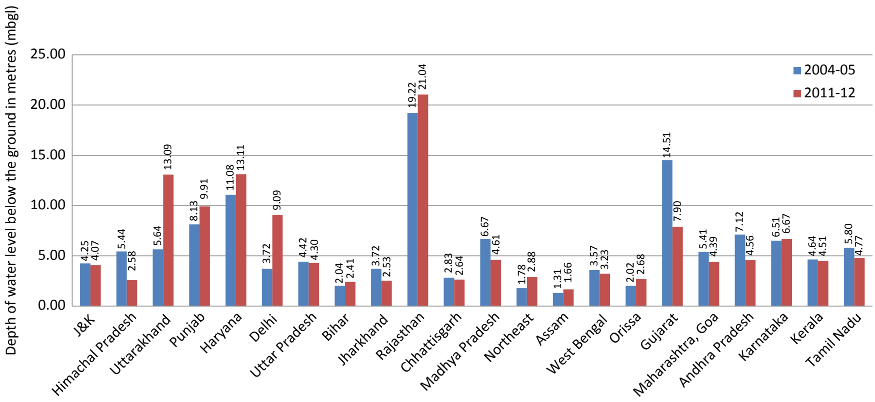


Fig. 6 Minimum groundwater levels in the Indian States across the IHDS rounds. *Source* Water Resource of India Portal, Central Ground Water Commission

References

Anderson, S., Francois, P., & Kotwal, A. (2015). Clientelism in Indian villages. *American Economic Review, 105*(6), 1780–1816.

Balasubramaniam, D., Chatterjee, S., & Mustard, D. B. (2014). Got water? Social divisions and access to public goods in rural India. *Economica, 81*(321), 140–160.

Bandiera, O., & Rasul, I. (2006). Social networks and technology adoption in northern Mozambique. *The Economic Journal, 116*(514), 869–902.

Banerjee, A., & Somanathan, R. (2007). The political economy of public goods: Some evidence from India. *Journal of Development Economics, 82*(2), 287–314.

Bardhan, P. K., & Mookherjee, D. (2000). Capture and governance at local and national levels. *American Economic Review, 90*(2), 135–139.

Bassi, N., & Kabir, Y. (2016). Sustainability versus local management: Comparative performance of rural water supply schemes. In *Rural water systems for multiple uses and livelihood security* (pp. 87–115). Elsevier.

Beaman, L., & Magruder, J. (2012). Who gets the job referral? Evidence from a social networks experiment. *American Economic Review, 102*(7), 3574–3593.

Besley, T., Pande, R., Rahman, L., & Rao, V. (2004). The politics of public good provision: Evidence from Indian local governments. *Journal of the European Economic Association, 2*(2–3), 416–426.

Betancourt, R., & Gleason, S. (2000). The allocation of publicly-provided goods to rural households in India: On some consequences of caste, religion and democracy. *World Development, 28*(12), 2169–2182.

Bhagat, R. B. (2013). Conditions of SC/ST households: A story of unequal improvement. *Economic and Political Weekly*, 62–66.

Blackden, M. C., & Wodon, Q. (2006). *Gender, time use, and poverty in Sub-Saharan Africa*. World Bank Working Paper, 73.

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science, 31*(6), 900–912.

Coffey, D., Spears, D., & Vyas, S. (2017). Switching to sanitation: Understanding latrine adoption in a representative panel of rural Indian households. *Social Science and Medicine, 188,* 41–50.

Comptroller and Auditor General of India. (2018). "Performance Audit of National Rural Drinking Water Programme", Union Government (Civil). Ministry of Drinking Water and Sanitaion. Report No. 15.

Dasgupta, I. (2005). Common property resources: Economic analytics. *Economic and Political Weekly*, *40*(16), 1610–1622.

Desai, S., & Vanneman, R. (2005). National Council of Applied Economic Research, New Delhi. India Human Development Survey (IHDS). *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*, 2010-06-29. http://doi.org/10.3886/ICPSR22626.v8.

Desai, S., & Vanneman R. (2012). National Council of Applied Economic Research, New Delhi. India Human Development Survey-II (IHDS-II), 2011–12. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*, 2015-07-31. http://doi.org/10.3886/ICPSR36151.v2.

Desai, S. B., Dubey, A., Joshi, B. L., Sen, M., Shariff, A., & Vanneman, R. (2010). *Human development in India*. New York: Oxford University. Downloaded from: https://ihds.umd.edu/sites/ihds.umd.edu/files/publications/pdf-files/11HDinIndia.pdf. Accessed May 27, 2015.

Dutta, P., Murgai, R., Ravallion, M., & van de Walle D. (2014). *Right to Work? Assessing India's Employment Guarantee Scheme in Bihar*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-0130-3. License: Creative Commons Attribution CC BY 3.0

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*(1), 39–57.

Filippini, M., Greene, W. H., Kumar, N., & Martinez-Cruz, A. L. (2018). A note on the different interpretation of the correlation parameters in the bivariate probit and the recursive bivariate probit. *Economics Letters, 167,* 104–107.

Galeotti, A., & Merlino, L. P. (2009). *Endogenous job contact networks*, Memio.

Gitto, L., Santoro, D., & Sobbrio, G. (2006). Choice of dialysis treatment and type of medical unit (private vs public): Application of a recursive bivariate probit. *Health Economics, 15*(11), 1251–1256.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives, 19*(1), 33–50.

Han, S., & Vytlacil, E. J. (2017). Identification in a generalization of bivariate probit models with dummy endogenous regressors. *Journal of Econometrics, 199*(1), 63–73.

Hutchings, P., Franceys, R., Mekala, S., Smits, S., & James, A. J. (2017). Revisiting the history, concepts and typologies of community management for rural drinking water supply in India. *International Journal of Water Resources Development, 33*(1), 152–169.

Ilahi, N. (2000). The intra-household allocation of time and tasks: What have we learnt from the empirical literature? Policy research report on gender and development working paper series, 13.

Jessoe, K. (2013). Improved source, improved quality? Demand for drinking water quality in rural India. *Journal of Environmental Economics and Management, 66*(3), 460–475.

Lakshminarayanan, S., & Jayalakshmy, R. (2015). Diarrheal diseases among children in India: Current scenario and future perspectives. *Journal of Natural Science, Biology, and Medicine, 6*(1), 24.

Larson, B., Minten, B., & Razafindralambo, R. (2006). Unravelling the linkages between the millennium development goals for poverty, education, access to water and household water use in developing countries: Evidence from Madagascar. *The Journal of Development Studies, 42*(1), 22–40.

Li, C., Poskitt, D. S., & Zhao, X. (2019). The bivariate probit model, maximum likelihood estimation, pseudo true parameters and partial identification. *Journal of Econometrics, 209*(1), 94–113.

Madanat, S., & Humplick, F. (1993). A model of household choice of water supply systems in developing countries. *Water Resources Research, 29*(5), 1353–1358.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

MDWS 2013, NRDWP Guidelines. (2013). Ministry of drinking water and sanitation, Government of India. http://www.mdws.gov.in/sites/default/files/NRDWP_Guidelines_2013.pdf. Accessed October 2015.

Ministry of Jal Shakti. *India water crisis*. Downloaded from: https://www.moneycontrol.com/news/business/budget-2019-all-eyes-on-jal-shakti-ministry-as-indias-water-crisis-deepens-4170371.html.

MODWS. Ministry Of Jal Shakti. Department Of Drinking Water and Sanitation. Downloaded from: https://jalshakti-ddws.gov.in/.

Montogomery, J. (1991). Social networks and labor market outcomes. *American Economic Review, 8*(5), 1408–1418.

Morris, S. (2007). The impact of obesity on employment. *Labour Economics, 14*(3), 413–433.

Mourifié, I., & Méango, R. (2014). A note on the identification in two equations probit model with dummy endogenous regressor. *Economic Letters, 125*(3), 360–363.

Mu, X., Whittington, D., & Briscoe, J. (1990). Modeling village water demand behavior: A discrete choice approach. *Water Resources Research, 26*(4), 521–529.

National Rural Drinking Water Programme failed to achieve target: CAG. The *Economic Times.* Downloaded from: https://economictimes.indiatimes.com/news/politics-and-nation/national-rural-drinking-water-programme-failed-to-achieve-target-cag/articleshow/65308525.cms. Accessed June 7, 2019.

Nauges, C., & Strand, J. (2007). Estimation of non-tap water demand in Central American cities. *Resource and Energy Economics, 29*(3), 165–182.

Nauges, C., & Van den Berg, C. (2009). Demand for piped and non-piped water supply services: Evidence from Southwest Sri Lanka. *Environmental & Resource Economics, 42*(4), 535–549.

Nauges, C., & Whittington, D. (2009). Estimation of water demand in developing countries: An overview. *The World Bank Research Observer, 25*(2), 263–294.

Oster, E., & Thornton, R. (2012). Determinants of technology adoption: Peer effects in menstrual cup take-up. *Journal of the European Economic Association, 10*(6), 1263–1293.

Panda, R. N. (2014). Tennessee Williams' Cat on a hot tin roof and summer and smoke: A study in sexual difference. *IUP Journal of English Studies, 9*(2).

Panda, S. (2015). Political connections and elite capture in a poverty alleviation programme in India. *The Journal of Development Studies, 51*(1), 50–65.

Persson, T. H. (2002). Household choice of drinking–water source in the Philippines. *Asian Economic Journal, 16*(4), 303–316.

Prasad, P., Mishra, V., & Sohoni, M. (2014). Reforming rural drinking water schemes. *Economic & Political Weekly, 49*(19), 59.

Prokopy, L. S. (2005). The relationship between participation and project outcomes: Evidence from rural water supply projects in India. *World Development, 33*(11), 1801–1819.

Ramachandrudu, M. V., & Snehalatha, M. (2010). Can WASH Services be improved by TAPping? Insights from WASHCost (India) Project 1.

Sarkhel, P. (2013). Employment guarantee and natural vulnerability: A study of MGNREGA in Indian Sundarbans. *Development and sustainability* (pp. 429–447). India: Springer.

Srinivasan, S., & Carattini, S. (2016). *Adding fuel to fire? Social spillovers and spatial disparities in the adoption of LPG in India*. Centre for International Environmental Studies.

Stalker, L., Abyankar, G. V., & Iyer, P. (2001). *Why some village water and sanitation committees are better than others: A study of Karnataka and Uttar Pradesh (India) (English)*. Water and Sanitation Program field note. South Asia. Washington, D.C.: World Bank Group.

Tewari, S., & Bapat, R. (2019). *One in four rural Indian households walks more than 30 minutes to get water*. IndiaSpend.com. Downloaded from: https://scroll.in/article/810363/the-great-indian-water-walkathon. Accessed June 7, 2019.

User charge on water in rural areas favoured. *The Hindu*. Downloaded from: https://www.thehindu.com/news/national/user-charge-on-water-in-rural-areas-favoured/article3456712.ece. Accessed June 7, 2018.

Vanaja, S. (2018). *Access to piped water, time savings and absenteeism in school: Evidence from India* (No. 2058-2018-5267). Downloaded from: https://ageconsearch.umn.edu/record/275954/files/2393.pdf. Accessed April 17, 2019.

Wilde, P. E., & Ranney, C. K. (2000). The monthly food stamp cycle: Shopping frequency and food intake decisions in an endogenous switching regression framework. *American Journal of Agricultural Economics, 82*(1), 200–213.

World Bank India. (2008, June). *Review of effectiveness of rural water supply schemes in India*. Washington, DC: The World Bank.

Public Finance and International Economics

Political Economy of Fiscal Reform in Indian States

Hiranya Mukhopadhyay

Abstract The political economy considerations for fiscal reforms in a big federal country with a multi-party political structure involve a complex process, especially at the subnational level. India, for example, is an ideal example where fiscal reforms in the states present a unique opportunity to study this complex process. There are some important political factors, with ambiguous empirical results, that will have profound effects on fiscal reforms. Fiscal populism is one such factor. It may be noted that the government is not a 'benevolent social planner' that maximizes the utility of the representative individual but is concerned only with winning the next election. As the ruling political party expects vote swing in its favor, it may not indulge into fiscal profligacy like providing free electricity, subsidizing power, or granting tax concessions, etc. which enables the state government to raise its own revenue as a percentage of its total expenditure. This may not necessarily be true, as, a party can become less reform oriented if there is no political uncertainty or pressure. Political alliance is another important variable. When both the federal government and the state are ruled by the same party, it is quite likely that the particular state will enjoy some fiscal benefits. As a result, own revenue as a percentage of total expenditure of the favored states might increase, and the states may be reluctant to initiate hard reforms. On the other hand, a reform-oriented federal government might influence the states with the same ruling political party to initiate bold reforms. An additional political economy impediment, namely the common-pool problem where the major interest group that benefits from the status quo (not initiating hard reforms), is well represented within the government. Practically, no serious attempt has been made so

This paper was prepared for the Adjunct Fellow Program of the Asian Development Bank Institute (ADBI). I thank seminar participants, especially Rana Hasan, Bruno Carrasco, and Peter Morgan, at the Asian Development Bank and ADBI for valuable suggestions. I would also like to thank Dr. Sudipto Mundle and an anonymous referee for important suggestions. The usual disclaimer applies. The views expressed in this paper are those of the author and do not necessarily reflect the views and policies of the Asian Development Bank or its Board of Governors or the governments they represent.

H. Mukhopadhyay (✉)
Central and West Asia Regional Department, Asian Development Bank,
Mandaluyong, Philippines
e-mail: hmukhopadhyay@adb.org

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_11

far to understand the political economy of reform process at the state level in India in a comprehensive manner. The purpose of this paper is to settle the ambiguities in the relationship between fiscal reform and the political economy determinants empirically in the context of the India states covering fourteen major states in India during the period 2001–02 and 2013–14.

> ….in place of the Platonic assumption that policymakers selflessly dedicate themselves to disinterested pursuit of the public good, public choice theory posits that they behave as standard economic theory assumes Homo economicus to behave: they maximize their personal economic well-being. Bureaucrats fight for turf, politicians decide whether or not to hand out pork by calculating whether it will enhance their prospects of re-election, pressure groups seek rents (e.g. from the imposition of import restrictions) by offering political funding or support, and politicians give into these pressures until they fear that the voters will perceive that they are being taken for a ride, ….
>
> —Williamson (1994)

1 Introduction

Fiscal reforms imply changing existing public institutions and reforming the delivery of public goods and services through changes in policies, legislation, and regulations. Political economy considerations for fiscal reforms in a large federal country with a multi-party political structure involve a complex process, especially at the subnational level. Fiscal reforms in the states (subnational governments) of India present a unique opportunity to study this complex process. This paper therefore focuses on the political economy of fiscal reforms in the states in India.

A principal conclusion of the empirical literature on the political business cycle is that models based on the manipulation of fiscal policy to win elections adapt well to the data and form a strong basis for a convincing theoretical model of electoral effects on economic outcomes (Drazen 2000). Khemanii (2000), focusing on national elections, did not find any evidence of populist measures to influence voters. Nonetheless, there are some important political factors, with ambiguous empirical results, that will have profound effects on fiscal reforms. Fiscal populism is one such factor. It may be noted that the government is not a 'benevolent social planner' that maximizes the utility of the representative individual but is concerned only with winning the next election. One important trade-off at the subnational level is the allocation between current transfers and capital expenditure within the discretionary class of expenditures. Capital expenditure may either be treated as a simple residual within the discretionary expenditure class after politically protected current transfers and loans to loss-making parastatals are taken care of. Alternatively, it may be the rational outcome of a utility maximization process by a government whose sole objective is to maximize its chances of being reelected. Rajaraman and Mukhopadhyay (2001) present a model following opportunistic models of public choice by linking timing and shares of components within the discretionary expenditure class to the electoral cycle, despite the decisive empirical rejection of the *opportunistic political business cycle* approach in the context of members of the Organization for Economic Co-operation and Development (Alesina and Roubini 1992). Alesina and Roubini (1992) primarily concentrated on the monetary opportunistic model of the political business cycle. As already stated, however, there is evidence of pre-electoral manipulations of fiscal policies.

The direction of the causal links between policy manipulations and winning elections are, however, not very clear. If the ruling political party expects vote swing in its favor, it may not indulge in fiscal profligacy such as providing free electricity, subsidizing power, or granting tax concessions, thus enabling the state government to maintain low deficits.[1] This may not necessarily be true, a party can become less reform oriented if there is no political uncertainty or pressure from the opposition political parties.

Needless to say, the proxy variable for fiscal populism will interact with important considerations, such as (i) distance from the next election (Alesina et al. 2006), (ii) whether the ruling political party is in their first or second term, (iii) the nature of the majority in the state assembly (strong government measured by simple, absolute or 2/3rd majority), and (iv) the voting history (whether power alternates between the two major political parties of a state with each election cycle).

Political alliance is another important variable. This refers to *discriminatory federal government policies*, such as higher investment in a particular federal public sector enterprise which has the potential of transforming a local economy in a state and therefore, can influence both the revenues and expenditure needs of a state favorably (Rao and Singh 2000).[2] When the federal government and the state are both ruled by the same party, the particular state is likely to enjoy some benefits.[3] As a result, own revenue as a percentage of the total expenditure of the favored states might increase, and the states might be reluctant to initiate hard reforms.[4] On the other hand, a reform-oriented federal government might influence the states with the same ruling political party to initiate bold reforms. Therefore, the relationship can be ambiguous.

---

[1] This decision not to engage in fiscal profligacy could also be justified under the argument that in the event of a high probability of reappointment, the government would be left with higher debt stock (Alesina and Tabellini 1987).

[2] The subnational governments depend on the federal government for fiscal transfers, and that may be a quite substantial proportion of total revenue of some subnational governments.

[3] Mukhopadhyay and Das (2003) found empirical support for discretionary fiscal policy in a paper on the persistence of horizontal imbalances in India.

[4] Pisauro (2003) notes that there exists a moral hazard problem created by the possibility of a bailout of local governments by the central (federal) government. This will reinforce our argument.

An additional political economy impediment, namely the common-pool problem where the major interest group that benefits from the *status quo* (not initiating hard reforms), is well represented within the government (Koptis 2009).[5]

Fiscal reforms are not easy and depend on a host of political economy factors.[6] Hence, understanding these factors in the design, acceptance, and implementation of reforms implies a sound knowledge of how to change the prevailing situation. Baring a few attempts (Khemani 2000; Rao and Singh 2000, 2007; Mukhopadhyay and Das 2003), no rigorous empirical attempt has been made so far to understand the political economy of the reform process at the state level in India.[7] As we have already stated, there are ambiguities about the relationship between fiscal reform and the political economy determinants, and the purpose of this research is to settle these issues empirically in the context of the India states.

2 Empirical Methodology and Data Requirements

We will run a panel data regression for the 14 major states in India during the period 2001–02 and 2013–14. The states are Andhra Pradesh, Bihar, Gujarat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal. To define a proxy for fiscal reform will be a challenging task. We could consider several options: (i) if there are steady improvements in own-tax revenue to gross state domestic product (GSDP) ratio; and (ii) if there are sustained improvements in revenue deficit to gross state domestic product.

Fiscal populism will be proxied by expected *vote swings*. But whether a party will ultimately stay in power or not is decided by the *swing in seats*. In a two-party system, only a change in popularity can cause the seat margin to change. In a multi-party system, a swing in the seat margin could be attributed to two factors: swing in votes and *split factor* (that is the change in the degree of unity between opposition parties[8] (see Butler et al. 1995).

Therefore, we may write,

---

[5]One implication is that the fiscal rules such as the Fiscal Responsibility and the Budget Management Acts in India might fail because of the common-pool problem. The policymakers often find it politically rewarding not to be fiscally disciplined (Wyplosz 2012).

[6]There is another line of argument that says targeted populist measures for social constituencies can help achieving fiscal reforms (Bijukumar 2004).

[7]Rao and Singh (2000, 2007) and Mukhopadhyay and Das (2003) primarily focus on the influence of political economy factors on India's federal fiscal system and horizontal imbalances across states. These papers do not focus on political economy considerations for endogenous fiscal reforms in the states in India. Khemani (2000) is the only paper that deals with this issue.

[8]The index of opposition unity (IOU) is a generalized measure of vote splitting:

IOU = (Vote of the largest opposition party/Sum of votes of all the opposition parties) * 100.

If there is only one opposition party (i.e., if the opposition is fully united), the IOU is clearly 100. Thus, the IOU can vary between zero and 100; the higher it is, the greater the unity of the opposition.

$$\text{Swing in votes} + \text{Split factor} = \text{Swing in Seats}$$

Therefore, the control variable for the political parties is the swing in votes and not the split factor as they cannot control the opposition unity. Note that, ex *ante*, political parties can only predict and control vote swings.

Discriminatory federal government policies will be a dummy variable in regression. It is one for the states where there was the same party in the federal government as well as in the state and zero otherwise.

Given these ambiguities, we explore the relationship directly between expected swing in votes and its impact on fiscal reforms (or fiscal populism) in Indian states, adjusted for other political factors. In many instances, despite reform motivations, states cannot reduce revenue deficit due to the high incidence of committed expenditure (salary, pension, and interest payment). Therefore, we have also controlled for this variable to avoid omitted variable bias.[9] We use elections and fiscal data from 14 major states over 2009–10 to 2015–16. Binary choice regression models (linear probability models, PROBIT, and LOGIT) have been used for empirical analysis. The data on state elections have been obtained from the Election Commission of India Web site (http://eci.nic.in/eci/eci.html). The Reserve Bank of India's database on the state budgets will be the main source of data for states' finances.[10]

3 Results

For each state, we used data for 3 years before the next election under the assumption that the ruling party's performance assessments for the next election and populist measures, if any, resume 2–3 years before elections.[11,12] Some key facts about the state elections are presented in Table 1.

In this paper, fiscal reform in a state is defined by improvements in revenue deficit[13] (as a percentage of gross state domestic product) over three consecutive years. We have also checked the robustness of the empirical results by assuming own-tax revenue (as a percentage of GSDP) instead of revenue deficit.[14] The dependent

---

[9]State-wise committed expenditure as a percentage of total own revenue for the selected years for each state is obtained from the RBI's annual database on states' finances. The range widely varied between 45.6% in Karnataka to 201% in West Bengal in FY2011. Relative positions of the states also remained unchanged. This is also reported in Rao et al. (2005).

[10]https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=State%20Finances%20:%20A%20Study%20of%20Budgets.

[11]Alesina et al. (2006) also argued that stabilizations are more likely to occur at the beginning of a term of office.

[12]State elections are exogenous in India, not endogenously determined by the policymakers (Ito 1990).

[13]Revenue Deficit = Current Expenditure − Total Revenue

[14]Regression results with own-tax revenue as an indicator of fiscal reform are reported in Appendix 1.

Table 1 Major facts about state elections

States	Election years	Changes in voting share in contested seats ($VS_t - VS_{t-5}$)	Ruling party	Selected years
1. Andhra Pradesh	2009–2014	−24.62	Indian National Congress (INC)	FY2011, FY2012, FY2013
2. Bihar	2010–2015	1.88	Janata Dal (United)	FY2011, FY2012, FY2013
3. Gujarat	2007–2012	−0.82	Bharatiya Janata Party (BJP)	FY2009, FY2010, FY2011
4. Haryana	2009–2014	−14.51	INC	FY2011, FY2012, FY2013
5. Karnataka	2008–2013	−13.86	BJP	FY2010, FY2011, FY2012
6. Kerala	2011–2016	−7.53	INC	FY2011, FY2012, FY2013
7. Madhya Pradesh	2008–2013	7.1	BJP	FY2010, FY2011, FY2012
8. Maharashtra	2009–2014	−18.41	INC	FY2011, FY2012, FY2013
9. Odisha	2009–2014	−0.75	Biju Janata Dal	FY2011, FY2012, FY2013
10. Punjab	2007–2012	−2.65	Shiromani Akali Dal (SAD)	FY2009, FY2010, FY2011
11. Rajasthan	2008–2013	−3.61	INC	FY2010, FY2011, FY2012
12. Tamil Nadu	2011–2016	−13	All India Anna Dravida Munnetra Kazhagam (AIADMK)	FY2011, FY2012, FY2013
13. Uttar Pradesh	2007–2012	−4.48	Bahujan Samaj Party (BSP)	FY2009, FY2010, FY2011
14. West Bengal	2011–2016	−4.98	All India Trinamool Congress (TMC)	FY2011, FY2012, FY2013

Source http://eci.nic.in/eci/eci.html

variable in the binary choice regression model and the explanatory variables are as follows:

(i) Pr ($Y_i = 1$; if fiscal reforms initiated; $= 0$ otherwise); where Y is an indicator variable that denotes the presence of fiscal reforms.
(ii) VSCH = Difference in vote shares (in contested seats only) between elections in periods t and $t - 5$. Note that VSCH is time-invariant for the individual state over three data points. This variable essentially represents a proxy (*best guess*) for the ruling party's expectation about the change in the vote shares.[15]
(iii) VSD = VSCH * (1/time) (where time takes three values 3 years from the next elections, 2 years, and 1 year). In other words, expectations are being refined as the election date is coming closer.
(iv) VSDGOV = VSD * (percentage of total seats in the assembly by the ruling government).
(v) Political alliance: $D_a = 1$ if both the state and the central (federal) governments are ruled by the same political party; 0 otherwise.
(vi) Voting pattern: $D_k = 1$ for Kerala; and 0 otherwise.[16]
(vii) Commitownrev: Total committed expenditure as a percentage of total own revenue for the state-wise selected years.

First set of empirical results is summarized in Table 2.

Before we discuss the empirical results, it is important to define the marginal effects in a binary choice regression model. In a linear probability model (PROBIT), the marginal effect of X_i on $Y_i (Y_i/\partial X_i)$ is simply β_i in a regression model of $Y_i = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$. However, in a PROBIT model of $Y_i = \Phi\ (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$, the marginal effect is $\beta_i f\ (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$, where Φ is the cumulative distribution function (CDF) and f is the probability density function (pdf). In the case of the linear probability model, the partial effect is constant. However, in the case of the LOGIT and PROBIT modes, we have to calculate them at different levels of the explanatory variables to get an idea of the range of variation of the resulting changes in probabilities.[17]

We carried out the endogeneity test and ruled out that VSCH and VSDGOV are endogenous in Model II (see Appendix 2). The marginal effects of VSCH, obtained from the PROBIT model, are in Tables 3 and 4.[18]

---

[15] Survey-based tracking of the mood of the electorate has been a regular feature in Indian elections. Most political parties sponsor opinion polls, openly or secretly, before elections. Moreover, both traditional and the new media facilitate dissemination of electorate views about the participating political parties.

[16] Election results in Kerala always alternate between two major political blocks.

[17] To evaluate the 'average' or 'overall' marginal effect, two approaches are frequently used. One approach is to compute the marginal effect of the sample means of the data. The other approach is to compute marginal effect at each observation and then to calculate the sample average of individual marginal effects to obtain the overall marginal effect. For large sample sizes, both the approaches yield similar results. However, for smaller samples, averaging the individual marginal effects is preferred. We followed the first approach in this paper.

[18] From a predictive analytics perspective, probit and logit perfectly predictively equivalent (Gunduz and Fukue 2013). There is not much difference.

Table 2 Regression results

	Model I			Model II			
	Linear probability model	PROBIT	LOGIT	Linear probability model	PROBIT	LOGIT	PROBIT
Constant	−0.36 (3.37)	−0.33 (−1.28)	−0.54 (−1.24)	0.36 (3.37)	−0.37 (−1.42)	−0.60 (−1.37)	−0.08 (−0.12)
VSCH: Change in vote share	−0.04 (−2.25)	−0.14 (−2.26)	−0.24 (−2.16)	−0.04 (−2.25)	−0.12 (−2.12)	−0.19 (−2.04)	−0.11 (−1.94)
VSD: Changes in vote share adjusted for distance from elections	0.06 (3.91)	0.26 (2.13)	0.43 (2.03)				
VSDGOV: Changes in vote share adjusted for distance from elections and strength of the ruling party in the assembly				0.10 (3.43)	0.37 (2.10)	0.61 (2.01)	0.37 (2.09)
COMMITOWNREV							−0.004 (−0.71)

Source Author's estimates
Notes (1) Figures in parentheses are either *t*-statistics or *z*-statistics for PROBIT and LOGIT
(2) LPM estimates are based on white heteroscedasticity-consistent standard errors and covariance
(3) Political alliance and voting pattern turned out to be statistically insignificant. Not reported in the table. COMMITOWNREV also turned out to be insignificant. See the last row of Table 2
(4) It is desirable to adjust the coefficients of the three models so that they are all on a comparable level

Table 3 Marginal effect of changes in vote share on fiscal reform probability (Model I)

	VSCH
Distance from elections: 3 years	
Average of VSCH	−7.16
Coefficient[a]	−0.05
Marginal effect	−0.03
Distance from elections: 1 year	
Average of VSCH	−7.16
Coefficient	0.12
Marginal effect	0.1

Source Author's estimates
[a]Coefficient in this case (3 years before elections) refers to coefficient of VSCH (−0.14) + coefficient of VSD divided by 3 (+0.26/3). See Table 2 for the original coefficient values

Table 4 Marginal effect of changes in vote share on fiscal reform probability without COMMITREV (Model II)

	VSCH
Distance from elections 3 years and the ruling party has 50% of seats	
Average of VSCH	−7.16
Coefficient	−0.06
Marginal effect	−0.03
Distance from elections 3 years and the ruling party has 80% of seats	
Average	−7.16
Coefficient	−0.02
Marginal effect	−0.01

Source Author's estimates

The results show that the expected lower vote share of the ruling party has a favorable impact on the probability of initiating fiscal reform (the marginal effect is negative), but this effect completely turns around to fiscal profligacy (the marginal effect is positive) closer to elections (1 year from elections in our sample). The probability of initiating fiscal reforms decline as the ruling party expects a decline in vote share.

Next, we consider whether the results change when we adjust the difference in voting shares for both distance from the next elections and the percentage of seats in the state assembly (VSDGOV). The results in Table 4 do not show any qualitative difference in marginal effects when elections are far away (3 years).

However, when we compare Models I and II, an interesting fact emerges in terms of changes in the ruling party's perception of fiscal reforms (Table 5).

The results are summarized in Fig. 1. Fiscal profligacy sets-in much earlies (2 years before elections) if the ruling party expects a lower share in votes and the ruling

Table 5 Comparison of Models I and II

Model I		Model II	
Distance from Elections	**3 Years**	**Distance from Elections**	**3 Years**
Coefficient of VSCH	−0.05	Share of seats (ruling Party)	70%
Distance from Elections	**2 Years**	Coefficient of VSCH	−0.03
Coefficient of VSCH	−0.01	**Distance from Elections**	**2 Years**
Distance from Elections	**1 Year**	Share of seats	70%
Coefficient of VSCH	**0.12 (POSITIVE)**	Coefficient of VSCH	**0.01 (POSITIVE)**

Source Author's estimates

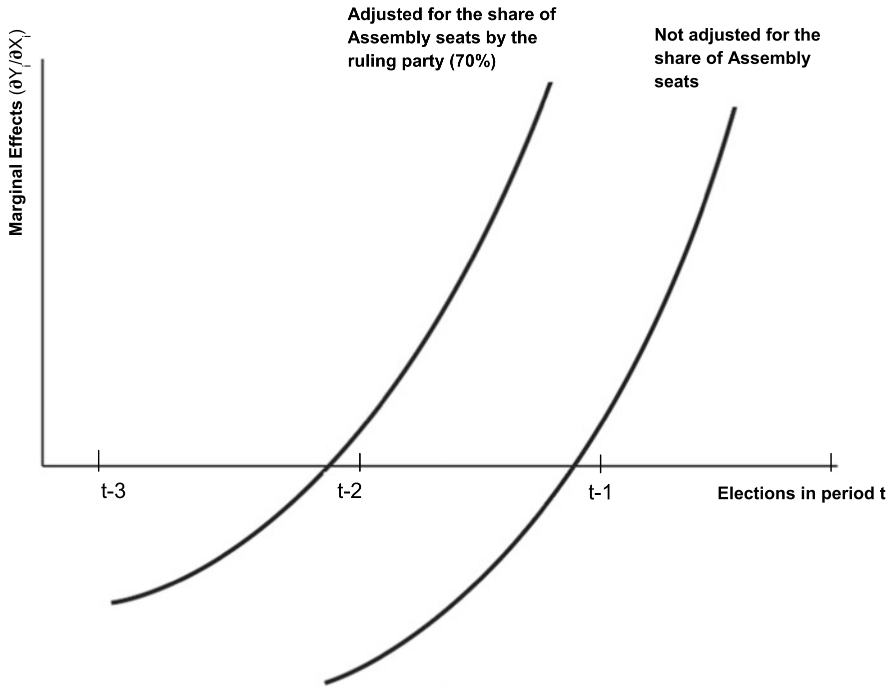


Fig. 1 Changes in the marginal effects of expected swing in the vote share

government has a big share in total assembly seats (70% is the turnaround point).[19] The marginal effect will become positive. This may be a result of the 'common-pool problem.' Drawing on the literature on fragmented government and public spending (Kontopoulos and Perotti 1997; Schaltegger and Feld 2009), we can also argue that each elected member of the ruling party will lobby initiatives to raise expenditure on items favoring his or her constituency in the expectation that this will brighten their chance of winning the next election. This pressure intensifies if the ruling party has a large number of elected members in the state assembly.

4 Conclusion

The main finding of this paper is that political economy variables are indeed important determinants of fiscal reforms and suggest the use of official populism to sway voters (different from Khemani 2000). Moreover, the onset of fiscal populism is faster when the ruling government (including the coalition partners—if any) has a large share of seats in the state assembly. As we have said, this could be the result of the 'common-pool problem.' However, we cannot determine whether this 'common-pool problem' is driven by the coalition governments in some states (Persson et al. 2007). Our empirical results are robust across different indicators of fiscal reform.

The results do not show whether the source of fiscal profligacy is higher current spending or tax benefits. Furthermore, our results could also be biased if a lower revenue deficit is planned to accommodate higher capital spending before elections, given an agreed fiscal deficit target. In other words, fiscal populism can also be manifested through higher capital spending (primarily, the construction of state and village roads in the constituencies selected for political patronage).[20]

Two future areas of research would provide a better understanding of this process. First, the impact of fiscal institutions on fiscal populism in the Indian context should be further investigated. Can an independent and effective fiscal institution break the link between electoral cycles and fiscal populism?

A second area of research concerns why voters respond to pre-elections fiscal populism if they are believed to be rational? The basic argument, first formalized by Rogoff and Sibart (1988) and Rogoff (1990), is that voters have imperfect information about relevant characteristics of potential policymakers, and what appears to be temporary fiscal stimulus have an effect on the election result. The second research question should therefore consider whether this hypothesis can be maintained in the era of new media (social media) activism?

[19]Bihar after 2010 elections is one such example. JD (U) and BJP together had 85% of the total assembly seats. However, the coalition ended on June 16, 2013. Mr. Nitish Kumar, however, continued to be the Chief Minister from JD (U) with the support of some other parties. Rashtriya Janata Dal decided to support JD (U). We could not find the evidence of fiscal reform, by our definition, despite an expected (but small) positive swing in votes in 2015 state assembly elections.

[20]Khemani (2000) found the electoral effect on state roads in big states is significant.

Appendix 1: Regression Results with Own-Tax Revenue as an Indicator for Fiscal Reform

	Model I	Model II
	PROBIT	PROBIT
Constant	−0.41 (−1.28)	−0.39 (−1.48)
VSCH: change in vote share	−0.12 (−1.95)	−0.11 (−1.89)
VSD: Changes in vote share adjusted for distance from elections	0.18 (2.24)	
VSDGOV: Changes in vote share adjusted for distance from elections and strength of the ruling party in the assembly		0.31 (2.17)
Marginal effect for 1 year before elections (compare with the marginal effect in Table 3)	0.047	

Source Author's estimates
Note Figures in parentheses are *t*-statistics

Appendix 2: The Test of Endogeneity

Test of endogeneity follows the procedure suggested in Wooldridge (2002, p. 122) and Bun and Harrison (2014). We use Model II with VSCH and VSDGOV as the explanatory variables. Note that VSDGOV = VSCH * (percentage of total seats in the assembly by the ruling party/Time). Time takes three values 3, 2, and 1. Thus, if VSCH is suspected to be endogenous, so does VSDGOV. Let us denote (percentage of total seats in the assembly by the ruling party/time) by GOVTIME. We use OLS to test the endogeneity of VSCH and VSDGOV. The procedure is as follows:

Step 1: Run two OLS regressions with dependent variables VSCH and VSDGOV. The explanatory variables (instruments) are COMMITOWNREV, GOVTIME, Alliance, and COMMITOWNREV multiplied by GOVTIME.[21] Save two residuals—$V1$ and $V2$.

Step 2. Run the original liner probability model (OLS) and PROBIT in Model II with the binary dependent variable as explained in the paper. The explanatory variables are VSCH, VSDGOV, $V1$, and $V2$. Notice that under the null hypothesis that VSCH and VSDGOV are exogenous, the coefficient of $V1$ = the coefficient of $V2 = 0$.

[21] See Wooldridge (2002) for the justification for including this interaction term in the set of instruments.

The estimated equations are as follows:
OLS:

$$\text{Binary Dependent Variable} = \underset{(2.88)}{0.32} - \underset{(-2.70)}{0.07}\,\text{VSCH} + \underset{(2.61)}{0.18}\,\text{VSDGOV} + \underset{(1.63)}{0.05}\,V1 - \underset{(-1.57)}{0.13}\,V2$$

$R^2 = 0.21$
PROBIT:

$$\text{Binary Dependent Variable} = \underset{(-1.37)}{-0.47} - \underset{(-2.59)}{0.21}\,\text{VSCH} + \underset{(2.43)}{0.64}\,\text{VSDGOV} + \underset{(1.63)}{0.15}\,V1 - \underset{(-1.56);}{0.40}\,V2$$

Following this procedure, we derived the F statistic of 1.40 with 2 and 37 degrees of freedom for OLS and the likelihood ratio of 2.91 with 2 degrees of freedom for PROBIT. Both are not significant at 5% level. Thus, we do not reject exogeneity of VSCH and VSDGOV.

References

Alesina, A., Ardagna, S., & Trebbi, F. (2006). *Who adjusts and when? On the political economy of reforms*. NBER Working Paper Series. No. 12049.

Alesina, A., & Roubini, N. (1992). Political cycles in OECD economics. *Review of Economic Studies, 59,* 663–668.

Alesina, A., & Tabellini, G. (1987). *A positive theory of fiscal deficit and government debt in a democracy*. NBER Working Paper Series, No. 2308.

Bijukumar, V. (2004). Economic reforms, populism and party politics in India. *Indian Journal of Political Science, 65*(2), 161–180.

Bun, M. J., & Harrison, T. D. (2014). *OLS and IV estimation of regressions models including endogenous interaction terms*. Amsterdam School of Economics, Discussion Paper, 2014/02.

Butler, D., Lahiri, A., & Roy, P. (1995). *India decides: Elections 1952–1995*. New Delhi: Books and Things.

Drazen, A. (2000). *The political business cycle after 25 years.* University of Maryland, Hebrew University of Jerusalem, and NBER, Mimeo.

Gunduz, N., & Fukue, E. (2013). *On the predictive analytics of the probit and logit link functions.* RIT Scholar Works. Rochester Institute of Technology, Mimeo (https://scholarworks.rit.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=2237&context=article).

Ito, T. (1990). The timing of elections and political business cycles in Japan. *Journal of Asian Economics, 1*(1), 135–146.

Khemani, S. (2000). *Political cycles in a developing economy: Effects of elections in the Indian States.* Development Research Group, The World Bank, Mimeo.

Kontopoulos, Y., & Perotti, R. (1997). Government fragmentation and fiscal policy outcomes: evidence from OECD countries. In *Fiscal Institutions and Fiscal Performance.* National Bureau of Economic Research (1999), 81–102.

Koptis, G. (2009). The political economy of fiscal reform in Central and Eastern Europe. *OECD Journal of Budgeting, 8*(3), 1–11.

Mukhopadhyay, H., & Das, K. (2003, April 5). Horizontal imbalances in India: Issues and determinants. *Economic and Political Weekly*.

Persson, T., Roland, G., & Tabellini, G. (2007). Electoral rules and Government spending in parliamentary democracy. *Quarterly Journal of Political Science, XX*, 1–34.

Pisauro, G. (2003). *Fiscal decentralization and the budget process: A simple model of common pool and bailouts*. SocietaItaliana di Economia Pubblica Working Paper 294.

Rajaraman, I., Mukhopadhyay, H., & Amar Nath, H. K. (2001). *The anatomy of sub-national fiscal reform.* National Institute of Public Finance and Policy, Mimeo.

Rao, M. G., Anand, M., & Choudhury, M. (2005). *Resource devolution from the centre to states: Enhancing the revenue capacity of states for implementation of essential health interventions*, MPRA Paper No. 24387, Munich University, Germany.

Rao, M. G., & Singh, N. (2000). *The Political economy of Centre-State fiscal Transfers in India*, presented at the Columbia University-World Bank Conference on Institutional Elements of Tax Design and Reform, February 18–19, 2000, at Columbia University.

Rao, M. G., & Singh, N. (2007). The political economy of India's fiscal federal systems and its reforms. *The Journal of Federalism, 37*(1), 26–44.

Rogoff, K. (1990). Equilibrium political budget cycles. *American Economic Review, 80,* 21–36.

Rogoff, K., & Sibert, A. (1988). Elections and macroeconomic policy cycles. *Review of Economic Studies, 55,* 1–16.

Schaltegger, C., & Feld, L. (2009). Do large cabinets favor large governments? evidence on the fiscal commons problem for Swiss cantons. *Journal of Public Economics, 93*(1-2), 35–47.

Williamson, J. (1994). *The political economy of policy reform* (edited), Peterson Institute for International Economics.

Woolbridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: The MIT Press.

Wyplosz, C. (2012). *Fiscal rules: Theoretical issues and historical experience*. NBER Working Paper, No. 17884.

Vertical Fiscal Imbalances and Its Impact on Fiscal Performance: A Case for Indian States

Madhurima Koley and Kumarjit Mandal

Abstract Vertical fiscal imbalance (VFI) is defined as the share of sub-national governments' own spending not financed through own revenues. Theoretical and empirical literatures on VFIs have identified them as an obstacle to sub-national accountability and good fiscal performance. India is a decentralized economy where a marked distinction is made between the spending and revenue responsibilities between the Centre and states. However, spending decentralization has not always complemented revenue devolution, giving rise to huge VFIs in the states. This paper attempts empirically to examine the relation between VFI and fiscal performance for Indian states. It provides stylized facts on size, financing and long-run trend of VFI for 24 major Indian states. Panel data estimation for all those states, covering the period 1995–96 to 2014–15, reveals that on an average the primary deficit of the state governments decline by 15 percentage point of Net State Domestic Product (NSDP) for each percentage point decline in VFI.

Keywords Vertical fiscal imbalance · Fiscal performance

JEL Classifications H62 · H77 · H79

1 Introduction

"Fiscal Decentralization is in vogue"—Oates (1999). Fiscal decentralization generally refers to the devolution of taxing and spending powers from the control of the central government authorities to the government authorities at the sub-national levels (regional, provincial, municipal, etc.). Increased fiscal decentralization is seen as an important means of increasing democratic participation in decision-making process. It is advocated on the grounds that sub-national governments have more

M. Koley (✉)
K.K. Das College, Kolkata 700084, India
e-mail: madhurima.koley@gmail.com

K. Mandal
Department of Economics, University of Calcutta, Kolkata, India
e-mail: kumarjitm@hotmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_12

information and therefore, can better match policies with citizens' preferences (Oates 1972). Individual local governments are presumably more close to the people and geography of their respective jurisdiction, and they possess knowledge of both local preferences and cost conditions that a central agency is unlikely to have and thus can supply appropriate form and level of public services. Again, fiscal decentralization may help in undoing the monopolizing power of the central government in collecting tax revenue; rather it may introduce competition among the sub-national government that would ultimately lead to scale efficiency in public sector (Brennan and Buchanan 1980). Finally, decentralization is likely to increase accountability and transparency in the delivery of public goods and services. Thus, fiscal decentralization is expected to boost public sector efficiency which, in turn, leads to welfare gains.

However, one of the major drawbacks of fiscal decentralization is vertical fiscal imbalance (VFI), which occurs when spending decentralization[1] outpaces revenue decentralization.[2] The sub-national governments rely on central government transfers and borrowing to finance their expenditures. The process of transfers from central government to sub-national governments gives rise to the problem of VFI. A review of theoretical and empirical literatures reveal that VFIs have often been identified as an obstacle to sub-national accountability and good fiscal performance. Recent econometric evidence (Eyraud and Lusiyan 2013) confirms that increases in VFIs are associated with general government fiscal balance.

India is also a decentralized economy where a marked distinction is made between the revenue and expenditure function between the Centre and the states. A detailed analysis of the financing structure of India reveals that the major and elastic tax sources belong to the central government while the state governments are generally bestowed with relatively inelastic sources of revenue. Hence, the revenue of the central government grows faster than the revenue from tax sources administered by the state governments (Thirteenth Finance Commission Report 2009). On the expenditure front, the central government is required to maintain macroeconomic stability and international trade relations and a large number of developmental functions having bearing on state government fiscal measures. On the other hand, state governments are assigned with the developmental and administrative functions, which involve substantial amount of revenue resources. The mismatch between the limited resource power and vast expenditure obligations of the states has accentuated the problem of VFI. In India, these vertical gaps are generally filled by transfers from the Centre to the states which include states' share in central taxes, grants from the Centre and loans from the Centre. These transfers are commonly known as vertical transfers.

Now, most of the papers addressing VFIs in the context of India lack any direct examination on the nature and pattern of VFIs across the states. The papers confirm that VFIs are inescapable, in India's case, and large VFIs are detrimental to fiscal

---

[1]Spending decentralization refers to assignment of expenditure functions to the lower tier of the government from the Centre.

[2]Revenue decentralization refers to devolution of revenue-raising capability to the lower tier of the government.

performance. However, no such attempts are made to ascertain the relation between VFI and fiscal performance for states in India. Thus, in the present study an attempt has been made to find the existence of such a relation in Indian context. For this purpose, econometric method involving panel data estimation technique has been used.

The rest of the paper is structured as follows: In Sect. 2, we provide a brief review of literature on the link between VFI and fiscal performance. Section 3 studies the evolution of VFI in the Indian scenario. Section 4 provides some stylized facts on the nature and pattern of VFI across 24 states in India. Section 5 delivers an economic estimation on the link between VFI and fiscal performance in states in India, while Sect. 6 concludes the paper.

2 Vertical Fiscal Imbalance and Fiscal Performance: A Review of Literature

A vertical imbalance refers to disproportionate allocation of expenditure and revenue between the different tiers government. Many a time, it is found that the devolution of spending responsibilities has not always gone hand in hand with the devolution of revenues resulting in vertical imbalances. Usually, the union government is allocated "redistributive" and "stability" functions, whereas the state governments are asked to carry out "allocative" functions. The state governments are assigned several responsibilities involving public expenditure on the grounds that these tiers of the government are nearest to the people, more sensitive to their needs, and thus would be better able to discharge such responsibilities. Moreover, vertical fiscal imbalance is accentuated by the fact that not only does the union government have major tax sources, the revenue from these sources grows faster than the revenue from tax sources administered by the state governments. Thus, VFI between different layers of government is built by provisions relating to powers of taxation. Herperger (1991) pointed out "perhaps the most striking feature about fiscal powers in nearly all federal systems is that the majority of major revenue have placed under federal authority". These factors put the union governments in an exalted position over the state governments in so far as financial resources are concerned. Consequently, there is an inverse relation between quantum of own revenue and size of expenditure at the two levels of the federal government, eventually creating vertical fiscal imbalance in the federation. Subsequently, the sub-national authorities have to rely on intergovernmental transfers and borrowing to finance their expenditure.

The normative literature on VFI posits that large VFIs may relax fiscal discipline. It has been argued that although some degree of mismatch between sub-national own revenue and sub-national expenditure is inevitable and even may be desirable, large gap presents risks. The theoretical literature often emphasizes the risks associated with large VFIs. Hagen and Eichengreen (1996) have pointed out that as VFI increases, the borrowing restrictions of the sub-national government increases, which in turn upsurges the debt exposure of the central government. They pointed out that if the expenditure of the sub-national government is financed by its own taxes, then there is significantly less restriction on borrowing. However, when the central government imposes restriction on borrowing by the sub-central government, then at the time of transitory shocks the government has to bail out the sub-central government, and hence, the debt exposure of the central government increases. In short, they are more heavily exposed to debt. Thus, VFI is often identified as an obstacle to sub-national accountability and good fiscal performance. A common view is large VFI may lead sub-national governments to overspend and/or lower their tax efforts.

The literature that views government resources as "common property" shows that fragmented policy-making effective through sub-national government may generate excessive fiscal deficits. This happens because interest groups having access to the common pool of government resources do not fully internalize the cost of expenditure programs that benefit their constituencies and disregard the externalities arising from their expenditure decisions. The sub-national governments may perceive that the cost of additional spending can be off-loaded onto others. Velasco (2000) advocated that if government's net assets (i.e. present value of future income stream minus outstanding debt) are the common property of all fiscal authorities, then a problem arises that is quite similar to the "tragedy of commons"—each agent under saves or overspends in case of fiscal policy. In this case, deficits and debts are incurred and accumulated even in contexts where there is no incentive for intertemporal smoothing.

The "soft budget constraint" literature also argues that VFI affects fiscal performance. When fiscal decentralization is not balanced in terms of tax and expenditure allocation, then to cope up with the imbalance, the lower-level (state) government resorts to funds channelled from the federal government in the form of grants and debt. The imbalance followed by pressures from the voters, civil servants and creditors is likely to be directed at the federal government, which has no choice but to bail them out. The sub-national governments, on their part, anticipating such bailout may follow lax policies. Since bailout in any one region is paid for by all other regions, a particular region may not have sufficient incentive to put in high effort. So, a "soft budget constraint" leads to over provision of public good. Rodden (2003) suggested such a situation of "soft budget constraint" occurs when a state government can "manipulate its access to funds in undesirable ways". More specifically, when federal governments fail to generate no bailout expectations, states have the incentive to overborrow or pay insufficient attention to the quality of public provision financed by their borrowing. This readiness to finance state deficits thus boils down to a commitment problem on behalf of the central government, which state governments seek to turn to its advantage. Goodspeed (2002) pointed out that the bailout of regional governments creates a regional "soft budget constraint" which results in two incentive effects: a common pool effect on taxpayers and an opportunity cost effect. The "soft budget constraint" lowers the opportunity cost of borrowing for the region but increases the tax cost, since a portion of the borrowing must be paid through increased taxes. When additional increased taxes are high enough to offset the lower opportunity cost resulting from bailout, then only efficient borrowing decisions may be undertaken by the sub-national governments. Otherwise, VFIs may lead to excessive and unproductive spending, and inefficient revenue mobilization.

The empirical literature on VFIs mainly draws on case studies. The literature is available for both developed and developing countries (Ma 1997; Rodden 2003; Karpowicz 2012). Only a few papers relate vertical imbalances to fiscal performance. Rodden (2002) provided evidence that higher reliance on intergovernmental transfers worsens the general government's overall balance, especially when sub-national governments have high borrowing autonomy. Similarly, Plekhanov and Singh (2007) found that the rules constraining sub-national borrowing improve fiscal performance when transfer dependency is high. In a sample of federations, Rodden and Wibbels (2009) showed that transfer dependency is associated with larger fiscal deficits, the negative impact being larger at high levels of decentralization. Jin and Zou (2002) found that transfers increase the size of the government at the sub-national, national and general government levels. Finally, according to De Mello (2007), transfer growth may become endogenous, with deficits bringing about more grants from the Centre, which in turn generate higher deficits. Thus, these papers find that large VFIs are generally associated with poor fiscal performance. In contrast, Baskaran (2010) found no effect of transfer dependency on public debt. Recently, Eyraud and Lusiyan (2013) echoed some concern about the process of fiscal decentralization. Ambitiously, the paper seeks to measure a causal link between VFI and aggregate government deficits. They concluded that the general government fiscal balance is found to increase by 1% of GDP for each 10 percentage point reduction in VFI.

3 On Vertical Fiscal Imbalance in States in India

India is a federal structure, in which a clear distinction is made between the union and state functions and sources of revenue, but the residual powers belong to the Centre. The Indian Constitution, in its Seventh Schedule, assigns powers and functions of the Centre and states. The schedule specifies the exclusive powers of the Centre (the union list) and the states (the state list), and those under joint jurisdiction (the concurrent list).[3] In brief, the functions of the central government are those required to maintain macroeconomic stability, international trade and relations, and those having implications for more than one state. On the other hand, the major subjects assigned to the states comprise public order, public health, agriculture, irrigation, land rights, fisheries and industries and minor minerals. The states also assume a significant role for subjects in the concurrent list, such as education and transportation, social security and social insurance.

As regards the assignment of tax power, it is based on the principle of separation: most broad-based taxes are assigned to the Centre, whereas in practice the states have

[3] A detailed list of union and state expenditure functions as mentioned in the Seventh Schedule of Indian Constitution is provided in Tables 4, 5 and 6 in Appendix.

a narrower tax base.[4] A scrutiny of the taxation powers contained in the union list and the state list reveals that major and elastic sources of tax revenue (like personal income tax, corporation tax, excise duty on manufactures) belong to the Centre, while relatively inelastic sources of revenue (like tax on purchase and sale of commodities) come under the purview of state governments. Moreover, the Centre enjoys almost unlimited powers to borrow in the domestic market as well as from abroad, whereas borrowing powers of the states are subject to various restrictions.

The consequence of this mismatch between the limited resources and vast expenditure obligations is vertical fiscal imbalance for the states. Rao (2000) figured out that the states raise 35% of total revenues, that finances 51% of their expenditures. The expenditure share of the state governments after giving grants to local governments is 35%. In social services, particularly in education and health sectors, the expenditure share of the states is more than 80%, and in economic services, it is about 50%. The constitutional provisions relating to allocation of financial resources between the Centre and the states might not be sufficient to cover for the assigned expenditure programmes of the state governments. To correct for such built-in vertical imbalance, the arrangement of transfer of funds from the Centre to the states through the mediation of Finance Commissions has been envisaged. These transfers are commonly known as vertical transfers. This distribution is in the form of share of states in central taxes, grants from the Centre and loans from the Centre. As pointed out by Rangarajan and Srivastava (2008) "…. (a) prior to transfers, Centre collects on average about 63–64% of the combined revenue receipts; after transfers, states get nearly 64% of the combined revenue receipts. (b) This enables the states to spend nearly 57% of the combined expenditure on an average on revenue account. The Centre spends about 43% of the combined revenue expenditure by retaining 36% of revenues after transfers by borrowing relatively more". Consequently, the states rely on central transfers to finance a significant portion of their expenditures. Thus, the higher the vertical imbalance for the states the larger will be the transfer of funds from Centre to states.

Now, most of the papers addressing VFI have focused mainly on the nature and trends of intergovernmental transfers (Rao 2000; Rangarajan and Srivastava 2008; Singh 2007; Mohan and Shyjan 2009; Rao and Srivastava 2014). Much of the literature is confined to the discussion of role of transfers as gap fillers to combat VFIs and the pattern of changes and variations across the states of such transfers. Although the papers confirm that VFIs exist among states, the literature lacks any direct examination on the nature and pattern of VFIs and its impact on fiscal performance across states in India.

The current study attempts, using a proper measure of VFI, to scrutinize the nature and pattern of VFI across states in India. An econometric estimation is carried out to ascertain the relation of VFI to fiscal performance in the context of Indian states.

[4] A detailed list of union and state taxes as laid down in Seventh Schedule of Constitution of India is provided in Tables 7 and 8 in Appendix.

4 A Measure of Vertical Fiscal Imbalance

The twin objective of analysing the nature and its impact on the fiscal discipline of states in India encompasses a proper measure of VFI. Various indicators of VFI are used in the empirical literature (Sharma 2012).

Boadway and Tremblay (2006) have defined VFI as a situation in which the size of transfers made by the federal government to the provinces falls well short of the amount of federal tax revenues relative to their expenditure responsibilities. Most of the papers have resorted to "transfer dependency" as an appropriate measure of VFI, where transfer dependency is defined as transfers as a share of either sub-national spending or sub-national revenue. While some authors have used the difference between own revenues and own spending as a measure of VFI (Bird and Tarasov 2004).

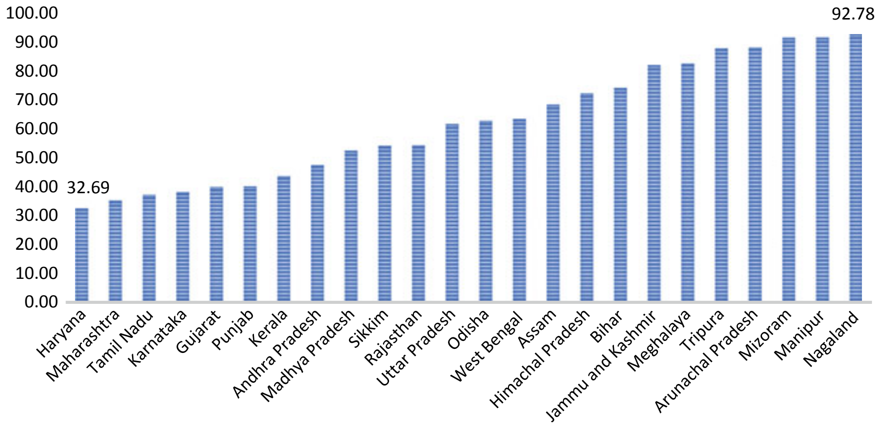
This paper uses the concept of VFI as propounded by Eyraud and Lusiyan (2013), where VFI is defined as the share of sub-national own spending not financed through own revenues, because of some its advantages over the other measures. Along with being easy to calculate in the context of Indian states, this measure extends the concept of "transfer dependency" to sub-national borrowing. Inclusion of sub-national borrowing is important since it is another kind of "soft" resource of financing the fiscal imbalances and important contributor of VFI dynamics. Further, this VFI indicator measures the mismatch between spending and revenue decentralization most appropriately. Thus,

$$\text{VFI} = 1 - \frac{\text{Sub-national Government (SNG) Own Revenue}}{\text{Sub-national Government (SNG) Own Expenditure}}$$

5 An Assessment of Vertical Fiscal Imbalance in Indian States

Before proceeding to examine the main issues of the study, let us look into some stylized facts on VFI, their dispersion across states, the mode of financing these imbalances and their long-run trends. We use data for 24 Indian states for a period of 1990–91 to 2015–16. Among the 24 states, 14 are general category states: Andhra Pradesh, Bihar, Gujarat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal, and 10 are special category states[5]: Arunachal Pradesh, Assam, Himachal Pradesh, Jammu and Kashmir, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura.

[5]Special category states are certain disadvantaged states with preferential treatment in the form of central assistance and tax break.



Source: Author's computation

Fig. 1 Average Vertical fiscal imbalance (in %)

Fact 1: There are sizeable differences in vertical fiscal imbalance across the states

The average VFI is 62.43% over the sample from 1990–91 to 2015–16.

Figure 1 presents a large dispersion in VFI across the 24 states. The variation is from about 32.69% in Haryana to 93% (approx.) in Nagaland. This implies the financing of the states' expenditure varies greatly across states.

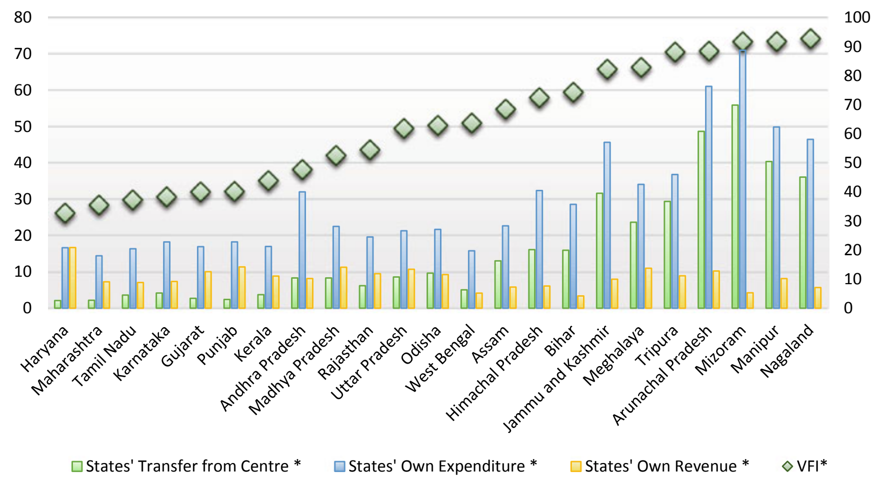
However, these cross-state differences in the structure of state finances may reflect factors such as the role of the states as provider of public services (especially health and education), regional imbalances and the presence of externalities amongst others. For example, VFIs are much higher for the special category states compared to the general category states.

Special category states are certain disadvantaged states having features like hilly and difficult terrain, low population density, economic and infrastructural backwardness, and non-viable nature of state finances. For this, they have low resource base and cannot mobilize resources for development. Hence, the gap between expenditure incurred and revenue raised by these states is wide and faces large vertical imbalances.

The dispersion of VFI across the general category states attributes to the wide variation in the revenue-raising capacity of the states, except for Andhra Pradesh (Fig. 2). As for the special category states, the heterogeneity is mostly related to the dispersed expenditure obligations of the states. Transfers from the Centre are pivotal for financing of the revenue gap.

Fact 2: Shares in central taxes and grants from Centre are essential in understanding VFI dispersion across states

To deal with VFI, the Finance Commission of India (under Article 280) recommends vertical transfers to states from the Centre. These transfers act as an instrument



* Percentage of NSDP current prices

Note: Sikkim has been dropped due to incomparable data

Source: Author's computation

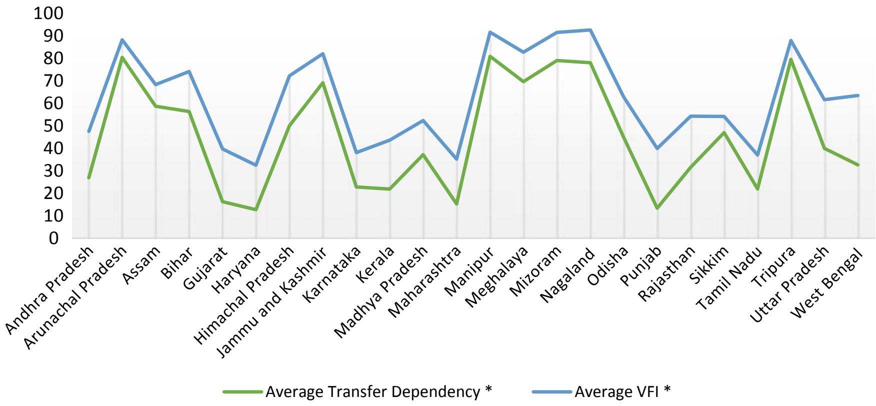
Fig. 2 VFI dispersion across states

to deal with the extant asymmetric decentralization of expenditure responsibility and revenue-raising authority of the states. In the context of Indian states, the transference of resources from Centre comes to occupy an important place in correcting the imbalances. The importance of Central contributions to states' resources is clear from Fig. 3. VFI and Transfer Dependency of the states (average over the period 1990–91 to 2015–16) have moved analogously.

In the process of devolution of resources from the Centre to states, there are three major modes of transfers, viz. states' share in central taxes, grants from the Centre and loans from the Centre.

The proportion of shares in central taxes to gross transfers has been highest for the general category states (Fig. 4). Now, the criteria for devolution of taxes across states, as used by the Finance Commission, are mainly population, distance, tax effort, fiscal discipline, etc.[6] Hence, the states which satisfy most of the criteria are the ones to accrue the highest proportion of share in central taxes to correct their imbalance. Thus, the general category states receive a greater proportion of transfers in the form of shares in central taxes. On the other hand, the share of grants in aid from Centre has been the highest for the special category states. The share of loans from Centre to gross transfers should not be overlooked. Even though they constitute

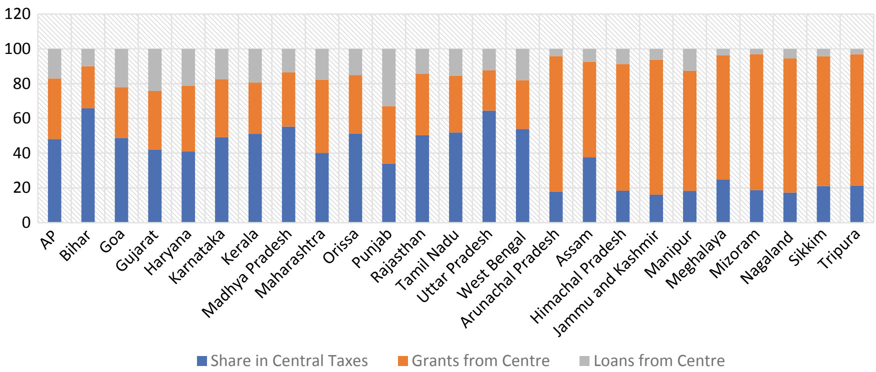
[6]The criterion for tax devolution has varied from one Finance Commission (FC) to another. For example, prior to 11th FC, multiple criteria like collection, backwardness, poverty ratio were for devolving taxes. The 11th FC used population, distance, area, index of infrastructure, tax effort and fiscal discipline as the basis of devolution. The 12th FC, on the other hand, dropped index of infrastructure from the criteria.



Transfers (Share in central taxes + grants from centre) as a share of States' Own Expenditure

Source: Author's computation

Fig. 3 Average VFI and average transfer dependency (in per cent)



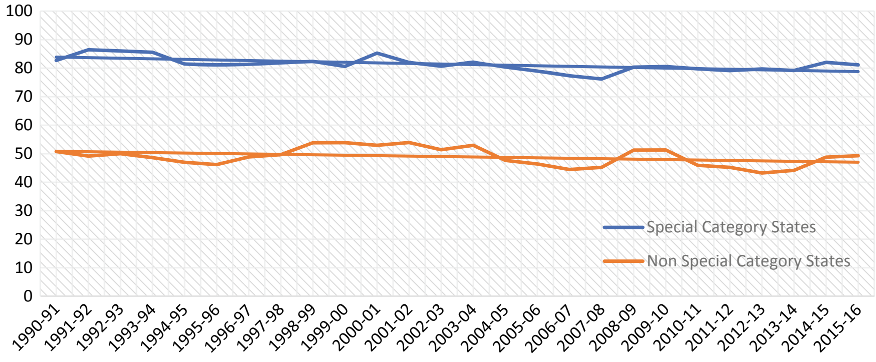
Source: Handbook of Statistics on State Government Finances, 2010 (Reserve Bank of India Publications)

Fig. 4 Components of gross transfers (Average over 1990–91 to 2009–10)

a small portion of gross transfers, the general category states depend more on them as compared to the special category states (Fig. 4).

Fact 3: VFIs have remained more or less the same in the last 20 years.

On an average, for the general category states, VFI declined by 2.93 percentage points from 1990–91 to 2015–16 and 1.89%, alternatively, for the special category states (Fig. 5).



Source: Author's computation

Fig. 5 Average vertical Fiscal imbalance—Evolution overtime (in %)

A fluctuating trend is noted for the general category states. VFI has been lowest in 2012–13 (43.24%).[7] As for the special category states, the trend has more or less been the same (Fig. 5).

Fact 4: Large vertical imbalances are associated with worst fiscal performance in the context of Indian states

As the literature posits, large VFIs relax fiscal discipline. The overall fiscal performance of the states is found to be positively associated with large vertical imbalances (Fig. 6). On an average, states' overall deficit is found to increase by 41 percentage point for every one per cent increase in VFI.

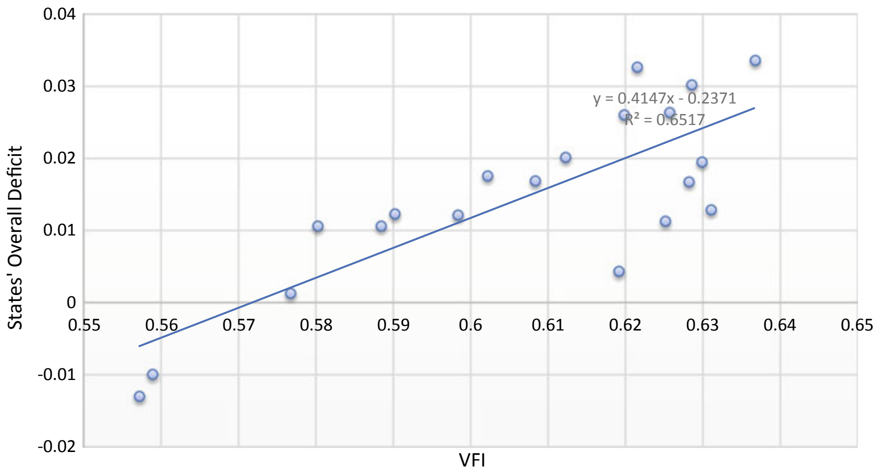
From this, it is quite evident that in almost all the states, the expenditure decentralization is not in tandem with the revenue decentralization in India.

Fact 5: "Finance follows function" is evident only for selected Indian states

Conventionally, the literature suggests that revenue decentralization should always follow expenditure decentralization. However, states' experience in India reveals the other way. For majority of the states, we find no causality. A reverse causality, revenue decentralization preceding expenditure decentralization, is found for six states.

"Finance follows function" is evident only for Meghalaya. The reason sorted for this can be due to the fact that devolution of revenue powers is easy to implement compared to expenditure decentralization. However, for majority of Indian states no specific pattern is observed, at least in the short run.

[7] The long-run trend of VFIs for individual states is provided in Fig. 7 in Appendix.



Source: Author's computation

Fig. 6 Fiscal performance and VFI

Granger non-causality test results[a]

	Expenditure decentralization → Revenue decentralization	Expenditure decentralization + Revenue decentralization
Revenue decentralization → Expenditure decentralization		Andhra Pradesh, Karnataka, Rajasthan, Arunachal Pradesh, Assam, Nagaland
Revenue decentralization + Expenditure decentralization	Meghalaya	

Notes (i) Decentralization figures in logarithms; lags chosen by Akaike information criteria (AIC)
(ii) X → Y: X Granger causes Y; X + Y: X does not Granger cause Y
(iii) Significant at 5% level
[a]To detect Granger causality in panel data set, Panel Granger causality test (Dumitrescu and Hurlin 2012) is applied. The detailed result is provided in Tables 9 and 10

Hence, as evident from the above discussion, all the states in India face large and widening vertical gaps. In this backdrop, it will be interesting to evaluate the impact of this persistent VFI on the fiscal performance of the states. In the next section, we provide an empirical estimation of link between VFI and fiscal performance in the context of states in India.

6 Econometric Evidence

(a) ***Methodology***:

A panel data estimation is carried out á la Eyraud and Lusiyan (2013) to assess the impact of VFI on the fiscal performance of the states.

An empirical relationship is specified linking the state government primary deficit to the VFI, spending decentralization and covariates. The main aim is to estimate the partial effect of VFI.

The following specification is applied to evaluate the relationship:

$$\mathrm{PD}_{it} = \alpha \times \mathrm{VFI}_{it} + \beta \times \mathrm{Decentralization}_{it} + \delta \times X_{it} + \varepsilon_{it} \quad (1)$$

where indices i and t denote states and years, respectively.

PD_{it} is the dependent variable, which is primary deficit of state government as a share of $\mathrm{NSDP}_{\text{current prices}}$. This is used as a proxy for fiscal performance.

VFI_{it} is the vertical fiscal imbalance. $\mathrm{Decentralization}_{it}$ is spending decentralization. X_{it} denotes control variables. The covariates include government debt and output gap. All these are the independent variables.

ε_{it} is time and state-specific error term.

The regression equation is applied on a sample of 24 states over the period 1995–96 to 2014–15. A two-way fixed effect estimation of Eq. (1) detected some degree of autocorrelation and group-wise heteroscedasticity in the residuals. Hence, to ensure that the statistical inference is valid, Eq. (1) is estimated with the least square dummy variable (LSDV) estimator with robust standard errors clustered at the state level.

The empirical analysis tests the following three hypotheses related to the impact of VFI on fiscal performance.

H1: Shifting the financing of state expenditures from transfers to own revenue improves fiscal performance.

In line with the theoretical literature and the facts, as obtained in the previous section, the VFI is included in Eq. (1) as a determinant of state government primary deficit. Thus, we can expect that raising the share of own revenue of the states to finance their expenditure will improve fiscal performance. The coefficient α represents the impact of VFI, keeping spending decentralization constant. Estimated α is expected to be positive. Thus, given VFI decreases, primary deficit of the state government also decreases.

H2: Too large vertical gaps are bad for fiscal performance, but too small as well.

H2 tests whether or not the effect of VFI is linear. As pointed out in Sect. 2, high VFIs present risks. Eyraud and Lusiyan (2013) also mentioned that a small VFI may have adverse effect on fiscal performance, because of diseconomies of scale.

This hypothesis is tested by including a quadratic term in the equation. According to Eyraud and Lusiyan (2013), "If there is an 'optimal' positive VFI maximizing the fiscal balance,[8] the VFI coefficient should be positive and squared VFI coefficient should be negative". This paper uses the concept of primary deficit instead of fiscal balance. Hence, the desired sign of the coefficient of VFI and squared VFI will be negative and positive, respectively.

H3: The effect of VFI is more pronounced in states with large horizontal fiscal imbalances (HFI).

[8]They defined fiscal balance as difference between general government total revenue and primary expenditure (i.e. expenditure net of interest payment).

HFI refers to differential fiscal capacities of sub-national economies. HFI is the VFI which is left over when the VFI is solved for the richest sub-national government (Bird and Vaillancourt 2006). Studies suggest VFI and HFI may interact with each other and can be detrimental to fiscal performance.

As apparent from the scatter diagram,[9] HFI among the 24 major Indian states has been persistently increasing over the years. It is also evident from the coefficients of variation of the states' HFI which has been on a rise over time. The persistent HFI can be due to differentials in revenue-raising capacity and cost disability factors among the states.

This hypothesis aims at testing the combined effect of VFI and HFI on fiscal performance of the states in India. The estimated coefficient is expected to be positive.

Given no common consensus on the measure of HFI, distance of per capita income from the average of highest three per capita income is taken as a measure of HFI. In Indian context, when the level of per capita incomes is highly unequal, per capita income distance is one of the main criteria for distribution of equalization grants.[10] It meets the horizontal equity criteria of transfers (Rangarajan and Srivastava 2008). Hence, per capita income deviation can be considered as an appropriate proxy for HFI.

(b) ***Data Description***:

To carry out the above-mentioned estimation, the paper employs a sample of 24 Indian states over the period 1995–96 to 2014–15. The relevance of choosing the sample period relies upon the fact that three-tier devolution of fiscal responsibility and resources has been constitutionalized in India from the year 1993 onwards.[11] Thus, any study assessing the impact of vertical fiscal imbalances on the fiscal performance of the states should take into account such provisions, as laid down in the constitution.

The variables used in the econometric analysis are as follows:

(i) $\mathrm{PD}_{it} = $ State government primary deficit as a share of $\mathrm{NSDP}_{\text{current prices}}$

As a proxy of fiscal performance, primary deficit is used. Primary deficit is the standard deficit minus the interest on the government debt. Equivalently, as traditionally measured, the primary deficit is government non-interest outlays minus total revenues.

(ii) $$\text{VFI(Vertical Fiscal Imbalance)} = 1 - \frac{\text{State Government Own Revenue}}{\text{State Government Own Expenditure}}$$

It is the share of state government's own expenditure (i.e. revenue expenditure excluding grants in aid and contributions and capital expenditure excluding discharge

[9] Refer Fig. 8 in Appendix.

[10] As suggested by the Finance Commission Reports.

[11] Two constitutional amendment (73rd and 74th in the year 1993 and 1994) established mandatory provisions for decentralization to local governments in India. It envisaged the panchayat and municipal bodies as institutions of self-government. The states endow the lower-tier governments with powers and authority to enable the latter to function as institutions of self-government.

of internal debt and repayment of loans to Centre) not financed with state government's own revenue (i.e. revenue receipts excluding share in central taxes and grants from Centre).

(iii) $$\text{Expenditure decentralization} = \frac{\text{Devolution and Transfer of Resources from Centre}}{\text{States' Own Expenditure}}$$

Devolution and transfer of resources from Centre include share in central taxes and grants from Centre to the state.[12]

(iv) The covariates used are lag debt-to-NSDP gap and lag output gap, where

Debt-to-NSDP gap ratio = State government's gross outstanding liabilities as a ratio of $\text{NSDP}_{\text{current prices}}$

$$\text{Output gap} = \frac{\text{Actual NSDP}_{\text{constant prices}} - \text{Estimated potential NSDP}}{\text{Estimated potential NSDP}}$$

Estimated potential NSDP is obtained by Hodrick–Prescott filter of $\text{NSDP}_{\text{constant prices}}$ for each of the 24 states over the period 1995–96 to 2014–15.

(v) Two alternate definitions of HFI are used:

(a) HFI = Average of highest three per capita $\text{NSDP}_{\text{current prices}}$ − State per capita $\text{NSDP}_{\text{current prices.}}$

(b) HFI = Fiscal capacity of each state − Median fiscal capacity of states concerned. Fiscal capacity of the states is derived from the use of representative tax system (RTS).[13]

All the data are obtained from Handbook of Statistics on Indian Economy, 2018 (Reserve Bank of India), Handbook of Statistics on Indian States, 2013–18 (Reserve Bank of India), State Finances: A Study of Budgets, 2013–18 (Reserve Bank of India) and Handbook of Statistics on State Government Finances, 2010 (Reserve Bank of India).

---

[12]The literature suggests various definitions of decentralization. One of the most used is "share of sub-national expenditure to general government expenditure". This definition of decentralization is inapplicable in the context of India due to the unavailability of comprehensive expenditure data at local and provincial governments. Thus, the use of this data can present an incomplete picture of the sub-national government. However, the expenditure responsibilities of all tiers of government, specifically the lower tiers, are mostly met through taxes and grants as recommended by the Central Finance Commission. Hence, the share of transfer of resources from Centre to own expenditure of the states might represent the degree of decentralization in India accurately.

[13]Complete methodology given in Appendix.

(c) ***Econometric Results***:

Baseline Results:

This section presents the main econometric results.

Tables 1 and 2 show the results in the context of India states. The sample of 24 states includes 14 general category and 10 special category states.[14]

With the distinction of Indian states into non-special and special category, it would be preferable to run separate regressions to estimate the impact of VFI on the fiscal performance of the states, due to their incomparable data. While Table 1 reports the results for non-special category states, Table 2 provides the results found for special category states.

Column (2) adds the covariates lag debt-to-NSDP ratio and lag output gap. Column (3) verifies H2.

The results from Tables 1 and 2 show that all coefficient signs (except for lag output gap) are according to the hypotheses postulated in the paper. VFI is found to exacerbate primary deficit. The primary deficit of the general category state governments is found to deteriorate by 12 percentage points of NSDP for each percentage point decline in VFI. The special category states get the highest estimated coeffi-

Table 1 Regression result on relationship between VFI and fiscal performance for 14 major general category states

Dependent variable: primary deficit–NSDP ratio			
	(1)	(2)	(3)
VFI	0.12*** (10.19)	0.14*** (12.72)	0.14*** (3.83)
Expenditure decentralization	−0.16*** (−9.91)	−0.14*** (−10.48)	−0.17*** (−10.33)
Lag debt–NSDP ratio		−0.04*** (−4.31)	−0.04*** (−6.54)
Lag output gap		0.01 (0.57)	0.002 (0.18)
VFI squared			−0.02 (−0.40)
Overall R^2	0.52	0.61	0.49
Within R^2	0.66	0.73	0.74
Number of observations	280	280	280
Number of states	14	14	14

*** (**,*) = Significant at 1(5, 10) per cent level. The robust *t*-statistic (heteroscedasticity corrected) is represented in the parenthesis

Column (2) reports random effect results, as suggested by the Hausman test

[14]One may argue that there are differences in the physical and economical characteristics of the two category states. Therefore, they are not directly comparable. Hence, we may get misleading results when the two category states are clubbed together. To get more accurate results, separate regression estimation is carried out for the general category states and special category states.

Table 2 Regression result on relationship between VFI and fiscal performance for 10 major special category states

Dependent variable: primary deficit–NSDP ratio			
	(1)	(2)	(3)
VFI	0.27*** (7.39)	0.24*** (5.25)	0.55** (2.47)
Expenditure decentralization	−0.35*** (−13.71)	−0.34*** (−11.70)	−0.36*** (−11.89)
Lag debt–NSDP ratio		−0.05*** (−4.07)	−0.06*** (−4.34)
Lag output gap		0.03 (0.81)	0.04 (1.04)
VFI squared			−0.24 (−1.50)
Overall R^2	0.46	0.29	0.18
Within R^2	0.63	0.62	0.63
Number of observations	200	150	150
Number of states	10	10	10

*** (**,*) = Significant at 1(5, 10) per cent level. The robust t-statistic (heteroscedasticity corrected) is represented in the parenthesis

Notes Due to the unavailability of data on real NSDP for Mizoram (1995–96 to 1998–99), the sample period is 2000–01 to 2014–15 in column (2) and (3)

cient of VFI. The primary balance of the special category states is found to improve by 27 basis points for every 1 percentage point decline in their respective VFI (as reported in Table 2). Hence, to improve fiscal performance of the states VFI should be reduced. The econometric analysis supports H1.

Spending decentralization is significantly negative suggesting higher is the devolution of resources to the states, given their expenditure responsibilities, the lower is the primary deficit of the general category states. Thereby, it validates H1.

Lag debt-to-NSDP ratio is significant. It implies the fiscal policy of the Indian states are constrained by debt sustainability considerations. Thus, debt in the previous period is accompanied by primary surplus in the current period to stabilize the debt. Lag output gap found to be statistically insignificant for both the categories of states indicating no effect of automatic stabilizers in the system.

The econometric analysis provides no support for H2. Squared VFI is insignificant, as implying too large or too small VFI has no such effect on the fiscal deficit of the states.

To summarize, the theoretical claim that VFI is detrimental to fiscal performance is empirically supported in the context of India. A decline in VFI is always found to be beneficial for the Indian states. A 1 percentage point reduction in the VFI reduces the primary deficit of the state governments by 15 percentage point[15] of NSDP, on

[15]Clubbing general and special category states, together. Result is provided in, Table 11.

average. Thus, to improve the fiscal position the central and the state government should make efforts to reduce VFI.

Result with HFI:

The econometric analysis conforms H3. HFI as an interacting variable makes the effect VFI on fiscal balance more interpretable.

As discernible from the regression result, VFI is more pronounced to fiscal performance in states with high HFI. However, the average effect of VFI on primary deficit of the states has remained relatively same in the presence of HFI. HFI also seems to have a significant positive effect on the primary balance, though its impact has been comparatively negligible (Table 3).

Robustness Analysis:

The main results found can be sensitive to how the variables are defined. This section re-estimates the main regression results with some alternate definitions of the regression variables. The reiteration of estimated results with alternative definitions shall provide additional insights to the present context of study.

Table 3 Regression results with HFI as interaction term

Dependent variable: Primary deficit–NSDP ratio				
	General category states		Special category states	
	(1)	(2)	(1)	(2)
VFI	0.12*** (11.64)	0.14*** (14.06)	0.28*** (5.72)	0.32*** (6.58)
Expenditure decentralization	−0.19*** (−11.63)	−0.19*** (−12.31)	−0.35*** (−11.46)	−0.34 (−10.95)
Lag debt–NSDP ratio	−0.04*** (−6.80)	−0.04*** (−7.01)	−0.04*** (−3.39)	−0.04*** (−3.23)
HFI		0.00*** (3.92)		0.00** (1.98)
VFI × HFI	0.00*** (4.96)	0.00*** (2.51)	0.00*** (3.69)	−0.00 (−1.47)
Overall R^2	0.52	0.77	0.40	0.41
Within R^2	0.76	0.57	0.65	0.66
Number of observations	280	280	150	150
Number of states	14	14	10	10

Fixed effect estimation as suggested by Hausman and Sargan–Hansen test of choosing between FE and RE model; Robust *t* statistics (heteroscedasticity corrected) clustered at state level in parenthesis; *** (**,*) = Significant at 1(5, 10) per cent level

Since the interaction term coefficient is found to be approximately equal to 0, even though statistically significant, the combined effect of VFI around the *centred* value of HFI is ignored

The rigorous sensitivity tests run as follows. It incorporates alternate indicators of VFI and fiscal performances.[16]

Lag output gap, again, is found to be insignificant in the context of Indian states possibly reflecting the absence of automatic stabilizing effect of fiscal policy. The inclusion of lag output, in place of lag output gap, as a covariate can elucidate the results on the effect of VFI on fiscal performance of Indian states.

Further, the literature on estimation of HFI across sub-national economies emphasizes on the differences in the fiscal capacity measure. The HFI measure applied in the main regression results is absolute measures which might lead to biased results due to variability in the data. Given non-concurrence on any particular definition of HFI, the robustness analysis incorporates some alternate (normalized) measures of HFI.

Results of robustness analysis[17] confirm the main regression results. The main variables are still found to be significant on re-estimation of baseline results in reference to the above-mentioned alternate measures of various regression variables. The coefficients of VFI are found to be negatively significant conforming that VFI is detrimental to fiscal performance. The primary deficit of general category states is found to improve by 9 (1) percentage points of NSDP for each percentage point decline in VFI for general category states.

As found in the main regression results, the special category states get the highest estimated coefficient for VFI. The econometric analysis again supports H1. However, the impact of VFI in the presence of HFI is found to be insignificant.

With alternate definitions of fiscal performance, the VFI coefficients are also found to be significantly high. For the special category states, HFI is again seemed to have a positive effect on the fiscal performance alternatives. It furthermore validates H3 for the special category states.

Results with HFI as interaction term give mixed outcomes. Primary deficit of the general category states is found to decline, in the presence of VFI, if the per capita NSDP of the states is greater than average per capita NSDP of top three richest states.

For special category states, VFI is found to affect primary deficits inversely, if deviation of each state's estimated fiscal capacity from median capacity is positive. Thus, for special category states, high VFI is found to reduce primary deficits in the presence of uneven fiscal capacity across states. This could be because of the fact that the greater the disadvantageous position of the special category states the higher the transfer of funds to the states from the Centre.

The dynamic panel data model takes into account the possibility that all explanatory variables are endogenous. The test for autocorrelation and over-identifying restrictions presents no evidence of model misspecification. The results of the model are corroborating the finding that VFI is detrimental to fiscal performance of the states. All the estimated coefficients are statistically significant and have the usual expected sign for the general category states (as reported in the previous results). An anomaly is noted for the special category states where lag debt–NSDP ratio and

[16]For alternate measures of VFI, fiscal performances and HFI, refer Appendix.

[17]Refer Appendix, for robustness analysis results.

lag primary deficit are insignificant. It implies past year deficit exerts no substantial influence on current year primary deficit of the special category states.

7 Conclusion

Thus, theoretical and empirical literatures have provided that VFI is detrimental to fiscal health. Given the financing structure of the sub-national governments, VFI is inevitable in almost all the sub-national governments around the world. However, VFIs come into greater scrutiny during difficult times. Attention to VFIs increases when deficits become chronic, necessitating frequent bailouts from the central government. Given the financing structure of the Indian states, the notion of VFI is also apparently prevalent in Indian states. This paper presents an analysis of the nature, pattern and financing structure of VFI in 24 states in India for the period 1995–96 to 2014–15. Further, an attempt is made to examine the exact relationship between VFI and fiscal performance across the above-mentioned sample.

It is found that VFI is quite dispersed across states in India. The analysis has considered two categories of states—general and special. The impact of VFI is comparatively lower in the general category states as compared to the special category states. As regards the financing of VFIs, states' share in central taxes is the major source of finance for the general category states, whereas grants from Centre act as active gap fillers for the special category states. Besides this, it has been found that VFI in India follows a more or less constant trend in the last 20 years. A panel data estimation, on Indian data, confirms the widely held view that closing the gap between sub-national expenditure and sub-national own revenue is associated with better fiscal outcomes. On an average, the primary deficit of the states is found to decline by 15 percentage point of NSDP for each 1 percentage point decline in VFI. The combination of horizontal and vertical imbalance also affects fiscal performance adversely. Thus, to improve the fiscal performance of the states, attempt must be made to reduce VFI.

However, this analysis is an initial approach to investigate the nature of VFI and its link with fiscal performance, in the context of Indian states. The empirical research in this area can be further strengthened in a number of directions. This analysis has provided only the nature of VFI in states in India. Attempts can be made to assess the causes of such pattern. For example, efforts can be carried out to find why VFIs are so widely dispersed across the states. A detailed study on the finances of individual's states, their revenue-raising efforts, the central policies on state-wise tax devolutions (if any), etc., may provide a better insight into the issue.

Appendix

See Tables 4, 5, 6, 7, 8, 9, 10 and Fig. 7.

Table 4 Developmental functions included in the union lists in the Seventh Schedule of the Constitution

S. No.	Entry No.	Heads
1	6	Atomic energy and mineral resources necessary for its production
2	22	Railways
3	23	Highways declared by or under law made by Parliament to be national highways
4	24	Shipping and navigation on inland waterways, declared by Parliament by law to be national waterways, as regards mechanically propelled vessels the rule of the road on such waterways
5	25	Maritime shipping and navigation including shipping and navigation on tidal waters provision of education and training for the mercantile marine and regulation of such education and training provided by States and other agencies
6	26	Lighthouses, lightships, beacons and other provision for the safety of shipping and aircraft
7	27	Ports declared by or under law made by Parliament or existing law to be major ports, including their delimitation and the constitution and powers of port authorities therein
8	28	Port quarantine, including hospitals connected therewith seamen's and marine hospitals
9	29	Airways aircraft and air-navigation provision of aerodromes; regulation and organisation of air traffic and of aerodromes; provision for aeronautical education and training and regulation of such education and training provided by States and other agencies
10	30	Carriage of passengers and goods by railways, sea or air, or by national waterways in mechanically propelled vessels
11	31	Posts and telegraph: telephones, wireless, broadcasting and other forms of communications
12	41	Trade and commerce with foreign countries; import and export across customs frontiers; definition of customs frontiers
13	42	Inter-state trade and commerce
14	52	Industries, the control of which by the union is declared by parliament by law to be expedient in the public interest
15	53	Regulation and development of oilfields and mineral oil resources; petroleum and petroleum products; other liquids and substances declared by Parliament by law to be dangerously inflammable
16	54	Regulation of mines and mineral development to the extent which such regulation and development under the control of the Union is declared by Parliament by law to be expedient in the public interest
17	56	Regulation and development of inter-State rivers and river valleys to the extent to which such regulation and development under the control of the Union is declared by Parliament by law to be expedient in the public interest

(continued)

Table 4 (continued)

S. No.	Entry No.	Heads
18	57	Fishing and fisheries beyond territorial waters
19	65	Union agenda and institutions for a. professional, vocational or technical training including the training of police officers; or b. the promotion of special studies or research; or c. scientific or technical assistance in the investigation or detection of crime
20	66	Coordination and determination of standards in institutions for higher education or research and scientific and technical institutions
21	68	Survey of India, the geological, botanical, zoological and anthropological surveys of India, meteorological organizations

Table 5 Developmental functions included in the state lists in the Seventh Schedule of the Constitution

S. No.	Entry No.	Heads
1	5	Local government, that is to say, the constitution and powers of municipal corporations, improvements trusts, district boards, mining settlement authorities and other local authorities for the purpose of local self-government or village administration
2	6	Public health and sanitation; hospitals and dispensaries
3	9	Relief of the disabled and unemployable
4	13	Communications, that is to say, roads, bridges, ferries, and other means of communication not specified in List I: municipal tramways; ropeways; inland waterways and traffic thereon subject to the provisions of List I and List II with regard to such waterways; vehicles other than mechanically propelled vehicles
5	14	Agriculture, including agricultural education and research, protection against pests and prevention of plant diseases
6	15	Preservation, protection and improvement of stock and prevention of animal diseases; veterinary training and practice
7	17	Water, that is to say, water supplies, irrigation and canals, drainage, embankments, water storage and water power subject to the provisions of entry 56 of List I
8	18	Land, that is to say, rights in or over land, land tenures including the relations of landlord and tenant, and the collection of rents; transfer and alienation of agricultural land; land improvement and agricultural loans; colonization
9	21	Fisheries
10	23	Regulation of mines and mineral development subject to the provisions of List I with respect to regulation and development under the control of the Union

(continued)

Table 5 (continued)

S. No.	Entry No.	Heads
11	24	Industries subject to the provisions of entries 7 and 52 of List I
12	25	Gas and gas-works
13	26	Trade and commerce within the State subjects to the provisions of entry 33 of List III
14	27	Production, supply and distribution of goods subject to the provisions of entry 33 of List III
15	32	Cooperative societies
16	35	Works, lands and buildings vested in or in the possession of the state

Table 6 Developmental functions included in the concurrent list in the Seventh Schedule of the Constitution

S. No.	Entry No.	Heads
1	17A	Forests
2	20	Economic and social planning
3	20A	Population control and family planning
4	23	Social security and social insurance; employment and unemployment
5	25	Education, including technical education, medical education and universities, subject to the provisions of entries 63, 64, 65 and 66 of List I; vocational and technical training of labour
6	27	Relief and rehabilitation of persons displaced from their original place of residence by reasons of the setting up of the Dominions of India and Pakistan
7	31	Ports other than those declared by or under law made by Parliament or existing law to be major ports
8	32	Shipping and navigation and inland waterways as regards mechanically propelled vessels, and the rule of the road on such waterways, and the carriage of passengers and goods on inland waterways subject to the provisions of List I with regard to national waterways
9	33	Trade and commerce in, and the production supply and distribution of - a. the products of any industry where the control of such industry by the Union is declared by Parliament by law to be expedient in the public interest and imported goods on inland waterways subject to the provisions of List I with regard to national waterways b. foodstuffs, including edible oilseeds and oils c. cattle fodder, including oilseeds and other concentrates d. raw cotton, where ginned or unginned and cottonseed; and e. raw jute

(continued)

Table 6 (continued)

S. No.	Entry No.	Heads
10	36	Factories
11	37	Boilers
12	38	Electricity

Table 7 Taxation heads assigned to the union in the Seventh Schedule of the Constitution

Entry in List I of the Seventh Schedule	Head
82	Taxes on income other than agricultural income
83	Duties of customs including export duties
84	Duties of excise on tobacco and other goods manufactured or produced in India except- a. alcoholic liquors for human consumption b. opium, Indian hemp and other narcotic drugs and narcotics; but including medicinal and toilet preparations containing alcohol or any substance included in sub-paragraph (b) of this entry
85	Corporation tax
86	Taxes on the capital value of the assets, exclusive of agricultural land of individuals and companies; taxes on the capital of companies
87	Estate duty in respect of property other than agricultural land
88	Duties in respect of succession to property other than agricultural land
89	Terminal taxes on goods or passengers carried by railway, sea or air: taxes on railway fares and freights
90	Taxes other than stamp duties on transactions in stock exchanges and future markets
91	Rates of stamp duty in respect of bills of exchange, cheques, promissory notes, bills of lading, letters of credit, policies of insurance, transfer of shares, debentures, proxies and receipts
92	Taxes on the sale or purchase of newspapers and on advertisements published therein
92A	Taxes on the sale or purchase of goods other than newspapers, where such sale or purchase takes place in the course of inter-State trade or commerce

(continued)

Table 7 (continued)

Entry in List I of the Seventh Schedule	Head
92B	Taxes on the consignment of goods (whether the consignment is to the person making it or to any other person), where such consignment takes place in the course of inter-State trade or commerce
97	Any other matter not enumerated in List II or List III including any tax not mentioned in either or both the Lists

Table 8 Taxation heads assigned to the states in the Seventh Schedule of the Constitution

Entry in List II of the Seventh Schedule	Head
45	Land revenue, including the assessment and collection of revenue, the maintenance of land records, survey for revenue purposes
46	Taxes on agricultural income
47	Duties in respect of succession of agricultural land
48	Estate duty in respect of agricultural land
49	Taxes on lands and buildings
50	Taxes on mineral rights subject to any limitations imposed by Parliament by law relating to mineral development
51	Duties of excise on the following goods manufactured or produced in the State and countervailing duties at the same or lower rates on similar goods manufactured or produced elsewhere in India: a. alcohol liquors for human consumption b. opium, Indian hemp and other narcotic drugs and narcotics; but not including medicinal and toilet preparations containing alcohol or any substance included in sub-paragraph (b) of this entry
52	Taxes on the entry of goods into a local area for consumption, use or sale therein
53	Taxes on the consumption or sale of electricity
54	Taxes on the sale or purchase of goods other than newspapers, subject to the provisions of entry 92A of List I
55	Taxes on advertisements other than advertisements published in the newspaper and advertisements broadcast by radio or television
56	Taxes on goods and passengers carried by road or on inland waterways

(continued)

Table 8 (continued)

Entry in List II of the Seventh Schedule	Head
57	Taxes on goods and passengers carried by road or on inland waterways. Taxes on vehicles, whether mechanically propelled or not, suitable for use on roads including tramcars, subject to the provision of entry 35 of List III
58	Taxes on animals and boats
59	Tolls
60	Taxes on professions, trades, callings and employments
61	Capitation taxes
62	Taxes on luxuries, including taxes on entertainments, amusements, betting and gambling
63	Rates of stamp duty in respect of documents other than those specified in the provision of List I with regard to rates of stamp duty

Table 9 Panel Granger non-causality test

	Revenue decentralization → Expenditure decentralization	Expenditure decentralization → Revenue decentralization
General category states	$\tilde{z} = 5.78$ (0.00)	$\tilde{z} = -0.00$ (0.00)
Special category states	$\tilde{z} = 3.34$ (0.00)	$\tilde{z} = 0.99$ (0.00)

H_0: X does not Granger − Cause Y
H_A: X does Granger − Cause Y for at least one state
p-values in parenthesis
Lag order selected by minimizing Akaike information criterion

Table 10 Result of Granger causality test (state-wise)

States	Revenue decentralization → Expenditure decentralization		Expenditure decentralization → Revenue decentralization	
	W_i	P_i	W_i	P_i
Andhra Pradesh	18.68	0.00		
Karnataka	7.67	0.01		
Rajasthan	5.14	0.03		
Arunachal Pradesh	11.92	0.00		
Assam	7.44	0.01		
Nagaland	4.52	0.04		
Meghalaya			63.07	0.00

States with only significant values reported
Significance at 5% level

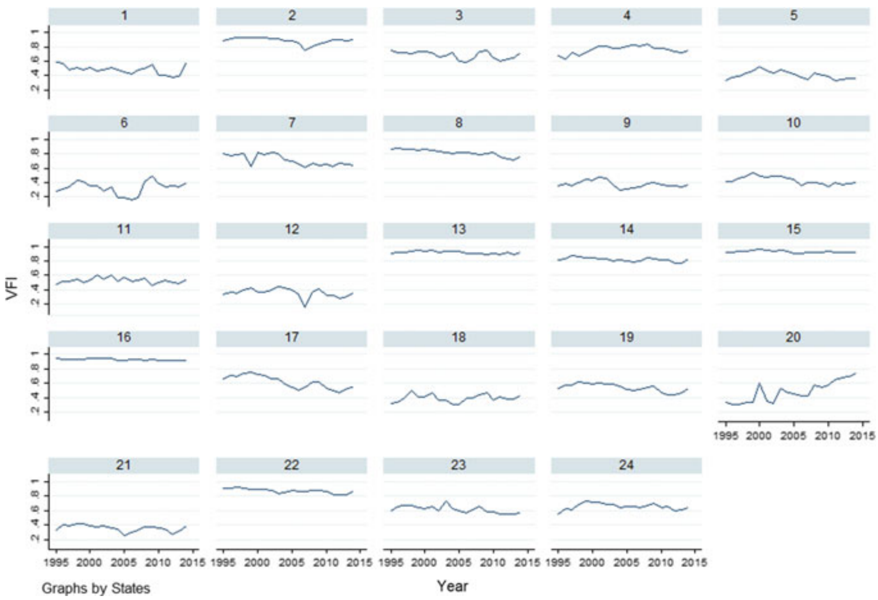


Fig. 7 VFI—Evolution overtime for individual states 1. Andhra Pradesh, 2. Arunachal Pradesh, 3. Assam, 4. Bihar, 5. Gujarat, 6. Haryana, 7. Himachal Pradesh, 8. Jammu and Kashmir, 9. Karnataka, 10. Kerala, 11. Madhya Pradesh, 12. Maharashtra, 13. Manipur, 14. Meghalaya, 15. Mizoram, 16. Nagaland, 17. Orissa, 18. Punjab, 19. Rajasthan, 20. Sikkim, 21. Tamil Nadu, 22. Tripura, 23. Uttar Pradesh, 24. West Bengal

Regression-based Fiscal Capacity Approach of Measuring Horizontal Fiscal Imbalances (HFI):
Under this approach, fiscal/taxable capacity is the predicted tax–NSDP ratio estimated from a regression. The estimation procedure takes into account the states' specific characteristics which include economic, demographic and institutional characteristics.

The basic specification (as given by Le et al. 2008) is

$$Y = f(\text{GDP, Population, Trade, Agr, Corr, Bureau})$$

where

Y	Tax or total fiscal revenue ratio to GDP
GDP	GDP per capita (constant 2000 $US)

Population: Rate of population growth or age dependency ratio as a share of the total population

Trade: Trade Openness

Agr: Agricultural value added

Corr: Corruption Index

Bureau: Bureaucracy quality.

The problem of extending the methodology at state level in India is simply unavailability of the above-mentioned data for all states and over time.

Hence, this paper employs the following specification:

$Y = f$ (Per capita $\text{NSDP}_{\text{current prices}}$).[18]

Predicted fiscal capacity $\left(\widehat{\text{CF}}_{it}\right)$ is then derived by running a panel fixed effect regression on the above-mentioned specification for the two groups of states, general and special category as:

$$\widehat{\text{CF}}_{it} = \hat{\beta}\,\text{NSDP}_{it}$$

Then, HFI is computed as:

$$\text{HFI}_{it} = \widetilde{\text{CF}}_t - \widehat{\text{CF}}_{it}$$

where $\widetilde{\text{CF}}_t$ is median fiscal capacity for each year.

See Table 11; Figs. 8 and 9.

Alternate definitions of VFI and fiscal performances:

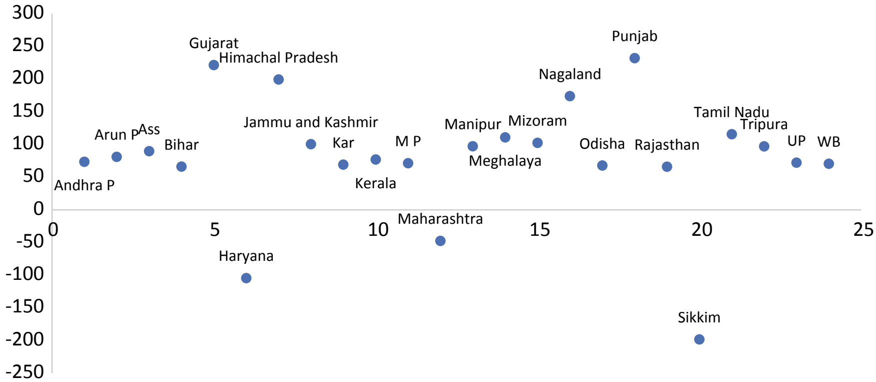
1. **VFI Measure**:

Table 11 Regression result on relationship between VFI and fiscal performance for 23 major Indian states

Dependent variable: Primary deficit–NSDP ratio			
	(1)	(2)	(3)
VFI	0.15*** (8.76)	0.13*** (9.84)	0.09* (2.47)
Expenditure decentralization	−0.26*** (−13.85)	−0.27*** (−16.28)	−0.27*** (−16.41)
Lag debt–NSDP ratio		−0.05*** (−6.36)	−0.05*** (−6.08)
Lag output gap		−0.01 (−0.70)	−0.00 (−0.60)
VFI squared			0.04 (1.02)
Overall R^2	0.17	0.13	0.14
Within R^2	0.57	0.64	0.69
Number of observations	460	460	460
Number of states	23	23	23

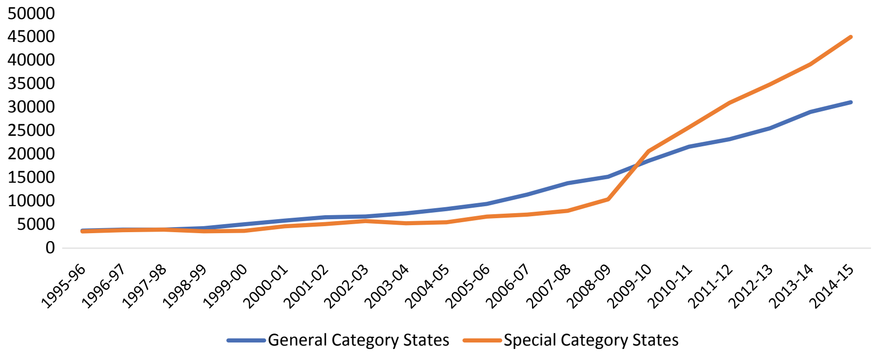
*** (**,*) = Significant at 1(5, 10) per cent level. The robust t-statistic (heteroscedasticity corrected) is represented in the parenthesis; Mizoram dropped due to the unavailability of data

[18] Refer Mukhopadhyay and Das (2003).



Source: Author's computation

Fig. 8 Coefficient of variation of HFI 1/ across states over the period 1995–96 to 2014–15



1/ HFI = Deviation of each states' per capita NSDP from the average of top three NSDP in each year

Source: Author's computation

Fig. 9 Trends in Average HFI 1/

(a) **VFI** = the ability of the state to finance their current expenditure from their own revenue source

$$\text{VFI} = \frac{\text{Own Revenue Source}}{\text{Current Expenditure}}$$

Refer Rao and Singh (2002), Table 2 in Appendix and page no. 9

(b) $$\mathrm{VFI} = \frac{\frac{\text{Own Tax Revenue}_{\text{SNG}} + \text{Non Tax Revenue}_{\text{SNG}}}{\text{Combined Revenues}}}{\frac{\text{Revenue Expenditure}_{\text{SNG}}}{\text{Combined Revenue Expenditure}}}$$

Refer Muddipi (1991).

Given VFI is detrimental to fiscal performance, the coefficients of VFI is expected to be negative.

2. **As indicators of fiscal performance, the following measures are applied**:

(a) Gross fiscal deficit as a proportion of total expenditure of the states.

This measure is used as a component index of fiscal performance index as recommended by the Twelfth Finance Commission (Dholakia 2005).

(b) Ratio of revenue deficit to revenue expenditure of the states

It reveals the proportion of current expenditure to the government financed by the current deficit.

3. **HFI Measures**:

HFI refers to differential fiscal capacities of the sub-national economics. It occurs when sub-national governments have different abilities to raise funds from their tax bases and to provide services. Given non-concurrence on any particular definition of HFI, the following alternate definitions are used.

(a) $$\mathrm{HFI} = \frac{\text{State per capita NSDP}_{\text{Current prices}}}{\text{Average of highest three per capita NSDP}_{\text{current prices}}}$$

Given the definition of HFI used in the main paper, it can be categorized as high or low. Thus, we use dummy variable approach for HFI.

(b) When HFI = Average of highest three per capita $\text{NSDP}_{\text{current prices}}$ − State per capita $\text{NSDP}_{\text{current prices}}$

$$\begin{aligned}\text{HFI Dummy} &= 1,\ \text{when Average of highest three per capita NSDP}_{\text{current prices}} \\ &\quad > \text{State per capita NSDP}_{\text{current prices}} \\ &= 0,\ \text{otherwise}\end{aligned}$$

(c) When HFI = Difference between the fiscal capacity of each state and median fiscal capacity

$$\begin{aligned}\text{HFI Dummy} &= 1,\ \text{when fiscal capacity of each state} > \text{median fiscal capcity} \\ &= 0,\ \text{otherwise}\end{aligned}$$

Recording HFI as a dummy variable makes the regression results more interpretable. It captures the influence of VFI on the fiscal performance of the states, in the presence of differential fiscal position among the states.

Robustness Analysis Results:

See Tables 12, 13, 14, 15, 16, 17, 18 and 19.

Table 12 Regression result on relationship between VFI (alternate measures) and fiscal performance for 14 major general category states

Dependent variable: primary deficit–NSDP ratio VFI measure: states' own revenue/own expenditure				
	(1)	(2)	(3)	(4)
VFI	−0.09*** (−9.15)	−0.11*** (−9.92)	−0.07*** (−4.02)	−0.10*** (−3.14)
Expenditure decentralization	−0.17*** (−9.98)	−0.14*** (−9.03)	−0.18*** (−10.39)	−0.17*** (−7.86)
Lag debt–NSDP ratio		−0.03*** (−3.64)	−0.04*** (−7.81)	−0.04*** (−6.19)
Lag output gap		0.00 (0.24)		
HFI 1/				−0.05 (1.40)
VFI × HFI			−0.03** (−1.99)	0.01 (0.30)
Within R^2	0.63	0.68	0.70	0.69
Overall R^2	0.20	0.52	0.45	0.61
Column (2) and (4) reports random effect results as suggested by Hausman test and Sargan–Hansen test of choosing between FE and RE models				
VFI measure : $\frac{\text{States' Own Tax and Non Tax Revenue/Combined Revenue}}{\text{States' Revenue Expenditure/Combined Expenditure}}$				
	(1)	(2)	(3)	(4)
VFI	−0.01*** (−4.34)	−0.01*** (−4.99)	−0.00 (−0.41)	−0.01 (−1.17)
Expenditure decentralization	−0.21*** (−13.56)	−0.22*** (−14.39)	−0.22*** (−13.92)	−0.23*** (−14.25)

(continued)

Table 12 (continued)

VFI measure : $\frac{\text{States' Own Tax and Non Tax Revenue/Combined Revenue}}{\text{States' Revenue Expenditure/Combined Expenditure}}$				
	(1)	(2)	(3)	(4)
Lag debt–NSDP ratio		−0.05*** (−6.29)	−0.04*** (−6.87)	−0.05*** (−6.96)
Lag output gap		−0.00 (−0.16)		0.13 (1.21)
HFI 1/				−0.03* (−1.67)
VFI × HFI			−0.01 (−1.26)	−0.00 (−0.04)
Within R^2	0.50	0.57	0.57	0.58
Overall R^2	0.13	0.12	0.13	0.17
Number of observations	280	280	280	280
Number of states	14	14	14	14

Fixed effect estimation as suggested by Hausman test and Sargan–Hansen test of choosing between FE and RE Model
*** (**,*) = Significant at 1(5,10) per cent level. Robust t-statistic (heteroscedasticity corrected) represented in parenthesis
1/HFI = as the ratio of state per capita NSDP to average of highest three per capita NSDP

Table 13 Regression result on relationship between VFI (alternate measures) and fiscal performance for 10 major special category states

Dependent variable: primary deficit–NSDP ratio VFI measure: states' own revenue/own expenditure				
	(1)	(2)	(3)	(4)
VFI	−0.25*** (−9.34)	−0.21*** (−5.66)	−0.20*** (−5.48)	−0.24*** (−3.87)
Expenditure decentralization	−0.35*** (−14.43)	−0.34*** (−11.48)	−0.35*** (−15.40)	−0.36*** (−14.34)
Lag debt–NSDP ratio		−0.04*** (−2.90)	−0.04*** (−3.88)	−0.04*** (−3.99)
Lag output gap		0.02 (0.44)		
HFI 1/				−0.02 (−1.05)
VFI × HFI			−0.03 (−0.78)	0.04 (0.21)

(continued)

Table 13 (continued)

Dependent variable: primary deficit–NSDP ratio VFI measure: states' own revenue/own expenditure				
	(1)	(2)	(3)	(4)
Within R^2	0.66	0.62	0.68	0.69
Overall R^2	0.49	0.30	0.34	0.39
VFI measure: $\frac{\text{States' Own Tax and Non Tax Revenue}/\text{Combined Revenue}}{\text{States' Revenue Expenditure}/\text{Combined Expenditure}}$				
	(1)	(2)	(3)	(4)
VFI	−0.06*** (−5.93)	−0.06*** (−4.34)	−0.00 (−0.41)	−0.01 (−1.17)
Expenditure decentralization	−0.36*** (−13.13)	−0.36*** (−11.64)	−0.22*** (−13.92)	−0.23*** (−14.25)
Lag debt–NSDP ratio		−0.05*** (−3.23)	−0.04*** (−6.87)	−0.05*** (−6.96)
Lag output gap		0.04 (0.85)		0.13 (1.21)
HFI 1/				−0.03* (−1.67)
VFI × HFI			−0.01 (−1.26)	−0.00 (−0.04)
Within R^2	0.60	0.60	0.57	0.58
Overall R^2	0.39	0.25	0.13	0.17
Number of observations	200	150	200	200
Number of states	10	10	10	10

Fixed effect estimation as suggested by Hausman test and Sargan–Hansen Test of choosing between FE and RE models

*** (**,*) = Significant at 1(5,10) per cent level. Robust *t*-statistic (heteroscedasticity corrected) represented in parenthesis

Notes Due to the unavailability of data on real NSDP for Mizoram (1995–96 to 1998–99), the sample period is 2001–01 to 2014–15 in col. (2)

1/HFI = as the ratio of state per capita NSDP to average of highest three per capita NSDP

Table 14 Regression result on relationship between VFI 1/ and fiscal performance (alternate measures) for 14 major general category states

Dependent variable: Gross fiscal deficit to total expenditure ratio				
	(1)	(2)	(3)	(4)
VFI	0.78*** (18.13)	0.14*** (12.72)	0.63*** (7.85)	0.67*** (5.93)
Expenditure decentralization	−0.82*** (−20.70)	−0.14*** (−10.48)	−0.83*** (−17.27)	−0.80*** (−12.42)
Lag debt–NSDP ratio		−0.04*** (−4.31)	−0.02 (−0.99)	−0.02 (−0.99)
Lag output gap		0.01 (0.57)		
HFI 2/				−0.07 (−1.17)
VFI × HFI			0.19** (2.11)	0.17 (1.08)
Within R^2	0.84	0.73	0.85	0.84
Overall R^2	0.85	0.61	0.76	0.85
Columns (2) and (4) report random effect results as suggested by Hausman test				
Dependent variable: Revenue deficit/revenue expenditure				
	(1)	(2)	(3)	(4)
VFI	0.96*** (16.54)	1.01*** (14.54)	1.01*** (9.37)	0.72*** (3.91)
Expenditure decentralization	−1.35*** (−21.02)	−1.26*** (−21.63)	−1.25*** (−19.39)	−1.28*** (−18.31)
Lag debt–NSDP ratio		0.10* (1.70)	0.10 (1.55)	0.11* (1.68)
Lag output gap		0.26** (2.40)	0.13** (1.97)	0.27** (2.15)
HFI 2/				−0.24* (−1.68)
VFI × HFI			0.01 (0.10)	0.39 (1.56)
Within R^2	0.78	0.79	0.79	0.80
Overall R^2	0.71	0.78	0.78	0.77
Number of Observations	280	280	280	280
Number of states	14	14	14	14

***(**,*) = Significant at 1(5,10) per cent level. Robust *t*-statistic (heteroscedasticity corrected) represented in parenthesis

Columns (2) and (3) report random effect results as suggested by Hausman test

1/VFI = 1—State Government's Own Revenue/State Government's Own Expenditure

2/HFI = the ratio of state per capita NSDP to average of highest three per capita NSDP

Table 15 Regression result on relationship between VFI 1/ and fiscal performance (alternate measures) for 10 major special category states

Dependent variable: Gross fiscal deficit to total expenditure ratio				
	(1)	(2)	(3)	(4)
VFI	0.80*** (13.46)	0.66*** (9.11)	0.61*** (5.51)	0.32 (1.45)
Expenditure decentralization	−0.78*** (−11.51)	−0.73*** (−15.55)	−0.74*** (−15.56)	−0.72*** (−14.77)
Lag debt–NSDP ratio		−0.05*** (−2.62)	−0.05*** (−2.62)	−0.06*** (−2.95)
Lag output gap		0.14*** (2.96)	0.14*** (2.97)	0.13** (2.22)
HFI 2/				−0.33* (−1.82)
VFI × HFI			0.03 (0.96)	0.43* (1.89)
Within R^2	0.77	0.75	0.75	0.76
Overall R^2	0.82	0.73	0.70	0.70
Column (1) reports random effect results as suggested by Hausman test				
Dependent variable: Revenue deficit/revenue expenditure				
	(1)	(2)	(3)	(4)
VFI	0.87*** (6.14)	0.83*** (5.24)	0.84*** (4.56)	−0.03 (−0.10)
Expenditure decentralization	−1.16*** (−9.17)	−1.12*** (−7.75)	−1.12*** (−7.59)	−1.11*** (−11.25)
Lag debt–NSDP ratio		−0.16** (−2.21)	−0.17** (−2.32)	−0.29*** (−6.28)
Lag output gap		0.19* (1.76)	0.18* (1.73)	0.13 (1.21)
HFI 2/				−0.73** (−2.67)
VFI × HFI			−0.02 (−0.32)	0.89** (2.57)
Within R^2	0.56	0.58	0.58	0.61
Overall R^2	0.63	0.60	0.59	0.48
Number of observations	200	150	150	150
Number of states	10	10	10	10

***(**,*) = Significant at 1(5,10) per cent level. Robust t-statistic (heteroscedasticity corrected) represented in parenthesis
Columns (1), (2) and (3) report random effect results as suggested by Hausman test
1/VFI = 1—State Government's Own Revenue/State Government's Own Expenditure
2/HFI = the ratio of state per capita NSDP to average of highest three per capita NSDP

Table 16 Regression results—sensitivity analysis

Dependent variable: Primary deficit–NSDP ratio					
	(1)	(2)	(3)	(4)	(5)
General category states					
VFI	0.13*** (13.24)	0.13*** (13.17)	0.15*** (15.67)	0.13*** (13.14)	0.14*** (14.21)
Expenditure decentralization	−0.20*** (−12.17)	−0.21*** (−12.43)	−0.20*** (−12.99)	−0.21*** (−12.33)	−0.21*** (−12.82)
Lag debt–NSDP ratio	−0.04*** (−8.53)	−0.04*** (−7.43)	−0.04*** (−7.55)	−0.04*** (−7.98)	−0.04*** (−7.94)
Lag output	0.01*** (7.54)	0.00*** (6.41)	0.00*** (6.49)	0.01*** (7.55)	0.01*** (7.35)
HFI			0.00*** (3.78)		0.00** (2.50)
VFI × HFI		0.00** (2.12)	−0.00*** (−2.88)	0.00 (1.61)	−0.00* (−1.88)
Within R^2	0.78	0.78	0.79	0.78	0.79
Overall R^2	0.44	0.46	0.51	0.45	0.45
Number of observations	280	280	280	280	280
Number of states	14	14	14	14	14

Table 17 Regression results—sensitivity analysis

Dependent variable: Primary deficit–NSDP ratio					
	(1)	(2)	(3)	(4)	(5)
Special category states					
VFI	0.25*** (5.28)	0.28*** (5.79)	0.32*** (6.67)	0.27*** (5.28)	0.28*** (5.40)
Expenditure decentralization	−0.36*** (−11.26)	−0.35*** (10.84)	−0.35*** (−10.49)	−0.36*** (−11.06)	−0.36*** (−11.03)
Lag debt–NSDP ratio	−0.04*** (−2.62)	−0.04*** (−2.77)	−0.04*** (−2.67)	−0.05*** (−2.93)	−0.04*** (−2.91)
Lag output	0.01** (2.20)	0.00 (0.42)	0.00 (0.71)	0.01** (2.40)	0.01** (2.45)
HFI			0.00** (2.13)		0.00 (0.33)
VFI × HFI		0.00 (0.004)	−0.00 (−1.68)	0.00 (1.68)	−0.00 (−0.06)
Within R^2	0.63	0.65	0.66	0.64	0.64
Overall R^2	0.24	0.38	0.36	0.24	0.24
Number of observations	150	150	150	150	150

(continued)

Table 17 (continued)

Dependent variable: Primary deficit–NSDP ratio					
	(1)	(2)	(3)	(4)	(5)
Number of states	10	10	10	10	10

Lag output in logarithms; fixed effect estimation as suggested by Hausman and Sargan–Hansen test; Robust *t* statistics clustered at state level in parenthesis; *** (**,*) = Significance at 1(5, 10) per cent level
HFI indicator used in columns
(2)/(3)—Deviation of each state's per capita NSDP from the average of top three NSDP in each year
(4)/(5)—Deviation of each state's estimated fiscal capacity from median fiscal capacity

Table 18 Regression results with HFI as interaction term

Dependent variable: Primary deficit-to-NSDP ratio						
	(1)	(2)	(3)	(4)	(5)	(6)
General category states						
VFI 1/	0.14*** (7.30)	0.10*** (2.99)	0.13*** (11.65)	0.16*** (10.51)	0.13*** (9.45)	0.13*** (11.79)
Expenditure decentralization	−0.17*** (−10.31)	−0.18*** (−10.62)	−0.14*** (−10.64)	−0.17*** (−10.61)	−0.15*** (−8.88)	−0.17*** (−10.21)
Lag debt–NSDP ratio	−0.05*** (−8.65)	−0.04*** (−8.53)	−0.04*** (−6.45)	−0.04*** (−8.34)	−0.04*** (−7.09)	−0.04*** (−8.30)
HFI		−0.04* (−2.00)		0.02** (2.46)		0.00 (0.46)
VFI × HFI	−0.03 (−1.32)	0.03 (0.69)	0.01 (1.22)	−0.04** (−2.35)	0.01 (1.23)	−0.01 (−0.31)
Within R^2	0.74	0.74	0.72	0.74	0.73	0.74
Overall R^2	0.58	0.62	0.62	0.46	0.61	0.51
Number of observations	280	280	280	280	280	280
Number of states	14	14	14	14	14	14
Column (5) reports random effect results as suggested by Hausman test						
Special category states						
	(1)	(2)	(3)	(4)	(5)	(6)
VFI 1/	0.29*** (4.66)	0.22* (1.90)	0.24*** (5.11)	0.23*** (4.95)	0.24*** (5.31)	0.27*** (5.64)
Expenditure decentralization	−0.34*** (−11.38)	−0.34*** (−11.08)	−0.34*** (−11.60)	−0.34*** (−11.42)	−0.34*** (−11.82)	−0.34*** (−12.12)
Lag debt–NSDP ratio	−0.05*** (−3.74)	−0.05*** (−3.61)	−0.05*** (−3.98)	−0.05*** (−3.96)	−0.05*** (−3.97)	−0.05*** (−3.86)
HFI		−0.08 (−0.91)		−0.01 (−0.34)		0.07 (1.14)
VFI × HFI	−0.03 (−1.66)	0.07 (0.60)	0.00 (0.56)	0.01 (0.44)	−0.03*** (−2.68)	−0.11 (−1.63)

(continued)

Table 18 (continued)

Dependent variable: Primary deficit-to-NSDP ratio						
	(1)	(2)	(3)	(4)	(5)	(6)
Within R^2	0.63	0.63	0.62	0.62	0.63	0.64
Overall R^2	0.35	0.35	0.30	0.30	0.31	0.33
Number of observations	150	150	150	150	150	150
Number of states	10	10	10	10	10	10

*** (**,*) = Significant at 1(5,10) per cent level. Robust *t*-statistic (heteroscedasticity corrected) clustered at state level in parenthesis
1/VFI = 1−States' Own Revenue/States' Own Expenditure
HFI indicator used in column
(1)/(2)—Ratio of each states' per capita NSDP to average of top three NSDP in each year
(3)/(4)—Dummy = 1, if deviation of each states' per capita NSDP from average of top three NSDP in each year is positive otherwise = 0
(5)/(6)—Dummy = 1, if deviation of each states' estimated fiscal capacity from median capacity is positive otherwise = 0

Table 19 System dynamic panel data estimation results 1/

	General category states	Special category states
VFI	0.14*** (9.13)	0.23* (1.76)
Expenditure decentralization	−0.14*** (−9.66)	−0.33** (−2.01)
Lag debt–NSDP ratio	−0.04*** (−3.37)	−0.02 (−0.30)
Lag output gap	0.00 (0.01)	0.03 (−0.39)
Lag of primary deficit	0.16* (1.85)	0.05 (0.18)
Number of observations	266	140
Number of states	14	10
AR(1) 2/	−2.31** [0.02]	−1.23 [0.22]
AR(2) 3/	−0.99 [0.32]	−0.08 [0.94]
Sargan 4/	10.73 [1.00]	8.27 [1.00]

Robust *t*-statistics (heteroscedasticity corrected) clustered at state level in parenthesis (…); *p*-values in parenthesis […]; *** (**,*) = Significance at 1(5, 10) per cent level
1/Arellano–Bover/Blundell–Bond linear dynamic panel data estimation (Arellano and Bond 1991; Arellano and Bover 1995)
2/Residual autocorrelation at lag 1
3/Residual autocorrelation at lag 2
4/Sargan test statistics of over-identifying restrictions

References

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58,* 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68,* 29–51.

Baskaran, T. (2010). On link between Fiscal decentralization and public debt in OECD countries. *Public Choice, 145*(3–4), 351–378.

Bird, R. M., & Tarasov, A. M. (2004). Closing the gap: Fiscal imbalances and intergovernmental transfers in developed federations. *Environment and Policy C: Government and Policy, 22*(1), 77–102.

Bird, R. M., & Vaillancourt, F. (2006). *Perspectives on Fiscal Federalism*. WBI Learning Series 35628, The World Bank, Washington D.C.

Boadway, R., & Tremblay, J. (2006). A theory of Fiscal Imbalance. *FinanzArchiv, 62*(1), 1–27.

Brennan, G., & Buchanan, J. (1980). *The power to tax: Analytical foundations of a fiscal constitution*. Cambridge: Cambridge University Press.

De Mello, L. (2007). *Local government finances: The link between intergovernmental transfers and net worth*. OECD Economics Department Working Paper Number 581.

Dholakia, A. (2005). Measuring Fiscal performance of states an alternative approach. *Economic and Political Weekly, 40*(31), 3421–3428.

Dumitrescu, E. V., & Hurlin, C. (2012). Testing for Granger non-causality in heterogenous panels. *Economic Modelling, 29*(4), 1450–1460.

Eyraud, L., & Lusiyan, L. (2013). Vertical fiscal imbalances and fiscal performance in advanced economies. *Journal of Monetary Economics, 60,* 571–587.

Goodspeed, T. (2002). Bailouts in federations. *International Tax and Public Finance, 9*(4), 409–421.

Hagen, J., & Eichengreen, B. (1996). Federalism, Fiscal Restraints, and European Monetary Union. *The American Economic Review, 86*(2), 134–138.

Herperger, D. (1991). *Distribution of powers and functions in federal systems*. Ottawa: Ministry of Supply and Services.

Jin, J., & Zou, H. (2002). How does fiscal decentralization affect aggregate, national, and subnational government size? *Journal of Urban Economics, 52,* 270–293.

Karpowicz, I. (2012). *Narrowing vertical Fiscal imbalance in Four European Countries*. IMF Working Paper WP/12/91. International Monetary Fund.

Le, T. M., Dodson, B. M., & Rojchaichaninthorn, J. (2008). *Expanding taxable capacity and reaching revenue potential: Cross-country analysis*. Policy Research Working Paper 4559. The World Bank, Washington, D.C.

Ma, J. (1997). *Intergovernmental Fiscal Transfer: A comparison of Nine Countries (Cases of the United States, Canada, the United Kingdom, Australia, Germany, Japan, Korea, India, and Indonesia)*. Prepared for Macroeconomic Management and Policy Division, Economic Development Institute. The World Bank.

Mohan, R., & Shyjan, D. (2009). *Tax devolution and grant distribution to States in India. Analysis and road map for alternatives*. Working Paper 419: www.sds.edu.

Muddipi, S. (1991). *Finance commissions and centre state financial relations*. Allahabad: Chugh-Publications.

Mukhopadhyay, H., & Das, K. K. (2003). Horizontal imbalances in India—Issues and determinants. *Economic and Political Weekly, 38*(14), 1416–1420.

Oates, W. (1972). *Fiscal federalism*. New York: Harcourt-Brace. Chapter five.

Oates, W. (1999). An easy on Fiscal federalism. *Journal of Economic Literature, 37*(3), 1120–1149.

Plekhanov, A., & Singh, R. (2007). How should government borrowing be regulated? Some cross-country empirical evidence. *IMF Staff Papers, 53*(3), 426–452.

Rangarajan, C., & Srivastava, D. K. (2008). *Reforming India's Fiscal transfer system: Resolving vertical and horizontal imbalances*. Madras School of Economics Working Paper 31/2008.

Rao, M. G., & Singh, N. (2002). The political economy of centre-state fiscal transfers in India. In J. McLaren (Ed.), *Institutional elements of tax design and reform* (pp. 69–123).Washington DC: World Bank.

Rao, C. B., & Srivastava, D. K. (2014). *Dependence of states on central transfers: State-wise analysis*. NIPFP Working Paper No. 2014-137. National Institute of Public Finance and Policy, New Delhi.

Rao, M. G. (2000). *Fiscal decentralization in Indian federalism*. Retrieved from https://www.imf.org/external/pubs/ft/seminar/2000/fiscal/rao.pdf.

Rodden, J. (2002). The dilemma of Fiscal Federalism: Grants and Fiscal performance around the world. *American Journal of Political Science, 46*(3), 670–687.

Rodden, J. (2003). Reviving leviathan: Fiscal Federalism and the growth of government. *International Organization, 57*(4), 695–729.

Rodden, J., & Wibbels, E. (2009). Fiscal decentralization and the business cycle: An empirical study of seven federations. *Economics and Policies, 22*(1), 37–67.

Sharma, C. K. (2012). Beyond gaps and imbalances: Restructuring the debate on intergovernmental fiscal relations. *Public Administration—An International Quarterly, 90*(1), 99–128.

Singh, N. (2007). *Fiscal; federalism and decentralization in India*. Paper Prepared for a World Bank project.

Thirteenth Finance Commission Report. (2009).

Velasco, A. (2000). Debts and deficits with fragmented fiscal policy making. *Journal of Public Economics, 76,* 105–125.

Data Source

Reserve Bank of India (RBI). *Handbook of statistics on state finances*. Mumbai: RBI.

Reserve Bank of India (RBI). *Handbook of statistics on Indian economy*. Mumbai: RBI.

Reserve Bank of India (RBI). *Handbook of statistics on Indian States*. Mumbai: RBI.

Reserve Bank of India (RBI). *State finances: A study of budgets*. Mumbai: RBI.

Exchange-Rate Pass-Through to Sectoral Import Prices in an Emerging Market Economy: *An Indian Evidence*

Pradyut Kumar Pyne and Saikat Sinha Roy

Abstract This paper investigates the degree of exchange-rate pass-through (ERPT) to prices of non-oil imports in India during reforms. The empirical literature provides ample evidence on incomplete pass-through to import prices across developed and developing economies. Theoretically, incomplete pass-through to import prices is explained in terms of exporters adjusting their markups in order to maintain market shares following currency depreciation. Even though the studies on India have sound theoretical basis, the empirical estimates are weak based on single equation models. In sharp contrast to earlier attempts, a simultaneous equation model incorporating both demand and supply sides is set up for estimation in this paper. The reduced form

This paper is based on the first author's unpublished M. Phil. dissertation submitted to Jadavpur University, which was adjudicated by Professor Sarmila Banerjee. She also adjudicated the Ph.D. thesis of the first author. Professor Banerjee has given the first lessons in Economics to the second author and has advised both the authors all through their respective careers. Both the authors express their immense gratitude to Professor Sarmila Banerjee by re-writing this paper based on the comments received from her on the M. Phil. dissertation.
The initial research was supported by the Centre for Advanced Studies, Department of Economics, Jadavpur University, Kolkata. The second author wishes to thanks the Centre for the generous research grant. An earlier version of this paper was presented at the International Conference on *Empirical Issues in International Trade and Finance* organized by Indian Institute of Foreign Trade, Kolkata during 23–24 December 2008, *45th Annual Conference of the Indian Econometric Society* held at Guwahati, India during 7–9 January 2009, 11th Annual Conference on Money and Finance organized by Indira Gandhi Institute of Development Research, Mumbai during 23–23 January 2009 and the *Dynamics, Economic Growth and International Trade (DEGIT) XIV Conference* held at University of California, Los Angeles, USA during 5–6 June 2009. The authors are indebted to K. L. Krishna, Ajitava Raychaudhuri, Soumyen Sikdar, Ashima Goyal, Ganti Subramanyam, David Mayer-Folkes, Bjarne S. Jensen, Areendam Chanda, Sanmitra Ghosh, Shantanu Bagchi, participants at these conferences, and an anonymous referee for their comments on an earlier draft. Help from Darpajit Sengupta with recent literature is gratefully acknowledged. However, responsibility for errors, if any, lies with the authors.

P. K. Pyne
Department of Economic Affairs, Ministry of Finance, New Delhi 110001, India
e-mail: pradyut.pyne@gmail.com

S. Sinha Roy (✉)
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: saikat.sinharoy@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_13

equation for import prices is estimated for disaggregated imports using pooled dataset covering eight major product groups during 1993–94 to 2017–18. The empirical results, which are robust, show incomplete exchange-rate pass-through to import prices with the degree of pass-through varying across import product groups. The results have significant implications for policy on trade and exchange rate in any emerging market economy.

Keywords Exchange rate · Import prices · Incomplete pass-through · Market structure · Simultaneous equation model

JEL Classification C33 · F14 · F33 · F41 · L16

1 Introduction

This paper investigates the degree of exchange-rate pass-through (ERPT) to prices of aggregate as well as disaggregated imports in India during reforms. Theoretically, devaluation of home currency leads to an increase in import prices in terms of home currency. Following the decline in home country currency value and the increase in home currency denominated price of imports, the domestic demand for imports declines. In order to maintain the existing market shares, especially in the short run, exporters adjust their markups. The higher the markup over production cost, the larger can be the room for adjustment in prices to changes in currency value. This explains why exchange rate movements often do not bring about proportional change in import prices and pass-through is incomplete. Empirically, the pass-through coefficient measures the degree of responsiveness of trade prices to unit changes in exchange rate. ERPT is complete if there is a proportional change in prices on account of a unit change in exchange rate and is incomplete if the change in price is less than proportional.

Goldstein and Khan (1985), while reviewing some of the earlier studies on the subject, find overwhelming evidence of incomplete exchange-rate pass-through to import prices.[1] Incomplete pass-through is found to be essentially a short-run phenomenon.[2] In addition, a number of studies have shown that the degree of pass-through has declined substantially since the 1990s.[3] However, most of these studies are with regard to developed country imports. With increasing global integration of developing economies, the responsiveness of prices of internationally traded goods

[1]A detailed review of the literature is available in Pyne (2008). Also see, for instance, Krugman (1987), Feenstra (1989), Kreinin et al. (1987), Hooper and Mann (1989), Feenstra et al. (1993), Goldberg and Knetter (1997), Campa and Goldberg (2002), Pollard and Coughlin (2004), Campa et al. (2005), Campa and Minguez (2006), among others.

[2]For instance, Faruqee (2004), Campa et al. (2005), Halpern and Koren (2007) among others, show that pass-through is complete in the long run. To factor in time-varying effects, Kim (1990) uses a varying parameter model to estimate ERPT. These contrast the findings of Ghosh and Rajan (2009).

[3]For instance, Marazzi et al. (2005) found a steady decline in the ERPT into US import prices during the 1990s.

and services to exchange rate movements has, of late, assumed importance in these countries. Recent studies highlight ERPT in emerging markets after the financial crises. For instance, Jašová et al. (2016) find exchange-rate pass-through in emerging economies decreased after the financial crisis, while exchange-rate pass-through in advanced economies has remained relatively low and stable over time. Moreover, it is shown that the declining pass-through in emerging markets is related to declining inflation both in the short and the long run.

Incomplete pass-through is generally explained in the existing literature by the existence of trade barriers, transaction and transportation costs, market power and imperfect substitutability between domestic substitute and foreign products. Market segmentation is one among the possible reasons for incomplete pass-through, which allows imperfectly competitive firms to price differently to different export market. To find out why exchange-rate pass-through is incomplete, Dornbusch (1987) explains the adjustment of relative prices to exchange rate movements in an industrial organization approach by using various models. The approach is to explain price adjustment in terms of market concentration, product homogeneity and substitutability, and relative market shares of domestic and foreign firms. All the models in this analysis predict that appreciation leads to a decline in the price of imports and vice versa. While, in the case of homogenous goods, domestic firms fully match the decline in price, the extent of decline in the relative price of differentiated imported brands depends on competition and on the relative number of home to foreign firms. For instance, as Halpern and Koren (2007) find import prices are higher for products of firms with greater market power. In contrast, Feenstra et al. (1993) find that pass-through tends to be highest for market with firms facing less competition and not experiencing a similar change in costs. However, Krugman (1987) explains incomplete pass-through in terms of supply dynamics resulting from the costs of rapidly adjusting the marketing and distribution infrastructure, and the demand dynamics resulting from the need of firms to invest in reputation.

Choice of currency in invoicing prices of commodities also plays a major role in exchange-rate pass-through.[4] Firms generally set the prices by examining the nature of fluctuations of prices in both home and destination market currencies. If the variance of export prices in firm's own currency is comparatively less than in local currency, then the firm opts for producer currency pricing and vice versa. Campa and Goldberg (2002) explain incomplete pass-through in terms of vital but limited role of macroeconomic variables among industries, while Campa and Minguez (2006) and Zorzi et al. (2007) explain in terms of the degree of openness. Lopez-Villavicencio and Mignon (2017) do not find convincing evidence that global factors cause a structural change in the degree of exchange-rate pass-through. Trade openness is found to increase ERPT in some sectors, though not universal. However, regionalization is found to reduce the pass-through in a more generalized way. Further, inflation is found to play an important role in determining the nature of pass-through (see, Taylor 2000; Choudhuri and Hakura 2001; Zorzi et al. 2007).

[4]Goldberg and Knetter (1997), Bacchetta and van Wincoop (2002), Engel (2005), and Campa et al. (2005) emphasize the importance of currency invoicing of the prices of traded commodities.

For India, there are some recent studies as well looking into exchange-rate pass-through to general price level and inflation, with evidence being divided on the extent and trend in ERPT during post-reforms.[5] Dholakia and Saradhi (2000) and Dash and Narasimhan (2011) show pass-through to import prices near complete or more than complete even in the short run. In contrast, using the profit maximization framework, Mallick and Marques (2006) show incomplete pass-through during 1980–2001, with pass-through coefficient across sectors varying between the two decades. Except for some including crude, food and oils imports, ERPT is found to be incomplete. Mallick and Marques (2008) find pass-through of changes in exchange rate and tariff rate to vary across products. ERPT into import prices is significant in 12 industries, whereas tariff pass-through is significant only in six industries, with full pass-through. However, ERPT is incomplete only in four industries, TRPT being incomplete in 36 industries, implying firms exporting to India maintaining their market share with regard to tariff changes as against exchange rate changes.

The above review shows that the empirical literature on the subject for emerging economies, in particular India, is rare. Studies on exchange-rate pass-through to import prices are even fewer. In the context of trade and exchange rate reforms in India and depreciation/appreciation of the Indian currency during the period that followed, a study of exchange-rate pass-through to sectoral import prices becomes necessary. Even though the studies by Mallick and Marques (2006, 2008) are exhaustive, they are not based on a macroeconomic framework. In that sense, the results thus arrived at in the earlier study might not be appropriate. The present study derives its justification from such deficiencies in the existing literature. The rationale for another study on ERPT to import prices is further based on the link between changes in import prices and inflation in an open economy. With evidence on increasing import dependence to maintain international competitiveness by Indian firms,[6] international materials and energy prices driving inflation in India,[7] inflationary trends in the recent past and RBI's stated policy of inflation targeting through monetary management since 2016, a study on exchange rate through to domestic prices, in general, and import prices, in particular, becomes necessary. As Bacchetta and van Wincoop (2002) show higher pass-through to import prices than to consumer prices, it is more important to understand the extent of pass-through to import prices over domestic prices.

The main objective of this paper is to estimate and analyse the responsiveness of India's import prices to exchange rate movements. The exercise is also to find out the differences in the degree of the exchange-rate pass-through at the aggregate level as well as across some selected non-oil sectors. In specific, the present study analyses ERPT to import prices of disaggregated product groups including chemicals, machinery and transport equipment, crude materials, manufactures classified by materials and food products. Following Mallick and Marques (2006) finding that exchange-rate pass-through to be in the opposite direction for fuel and food product

[5]See, for instance, Ghosh and Ranjan (2007), Khundrakpam (2008) and Patra et al. (2018).

[6]See Ghosh and Sinha Roy (2018).

[7]See Sharma et al. (2011). Balakrishnan and Parameswaran (2019) using structuralist models show that oil prices is one among the many factors driving inflation in India.

imports, the latter is included in this analysis as the sectors are undergoing wide-ranging reforms. However, pass-through to fuel prices is kept outside the purview of this analysis.

The paper is organized as follows. Apart from this section including a review of existing literature, Sect. 2 delineates some stylized facts on exchange-rate pass-through in the context of trade and exchange rate reforms in India. The following two sections provide with empirical model specification and data used, and estimation results on exchange-rate pass-through. Section 5 concludes summarizing the major findings of the study along with their implications for policy.

2 Some Snapshots on India's Trade and Exchange Rate Reforms

Trade reforms, though in a piecemeal way, were initiated during the 1980s, whereas exchange rate reforms towards a market-determined exchange rate of the rupee started in the early 1990s.[8] This is, however, not to say that there were no exchange rate reforms prior to the 1990s. The major trade policy changes in the post-1991 period included removal of quantitative restrictions and replacement of such restrictions by tariffs, substantial reduction in the tariff rates as also their dispersion and simplification of procedures relating to trade. The WTO furthered this process of trade liberalization in India. As a result, the level of protection to Indian industry declined significantly with a decline in average import tariff across various product groups between 1996 and 2017 (see Table 1). The coefficient of variation for total duty rates, however, varied since then, despite significant rationalization of import duty structures. A further step towards rationalizing the incentive structure for a neutral trade regime involved the removal or restructuring of a large number of tax benefits and exemptions. Besides these, the policies also aimed to reduce transactions cost in trade through a number of measures bring about procedural simplifications. Trade policy reforms in India, on the whole, aimed at reducing distortions in relative prices, removing anti-export bias, improving competition thereby enhancing international competitiveness.

There was a spurt in the growth of trade in the post-reforms period. From the mid-1980s, India's merchandise exports grew at a high rate.[9] The annual average growth rate of exports during 1990s in US Dollar term was about 8.6%, while imports experienced a slightly higher average growth rate of 9.6%. After 2001–02, as evident from Fig. 1, import growth since 2001–02 was high, and it was higher than export growth during the period. The growth rate of merchandise imports peaked at 42.7% in 2004–05. The share of India's imports in global trade increased from 0.9% in 1984–85 to 1.5% in 2005–06. The high growth phase of India's merchandise trade lasts till

[8] See Panagariya (2005), Sinha Roy (2005), Sengupta and Sinha Roy (2018) for the purpose.

[9] Sinha Roy (2005) provides similar evidence. However, the study does not attribute growth in post-reform exports to trade liberalization alone.

Table 1 Average import duty rates in India

	1996	2000	2005	2010	2017
MFN applied rate (simple average)					
Food and live animals	30.6	37.8	34.3	31.8	32.6
Beverages and tobacco	159.1	101.8	88.0	93.6	91.5
Crude materials, inedible, except fuels	25.1	24.3	16.9	11.0	12.1
Machinery and transport equipment	34.9	27.9	15.1	8.2	9.8
Manufactured goods (classified by material)	41.0	33.2	15.6	8.3	9.4
Miscellaneous manufactured articles	45.2	34.1	14.5	9.2	9.4
Effective applied rate (simple average)					
Food and live animals	33.0	36.0	36.3	29.0	30.1
Beverages and tobacco	198.7	114.8	102.9	106.3	103.7
Crude materials, inedible, except fuels	24.5	23.8	15.1	6.9	7.3
Machinery and transport equipment	34.0	27.4	13.8	6.3	7.0
Manufactured goods (classified by material)	38.0	34.2	15.3	7.3	7.7
Miscellaneous manufactured articles	38.6	31.7	13.9	7.6	7.1

Source TRAINS, WITS, World Bank

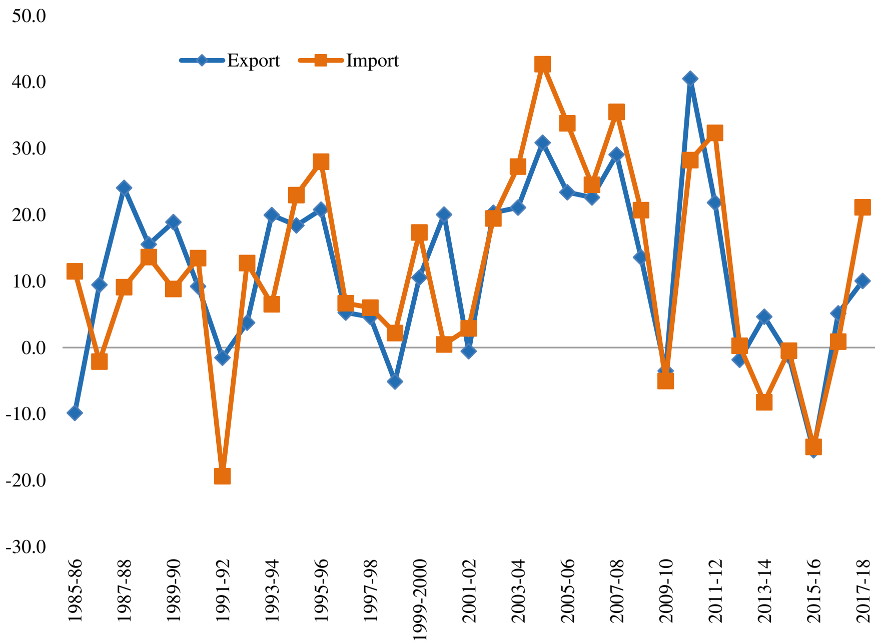


Fig. 1 Post-reforms growth of India's merchandise exports and imports (in %)

2008–09, from when the global economy slowed down with a sharp downturn in 2009–10. While slack in global growth moderated merchandise export growth after 2011–12 (Fig. 2), for import growth it could be crude oil prices, which declined from US$112 per barrel in 2011–12 to US$46 per barrel in 2015–16 (Fig. 3).

Accompanying high growth, India's merchandise imports underwent changes in direction and composition. During the reforms period, India's imports are being sourced from a wider range of countries. Data from RBI sources show that the shares

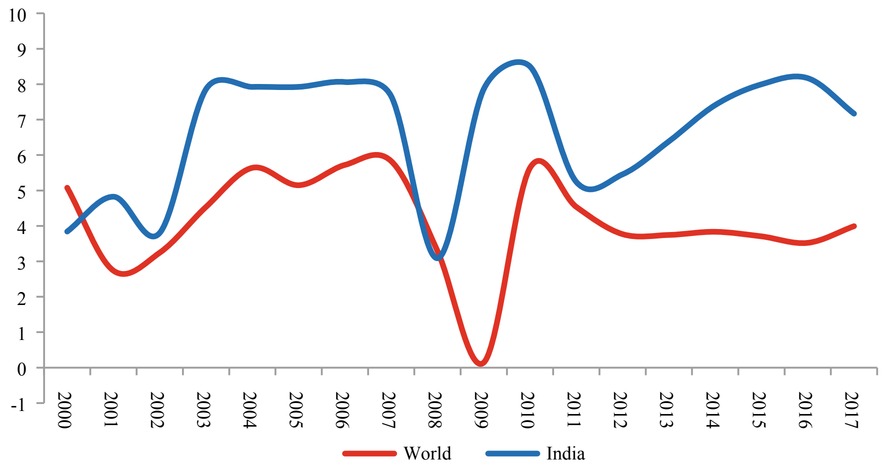


Fig. 2 Real growth of GDP (%). India's growth rates for the financial years are as per Central Statistics Organisation database. Growth rates from 2005–06 are at 2011–12 base year, while for 2000–01 to 2004–05 are at 2004–05 base. World growth rates are based on data from World Economic Outlook database of IMF

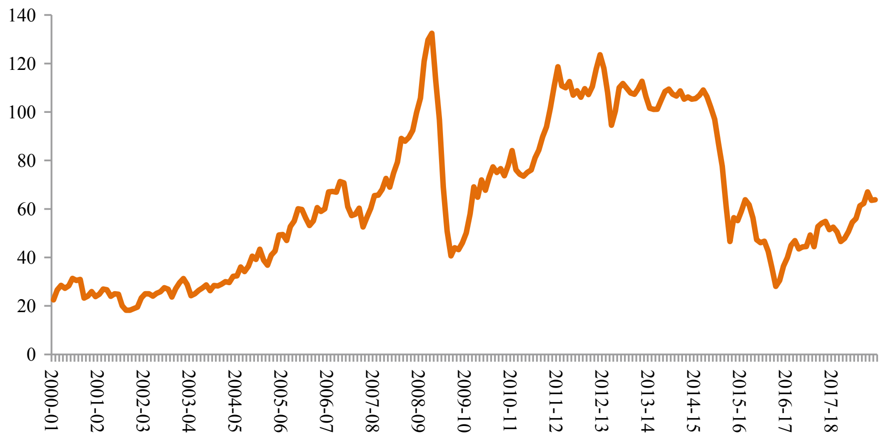


Fig. 3 Crude oil price (Indian basket) in US$ per barrel

of India's traditional trading partners like Germany, Japan, UK and Australia have declined along with the emergence of new import partners from East Asia including China. With regard to commodity composition of imports, the share of electronic items in total imports recorded high increase since 2011–12. Apart from agriculture and allied products imports, transport equipment, ores & minerals, and chemical and related products also increased their respective shares in total imports as against that for gems and jewellery and crude petroleum and products.

2.1 Exchange Rate: Policy Changes and Movements

In 1991, following India's balance of payment crisis, a series of currency reform measures was undertaken in addition to other structural reform measures. A transition to a market-determined exchange rate regime was felt necessary to effectively initiate reform measures in order to change the incentive structure of different sectors of the economy. Reserve Bank of India (RBI) undertook a downward adjustment of the exchange rate of about 18% in two stages on 1 July 1991 and 3 July 1991. This sets the stage for more fundamental changes in the exchange rate regime as recommended by the High Level Committee on Balance of Payments. Such downward adjustments in the currency were followed by the introduction of the Liberalized Exchange Rate Management System (LERMS) in March 1992 and the adoption of, for the first time, a dual exchange rate system in India. This dual exchange rate system implied a depreciation of about 11% for transactions routed through market. However, the dual exchange rate imposed an implicit tax on export proceeds and remittances, which hindered supplies of foreign exchange in the foreign exchange market. Consequently, downward pressures had been building up and a regime change was overdue. Subsequently, in March 1993, the dual exchange rate system gave way to a unified exchange rate system, which along with removal of exchange restriction on imports through the abolition of foreign exchange budgeting are the initial steps towards current account convertibility. The final step towards current account convertibility was taken in August 1994 by further liberalization of invisible transactions and exchange control regulations up to a specified limit.

Since then, the exchange value of the Indian rupee is determined through demand for and supply of foreign exchange in the market and is not administered. However, the Reserve Bank of India continued to actively intervene trading in the foreign exchange market with an aim of 'containing volatility' and thus influencing the currency value. The Indian rupee exhibited a reasonable stability up to mid-1997, when the currency experienced a mild attack of contagion emanating from East Asian currency crisis. During 1998, exchange rate management of the RBI continued its focus on smoothing excessive volatility in the exchange rate and maintaining orderly market condition. The depreciating trend of Rupee–US Dollar exchange rate till early 2000 got reversed in 2003–04 (Fig. 4). However, after sharp appreciation in 2007–08, Rupee–US Dollar exchange rate started depreciating with increased

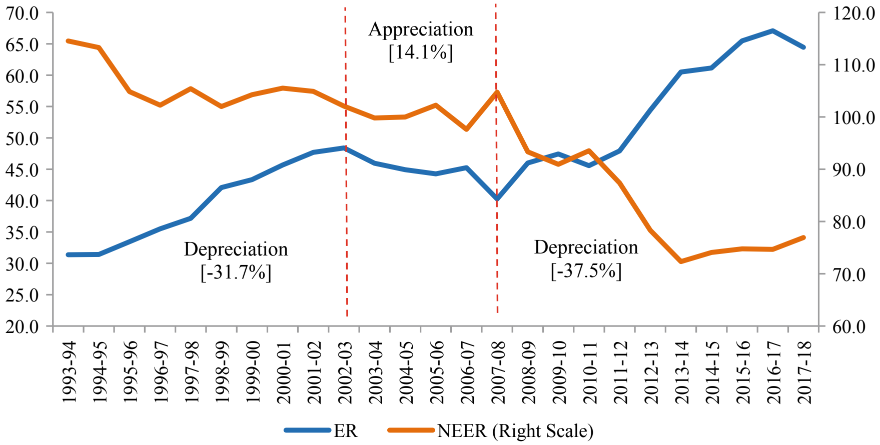


Fig. 4 NEER and Rupee–US dollar annual exchange rate

volatility.[10] Nonetheless, following depreciation in the rupee currency after the mid-1980s, both NEER and REER depreciated from mid-1980s to early 1990s. However, the nominal effective exchange rate appreciated during 1993–94 and thereafter again in 1997–98 (see Table 2). Both NEER and REER moved synchronously with NEER and REER depreciating, on the average, during the entire period.

Table 2 Depreciation (−ve)/appreciation (+ve) in India's REER and NEER

Year	REER	NEER	Year	REER	NEER
1994–95	4.3	−1.1	2006–07	−1.6	−4.5
1995–96	−5.9	−7.5	2007–08	8.4	7.3
1996–97	−1.4	−2.5	2008–09	−8.7	−10.9
1997–98	4.1	3.1	2009–10	4.2	−2.6
1998–99	−7.7	−3.2	2010–11	8.5	2.9
1999–00	3.2	2.2	2011–12	−2.1	−6.6
2000–01	4.3	1.2	2012–13	−4.3	−10.4
2001–02	0.8	−0.6	2013–14	−2.2	−7.7
2002–03	−2.7	−2.7	2014–15	5.5	2.4
2003–04	1.4	−2.2	2015–16	2.9	0.9
2004–05	0.5	0.2	2016–17	2.2	−0.1
2005–06	2.4	2.2	2017–18	4.5	3.1

Source Reserve Bank of India database

[10]Sengupta and Sinha Roy (2018) provide a brief account of RBI's attempts at maintaining the rupee exchange value.

2.2 Movements in India's Import Prices

In general, the 1990s is described as the era of price stability all over the world. The opening up of Indian economy during the 1990s, along with high depreciation and fluctuations in the global economy, increased macroeconomic uncertainties and added pressure on prices to rise. During this period, import prices of all product groups show increasing tendencies (Fig. 5), which became more prominent after 2008–09. Unit value index for 'machinery and transport equipment' was driving the general unit value index to increase sharply, especially after 2008–09, despite a significant reduction in import duty (see Table 1). Unit values for 'food articles' also increased sharply. For all other non-oil imports, unit value index shows an upward trend. Further, as Fig. 6 shows, the two series have moved together, but in opposite directions, in most years during 1993–94 to 2017–18. This evidence on synchronous turning points for NEER and import prices calls for a detailed econometric analysis.

3 Empirical Estimation

This section provides an econometric estimate of exchange-rate pass-through to import prices in India. This estimation assumes importance in the context of the observation made in the earlier section that India's import prices moved together with exchange rate. To recapitulate, import prices in domestic currency increased during years of depreciation, while, in years of appreciation, import prices fell. It can thus be hypothesized that changes in exchange rate are passed through to Indian

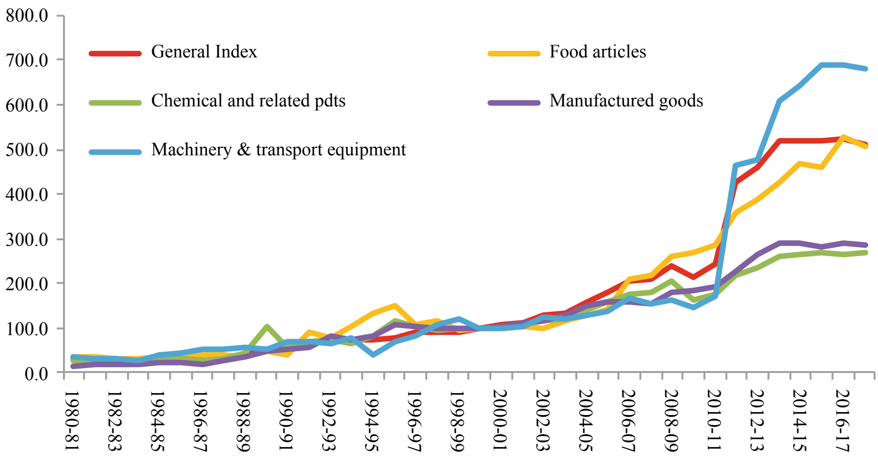


Fig. 5 Movements of unit values of Indian imports

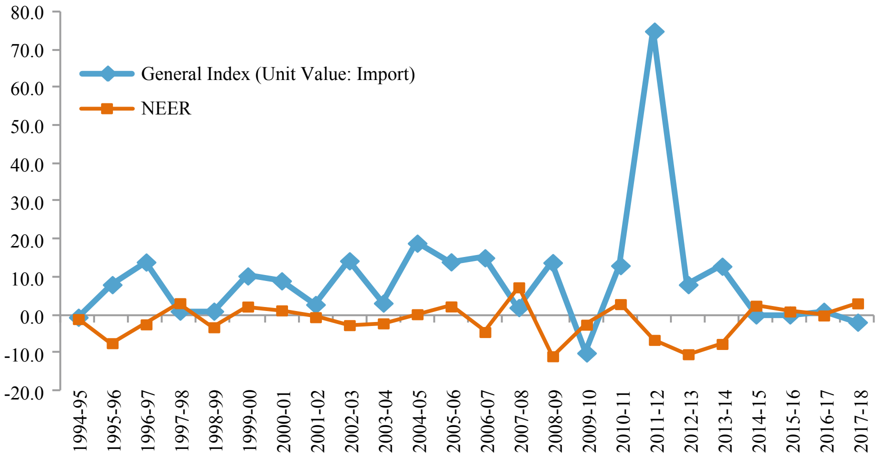


Fig. 6 Growth of import price and NEER

import prices, but the extent to which the import prices have responded to such changes need to be estimated.

To estimate the exchange-rate pass-through to prices in general and to trade prices in particular, economists use various models and methods. Earlier studies, in general, applied a single equation to estimate exchange-rate pass-through to trade prices for various economies. Most of these studies, as reviewed in Goldstein and Khan (1985) and Goldberg and Knetter (1997), provide evidence on exchange-rate pass-through either directly through bivariate estimation results or indirectly in terms of supply price elasticity estimates. As argued earlier in the paper, there is a need for an alternate empirical model based on a macroeconomic framework, which is detailed out hereafter.

3.1 The Empirical Model

In this study, the econometric model is based on a simultaneous equation imperfect substitute model as in Goldstein and Khan (1985). In the model, the quantity of import demand depends on the price of imports in domestic currency and price of domestic substitute commodities. The domestic demand for imports can also be an important factor influencing import demand. Another influencing factor is the trade openness of the economy. The import demand function can be specified as:

$$M_{\mathrm{d}} = f(P_{\mathrm{s}}/P_{\mathrm{m}}, Y) \tag{1}$$

where M_{d} is the quantity demanded for imports, P_{m} is the price of imports in domestic currency, P_{s} is the price of domestic substitute commodities in terms of domestic currency and Y is the gross domestic product of the economy. Equation (1) can also be written in log-linear form as:

$$\ln M_{\mathrm{d}} = \alpha_0 + \alpha_1 \ln P_{\mathrm{m}} + \alpha_2 \ln P_{\mathrm{s}} + \alpha_3 \ln Y + \varepsilon_1 \tag{2}$$

with $\alpha_1 < 0$ and $\alpha_2, \alpha_3 > 0$.

On the other hand, even if Indian imports account for a very small share of the world trade, the export supply by the foreign exporters to India cannot be assumed to be infinitely elastic. Assuming law of one price not to hold, export supply depends on the price of exports in exporters' currency as well as exporters' domestic price. The export supply equation can be specified as:

$$X_{\mathrm{s}}^* = g\left(P_{\mathrm{x}}^*, P^*\right) \tag{3}$$

where X_{s}^* is the quantity of export supply by the foreign exporters, P_{x}^* is the price of the export items in foreign currency and P^* is exporters' domestic price. Producers' currency pricing is assumed here as a most common case in international trade practices. This is not necessary due to stability in producers' currency, but it may be optimal to do so. Moreover, India is not a major importer of most of these products.

Given exchange rate e defined in terms of rupee price of the foreign currency, price of imports in domestic currency $P_{\mathrm{m}} = P_{\mathrm{x}}^* \cdot e$. This is valid in case of free trade and assuming zero transport cost. And in case of protected trade, if there is an ad valorem tariff t on unit value of imports, the price of imports at the domestic market will be $P_{\mathrm{m}} = P_{\mathrm{x}}^* \cdot e \cdot (1 + t)$.

Therefore

$$P_{\mathrm{x}}^* = \frac{P_{\mathrm{m}}}{e(1 + t)} \tag{4}$$

In log-linear form, the export supply equation can be written as:

$$\ln X_{\mathrm{s}}^* = \beta_0 + \beta_1 \ln P_x^* + \beta_2 \ln P^* + \varepsilon_2$$

with $\beta_1 > 0$ and $\beta_2 < 0$. The export supply equation can also be written in an alternative form as

$$\ln X_{\mathrm{s}}^* = \beta_0 + \theta_1 \ln P_{\mathrm{m}} + \theta_2 \ln e + \theta_3 \ln(1 + t) + \beta_2 \ln P^* + \varepsilon_2 \tag{5}$$

where θ_1, θ_2 and θ_3 are different components of β_1 and in Eq. (6), $\theta_1 > 0$, $\theta_2 < 0$ and $\theta_3 < 0$.

In equilibrium, demand for imports by the importing nation and supply of exports by the exporting nation matches, i.e.

$$M_{\mathrm{d}} = X_{\mathrm{x}}^* = M(\text{say}) \tag{6}$$

From the equilibrium condition, the following reduced form equation can be derived.

$$\ln P_{\mathrm{m}} = \delta_0 + \delta_1 \ln e + \delta_2 \ln(1+t) + \delta_3 \ln P^* + \delta_4 \ln P_s + \delta_5 \ln Y + \mu \quad (7)$$

where $\delta_0 = \left(\frac{\alpha_0-\beta_0}{\beta_1-\alpha_1}\right)$, $\delta_1 = -\left(\frac{\theta_2}{\beta_1-\alpha_1}\right)$, $\delta_2 = -\left(\frac{\theta_3}{\beta_1-\alpha_1}\right)$, $\delta_3 = -\left(\frac{\beta_2}{\beta_1-\alpha_1}\right)$, $\delta_4 = \left(\frac{\alpha_2}{\beta_1-\alpha_1}\right)$, $\delta_5 = \left(\frac{\alpha_3}{\beta_1-\alpha_1}\right)$ and $\mu = \left(\frac{\varepsilon_1-\varepsilon_2}{\beta_1-\alpha_1}\right)$.

The coefficient δ_1 provides us the degree of pass-through to import prices expressed in domestic currency. The sign of this coefficient is expected to be positive. This can also be referred to as the elasticity of domestic currency import prices with respect to exchange rate. Instead of the equilibrium model, a disequilibrium model could have been estimated where import prices respond to excess demand.

For estimation purpose, Eq. (8) can be rewritten for ith sector in tth time period as:

$$\ln P_{mit} = \delta_0 + \delta_1 \ln e_t + \delta_2 \ln(1+t) + \delta_3 \ln P^*_{it} + \delta_4 \ln P_{sit} + \delta_5 \ln Y_{it} + \mu \quad (8)$$

with $\delta_1, \delta_2, \delta_3, \delta_4 > 0$. The parameter of interest is δ_1.

3.2 The Data and Method

For the purpose of estimation, the data requirements are immense. There is a need for appropriate data, especially with regard to data on import and other prices. The choice of sectors, as has been stated earlier, is restricted by data availability and comparability over time. The analysis covers such non-oil imports into India including chemicals, machinery, transport equipment, manufactures classified by materials, and food products apart from crude materials, which together account for about 70% of total imports.[11] The study uses annual data for the time period 1993–94 to 2017–18. Such data limitations restrict the use of any time series method for estimation purposes.

The import prices used in the estimation is the rupee price-based unit value index of imports for the products under study (refer Appendix 1 for the list of products). Data for that are taken from the '*Handbook of Statistics on Indian Economy*' published by Reserve Bank of India. The base year of the data on unit value index is 1999–00 = 100. The foreign price (P^*) of each commodity group is proxied by, though restrictive, the producer price index in the USA. On account of paucity in data on price for domestic substitute commodities, wholesale price index is used, which is availed from the database of the Office of Economic Adviser, Department for Promotion of Industry and Internal Trade.

[11]The share of non-oil imports peaked at 78% of total imports in 2015–16. The share was low at 63% in 2013–14. This wide variance is on fluctuations in crude oil prices (Indian basket).

The type of exchange rate indexes is also important for estimation purposes, as is evident from Pollard and Coughlin (2006). The exchange rate data used in the study is nominal effective exchange rate (NEER) based on 36-country bilateral trade weights, with 2004–05 as the base year. As NEER is measured in terms of foreign currency, its increase implies appreciation and appreciation will have a negative influence on the import prices in home currency terms. MFN applied tariff rates across products are collected from TRAINS database and used as an important control in the regression estimates.

Gross domestic product (GDP) of the economy is generally used to specify demand for imports. As the study is at a disaggregated level, we alternately use the sectoral gross value added and value of output for these products in place of GDP. The data on value of output for these products are collected from the Annual Survey of Industry (ASI) database of the Central Statistics Organisation, despite some ambiguities with respect to concordance among product groups of different classification. ASI database provides data according to the National Industrial Classification (NIC). A concordance table between products under study and NIC product groups across difference revision used in this study is presented in Appendix 2.

Panel data technique is used in empirical estimation of ERPT by various studies including Gagnon and Knetter (1995), Feenstra et al. (1996), and Goldberg and Knetter (1999). Such method is used by Mallick and Marques (2006, 2008) for India's import prices and Sinha Roy and Pyne (2014) for India's export prices. This method is at variance with time series estimation of pass-through coefficients often used as in, among others, Athukorala and Menon (1994). With observed endogeneity among macroeconomic variables, McCarthy (1999) and some recent literature estimate ERPT in the VAR framework. The VAR framework along with the vector error correction model (VECM) used by Campa et al. (2005) accounts for non-stationarity of marcoeconomic variables.[12] However, given the data frequency used and the period covered in this study, time series methods including VAR and VECM are inappropriate as it is difficult to obtain significant t-ratios or F-statistics from regressions.

A pooled regression estimate is carried out instead.[13] A pooled estimator ignores the longitudinal or panel aspect of a dataset and treats the observations as cross section. Assuming different distributional patterns across time and cross sections, dummy variables with respect to time and/or cross section can be used as other control variables with interest variables to examine the impact on the dependent variable in

[12]In this study, import price estimation is carried out using the reduced form equation from the import sub-system of a larger macroeconomic model. In this import sub-system, import volume and prices are endogenous while other variables are exogenous, even if they are endogenous in the macroeconomic model. In contrast to this structural simultaneous equation system approach based on a sound theoretical basis, as outlined in Sect. 3.1, the use of VAR as an alternate empirical strategy would require the use of a larger macroeconomic system. A VAR framework runs the risk of being 'atheoretical', as is evident in literature. Qin (2011) looks in detail at the atheoretical nature of VAR models as against structural models. In that sense, in addition to other issues including time series properties of variables in small samples, the use of VAR models is avoided in this econometric exercise.

[13]For a brief exposition on the application of pooled estimation, see Asteriou and Hall (2016).

a regression. One other important aspect in regression with such datasets is that it eliminates the possibility of correlation in error terms among different observations.

In the estimation specification that follows, the variable 'y' has ith cross section at tth time period and there is the vector of independent variable 'x'. Each explanatory variable is assumed to change over time, at least for some cross-sectional groups.

$$y_{it} = \beta_0 + \beta_1 D_t + \beta_2 x_{it} + a_i + u_{it} \tag{9}$$

The variable D_t is the dummy variable to control time-specific effects while a_i (unobserved effect) controls for the cross-sectional-specific fixed effects. β_2 is the parameter to estimate.

4 Estimation Results

This section provides an econometric estimate of exchange-rate pass-through to import prices in India. The degree of exchange-rate pass-through to import prices across sectors is estimated using the model presented in Sect. 3.1. Estimations have been carried out using the pooled estimation method with and without fixed effects across products. The model suits the data well and the estimation results thus obtained are robust. Results for aggregate imports, both with and without fixed effects, are shown in Table 3, while the disaggregated results are presented in Table 4.

The results presented show that the sign of the coefficients of exchange rate in both the models is as expected and significant.[14] In Table 3, columns 1 and 2 show the results without product fixed effects, while columns 3–5 show results with fixed effect. The coefficient of nominal effective exchange rate is high and significant at 1% level in all the model specifications. The result suggests that India's exchange-rate pass-through to import prices is close to complete. Tariff pass-through to import prices is mostly incomplete, and close to complete in a couple of cases (specifications 3 and 4 are important in this regard). Non-inclusion of tariff as a control biases the pass-through coefficient as is observed in column (2) of Table 3. Producers' prices are also positive and significant in determining import prices, while domestic price is not significant.

As an indicator representing domestic demand, GDP is used. GDP is found to be positive and significant in all the model specifications. Sectoral GVA and value of output of the product or product groups are used in alternate model specifications. To eliminate ambiguity with respect to concordance of product groups over different classifications, NIC 2008 is used to arrive at the value of output (2008–09 to 2016–17). These variables also turn out to be significant in determining India's aggregate import prices. However, use of such alternate measures of GDP biases the ERPT coefficients,

[14] In the model estimated, ln(NEER) is a time-variant and sector-invariant variable. This variable may have an endogenous correlation with the error term in the equation. In such a case, instrumental variable method of estimation is appropriate.

Table 3 Estimation of exchange-rate pass-through to aggregate import prices

ln(UVI-import)	(1)	(2)	(3)	(4)	(5)
ln(NEER)	−0.51***	−1.17***	−0.92***	−1.09***	−1.19***
	(0.07)	(0.19)	(0.20)	(0.18)	(0.15)
ln(PPI)	0.45***	0.67***	0.77***	0.77***	0.72***
	(0.08)	(0.10)	(0.10)	(0.10)	(0.12)
ln(WPI)	0.01	0.00	−0.05	−0.04	0.01
	(0.02)	(0.05)	(0.05)	(0.06)	(0.05)
ln(Tariff)	0.16***		0.98***	0.96***	0.53**
	(0.06)		(0.27)	(0.27)	(0.29)
ln(GDP)	0.28***	0.13***	0.25***		
	(0.03)	(0.04)	(0.05)		
ln(GVA: sectoral)				0.21***	
				(0.05)	
ln(output: product)					0.15***
					(0.04)
Observations	557	557	557	557	389
Sector fixed effect	No	No	Yes	Yes	Yes
Adjusted R^2	0.996	0.714	0.721	0.719	0.779

Standard errors in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

with their values surpassing unity. On the other hand, it is also important to note that the inclusion of sector-specific dummies improves the extent of pass-through.

However, the extent of exchange-rate pass-through to import prices, as is evident in Table 4, shows significant variations in the degree of pass-through to import prices across sectors. Estimation results for the alternate model specifications are presented in Appendix 3 as robustness checks for these estimates. For all the sectors, the sign of the coefficient is mostly as expected. However, there are variations in the level of significance of the pass-through coefficients across sectors. For food and food products, import prices are significantly responsive to exchange rate and import tariffs. Both exchange rate and tariff changes are found to have been more than proportionately passed on to import prices. The ERPT result for this sector is in contrast to that of Mallick and Marques (2006, 2008).

For chemical and manufacturing goods classified by materials, exchange-rate pass-though is incomplete, while for machinery and transport equipment exchange-rate pass-though is complete. The results tend to hold good even in alternate specifications. For the heterogeneous sector ‘manufacturing goods classified by materials’, the extent pass-through to import prices is found to vary widely across sub-sectors, being complete and significant for textiles to being incomplete and insignificant for iron and steel (see Appendix 4). For crude materials, also a heterogeneous product group, the pass-through coefficient is found to be low and insignificant. Like all other

Table 4 Estimation of exchange-rate pass-through to import prices at the product level

	(1)	(2)	(3)	(4)	(5)
	Food products	Crude materials	Chemical products	Manufactured goods by materials	Machinery and transport
ln(NEER)	−1.57***	−0.22	−0.58**	−0.53**	−1.26**
	(0.39)	(0.36)	(0.23)	(0.22)	(0.57)
ln(PPI)	1.97***	1.04***	0.49***	0.41***	−2.17***
	(0.48)	(0.14)	(0.10)	(0.12)	(0.60)
ln(WPI)	0.02	−0.24***	0.01	0.32**	−0.51
	(0.22)	(0.08)	(0.12)	(0.14)	(0.48)
ln(Tariff)	4.22***	−0.11	−0.20	0.01	−0.65
	(1.28)	(0.64)	(0.36)	(0.29)	(0.98)
ln(GDP)	−0.00	0.40***	0.18**	0.08	0.69***
	(0.18)	(0.10)	(0.07)	(0.07)	(0.17)
Observations	80	115	110	84	103
Sector fixed effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.876	0.888	0.921	0.862	0.680

Standard errors in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

sectors, except food products, the tariff pass-through is also found to be insignificant for crude materials. This result tends to show that as India's production and imports have become raw material import-intensive,[15] exchange rate changes are necessarily not passed on to importers' price, especially when exchange rate is found to have depreciated on the average.

The observed variations in the degree of pass-through across sectors can be largely on account of sector-specific factors including market concentration and competition, the degree of (non) homogeneity of products, among others. The regression estimations have only controlled for these factors in terms of product-specific fixed effects. Moreover, trade openness is only factored in using tariff rates. Even though Mallick and Marques (2006, 2008) show that variations in sectoral pass-through on account of the sector's import orientation as well as effective rate of protection, the above results in this study often do not conform to such a pattern. The source countries of these imports are treated in aggregate, source country-specific fixed effects could have provided nuanced pass-through results. The varying nature of exchange-rate pass-through to import prices across sectors can also be explained in terms of exporting countries interest in maintaining their share in India's market, if India is a large importing country in case of certain products. If source country specific fixed effects are accounted for, a better explanation of variations in pass-through coefficients could have been provided.

On the whole, the evidence shows incomplete exchange pass-through to Indian import prices. The results show, in contrast to Mallick and Marques (2008), that exchange rate is a more effective policy instrument than tariff in maintaining import

[15] Ghosh and Sinha Roy (2018).

market share across sectors. The pass-through coefficient is found to be high in some sectors, and near complete in some others. For instance, as exchange rate depreciation leads to proportionately higher food prices and import prices of machinery. These results contribute to the existing understanding of exchange rate changes being passed on to import prices and inflation in general. High or near-complete pass-through leads unit costs to rise in certain sectors and contributes to building up of inflationary pressure in the economy. This effect, despite exchange-rate pass-through to fuel prices not holding good, has implications for targeting inflation.

5 Conclusions

This paper has investigated the degree of exchange-rate pass-through to prices of imports across sectors in India during reforms. For the purpose of analysis, certain non-oil imports including chemicals, machinery, transport equipment, manufacture classified by materials, crude materials and food processing into India are considered, which account for about 70% of the country's total imports. It is observed that, during post-reforms, import prices have moved synchronously with exchange rate indicating the possibility of exchange-rate pass-through to import prices. An econometric model based on simultaneous equation imperfect substitutes model is made use of as an empirical strategy, which is distinctly different from econometric models being used for estimation purposes in the literature. Pooled regression technique is used as the econometric method on a time-comparable dataset carefully compiled from different sources. The model has suited the data well and the results are thus robust.

The econometric results show significant exchange-rate pass-through to import prices in India. It can be observed that significant exchange-rate pass-through to import prices occurs when all the sectors are taken into account, but pass-through to import prices is incomplete. However, the degree of exchange-rate pass-through is found to vary across import sectors. In the case of chemicals and manufacturing classified by materials, the degree of exchange-rate pass-through to import prices is incomplete, while for 'food & food products' and 'machinery & transport equipment' exchange-rate pass-through are more than proportionate. The pass-through coefficient is low and insignificant in the case of 'crude materials other than fuel'. As pass-through coefficient is varying across sectors, factors such as market concentration of the exporting country and product characteristics might be of importance in this analysis. Estimations are controlled for the possible effect import tariff on import prices which is found as positive and significant determinant of aggregate import prices but insignificant in the case of most sectors. The above results, however, do not account for the nature of pass-through during the post-financial crises period. This may be an important gap in the analysis. Nonetheless, these results have important implications for policy especially in the context of rising import intensity in production and consumption as well as the policy to keep inflation within bounds.

Appendix 1

See Table 5.

Table 5 List of products/product groups under study

	Product/product groups		Product/product groups
1	**Food and live animals**	6	**Manufactured goods classified chiefly by material**
	Dairy products		Paper, paperboard and articles thereof
	Cereals and cereal preparations		Textile yarn
	Fruits and nuts		Iron and steel
	Spices		Copper
2	**Beverages and tobacco**		Aluminium
	Beverages	7	**Machinery and transport equipment**
3	**Crude materials, inedible, except fuels**		Power generating machinery and equipment
	Crude rubber including synthetic and reclaimed		Machinery specialized for particular industries
	Textile fibres and their wastes		Metal working machinery
	Crude fertilizers		General industrial machinery and equipment
	Minerals		Office machines and automatic data
	Ores and concentrates of base metals		Electrical machinery
4	**Animal and vegetable oil**		Transport equipment
5	**Chemicals and related products**	8	**Miscellaneous manufactured articles**
	Organic chemicals		Professional, scientific and controlling instruments
	Inorganic chemicals		Photographic apparatus, etc
	Dyeing, tanning and colouring materials		
	Medicinal and pharmaceutical products		
	Fertilisers, manufactured		
	Artificial resin and plastic material and cellulose ester		

Appendix 2

See Table 6.

Table 6 Concordance between products under NIC products

S. No.	Product	NIC-1987 (1993–94 to 1997–98)	NIC-1998 and 2004 (1998–99 to 2007–08)	NIC-2008 (2008–09 to 2016–17)
1	**Food and live animals**		**15**	**10**
	Dairy products		152	105
	Cereals and cereal preparations		153	106
	Fruits and nuts		1513	103
	Spices		01	01
2	**Beverages and tobacco**	**(220–227)**	**(155 + 16)**	**11 + 12**
	Beverages		155	11
3	**Crude materials, inedible, except fuels**		**(10 + 11 + 13)**	**(05 + 06 + 07)**
	Crude rubber including synthetic and reclaimed		**25**	22
	Textile fibres and their wastes	**(230–243)**	171	1311
	Crude fertilizers		2412	2012
	Minerals		26	23
	Ores and concentrates of base metals		13	7
4	**Animal and vegetable oil**	**(210 + 211 + 212)**	**1514**	**104**
5	**Chemicals and related products**		**24**	**20**
	Organic chemicals		242	202
	Inorganic chemicals		241	201
	Dyeing, tanning and colouring materials		2422	2022
	Medicinal and pharmaceutical products		2423	210

(continued)

Table 6 (continued)

S. No.	Product	NIC-1987 (1993–94 to 1997–98)	NIC-1998 and 2004 (1998–99 to 2007–08)	NIC-2008 (2008–09 to 2016–17)
	Fertilisers, manufactured		241	201
	Artificial resin and plastic material and cellulose ester		2520	222
6	**Manufactured goods classified chiefly by material**		**26**	**23**
	Paper, paperboard and articles thereof		21	17
	Textile yarn		171	131
	Iron and steel		271	241
	Copper		131	072
	Aluminium		120	072
7	**Machinery and transport equipment**		**29**	**28**
	Power generating machinery and equipment		2813	2513
	Machinery specialized for particular industries		29	332
	Metal working machinery		289	259
	General industrial machinery and equipment		291	281
	Office machines and automatic data		33	26
	Electrical machinery		31	27
	Transport equipment		**(34 + 35)**	29 + 30
8	**Miscellaneous manufactured articles**		36	**32**
	Professional, scientific and controlling instruments		73	72
	Photographic apparatus, etc		9309	742

The bold indicates aggregate product group, while the ones without bold are product sub-groups

Appendix 3

See Tables 7, 8, 9, 10 and 11.

Table 7 Estimation results for '*Food and live animals*'

	(1)	(2)	(3)
ln(NEER)	−1.57***	−1.56***	−1.51***
	(0.39)	(0.38)	(0.38)
ln(PPI-US)	1.97***	2.07***	1.88***
	(0.48)	(0.46)	(0.43)
ln(WPI)	0.02	0.08	−0.59**
	(0.22)	(0.21)	(0.28)
ln(Tariff)	4.22***	4.27***	4.20***
	(1.28)	(1.29)	(1.28)
ln(GDP)	−0.00		
	(0.18)		
ln(GVA-sectoral)		−0.06	
		(0.14)	
ln(output: product)			0.31**
			(0.14)
Observations	80	80	76
Sector fixed effect	Yes	Yes	Yes
Adjusted R^2	0.876	0.877	0.873

Standard errors in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 8 Estimation results for '*Crude materials, inedible, except fuels*'

	(1)	(2)	(3)
ln(NEER)	−0.22	−0.32	−0.94***
	(0.36)	(0.34)	(0.23)
ln(PPI-US)	1.04***	1.03***	0.83***
	(0.14)	(0.14)	(0.17)
ln(WPI)	−0.24***	−0.25***	−0.34***
	(0.08)	(0.08)	(0.08)
ln(Tariff)	−0.11	0.26	−1.20**
	(0.64)	(0.70)	(0.56)
ln(GDP)	0.40***		
	(0.10)		
ln(GVA-sectoral)		0.43***	

(continued)

Table 8 (continued)

	(1)	(2)	(3)
		(0.11)	
ln(output: product)			0.29***
			(0.10)
Observations	115	115	69
Sector fixed effect	Yes	Yes	Yes
Adjusted R^2	0.888	0.888	0.919

Standard errors in parentheses, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 9 Estimation results for '*Chemicals and related products*'

	(1)	(2)	(3)
ln(NEER)	−0.58**	−0.63***	−0.88***
	(0.23)	(0.22)	(0.22)
ln(PPI-US)	0.49***	0.47***	0.50***
	(0.10)	(0.10)	(0.11)
ln(WPI)	0.01	0.01	−0.14
	(0.12)	(0.12)	(0.20)
ln(Tariff)	−0.20	−0.06	−0.87**
	(0.36)	(0.41)	(0.35)
ln(GDP)	0.18**		
	(0.07)		
ln(GVA-sectoral)		0.20**	
		(0.08)	
ln(output: product)			0.10***
			(0.03)
Observations	110	110	93
Product fixed effect	Yes	Yes	Yes
Adjusted R^2	0.921	0.920	0.913

Standard errors in parentheses, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 10 Estimation results for '*Manufactured goods classified by material*'

	(1)	(2)	(3)
ln(NEER)	−0.53**	−0.53**	−0.30
	(0.22)	(0.21)	(0.32)
ln(PPI-US)	0.41***	0.41***	0.53*
	(0.12)	(0.12)	(0.29)
ln(WPI)	0.32**	0.31**	0.95***
	(0.14)	(0.14)	(0.23)
ln(Tariff)	0.01	0.08	−0.32
	(0.29)	(0.32)	(0.38)
ln(GDP)	0.08		
	(0.07)		
ln(GVA-sectoral)		0.09	
		(0.07)	
ln(output: product)			−0.15
			(0.13)
Observations	84	84	46
Product fixed effect	Yes	Yes	Yes
Adjusted R^2	0.862	0.863	0.866

Standard errors in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 11 Estimation results for '*Machinery and transport equipment*'

	(1)	(2)	(3)
ln(NEER)	−1.26**	−1.27**	−2.53***
	(0.57)	(0.55)	(0.57)
ln(PPI-US)	−2.17***	−2.19***	−0.98
	(0.60)	(0.59)	(0.98)
ln(WPI)	−0.51	−0.54	−0.18
	(0.48)	(0.48)	(0.77)
ln(Tariff)	−0.65	0.34	−2.45
	(0.98)	(1.10)	(1.70)
ln(GDP)	0.69***		
	(0.17)		
ln(GVA-sectoral)		0.81***	
		(0.19)	
ln(output: product)			0.01
			(0.19)
Observations	103	103	86
Adjusted R^2	0.680	0.687	0.595

Standard errors in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 12 Estimation results for textiles and iron and steel

Variables	(1)	(2)	(3)	(4)
	Textiles	Textiles	Iron and steel	Iron and steel
ln(NEER)	−1.46***	−1.09***	−0.22	−0.22
	(0.36)	(0.37)	(0.21)	(0.23)
ln(PPI)	1.26***	1.26***	0.40**	0.33
	(0.28)	(0.27)	(0.14)	(0.20)
ln(WPI)	0.06	0.00	0.66	0.70
	(0.18)	(0.16)	(0.37)	(0.41)
ln(tariff)		0.80*		−0.18
		(0.42)		(0.34)
ln(GDP)	0.00	0.18	−0.11	−0.13
	(0.06)	(0.10)	(0.16)	(0.20)
Constant	5.22*	−2.65	2.23	3.60
	(2.88)	(4.63)	(1.87)	(3.93)
Observations	25	25	15	15
Adjusted R^2	0.95	0.96	0.85	0.83

Robust standard errors in parentheses, ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Appendix 4

See Table 12.

References

Asteriou, D., & Hall, S. G. (2016). *Applied econometrics*. London: Palgrave (Macmillan Education).
Athukorala, P., & Menon, J. (1994). Pricing-to-market behavior and exchange rate pass-through in Japanese exports. *Economic Journal, 104*(423), 271–281.
Bacchetta, P., & vanWincoop, E. (2002). *A theory of the currency denomination of international trade*. NBER Working Paper No. 9039, NBER, Cambridge, MA, July.
Balakrishnan, P., & Parameswaran, M. (2019). *The Dynamics of Inflation in India.* Working Paper no. 485, Thiruvananthapuram: Centre for Development Studies, March.
Campa, J. M., & Minguez, J. M. G. (2006). Differences in exchange rate pass-through in the Euro Area. *European Economic Review, 50,* 121–145.
Campa, J. M., Goldberg, L. S., & González-Minguez, J. M. (2005, September). *Exchange rate pass-through to import prices in the Euro Area.* NBER Working Paper No. 11632, NBER, Cambridge, MA.
Campa, J., & Goldberg, L. S. (2002, May). *Exchange rate pass-through into import prices: A macro or micro phenomenon?* NBER Working Paper No. 8934, NBER, Cambridge, MA.
Choudhri, E. U., & Hakura, D. S. (2001). *Exchange rate pass-through to domestic prices: Does the inflationary environment matter?* International Monetary Fund Working Paper 01/194.
Dornbush, R. (1987). Exchange rates and prices. *American Economic Review, 77*(1), 93–106.

Dash, A. K., & Narasimhan, V. (2011). Exchange rate pass-through: How much do exchange rate changes affect the prices of Indian exports and imports. *South Asia Economic Journal, 12*(1), 1–23.

Dholakia, R. H., & Saradhi, R. V. (2000). Exchange rate pass-through and volatility: Impact on foreign trade. *Economic and Political Weekly, 35*(47), 4109–4116.

Engel, C. (2005, March). *Equivalence results for optimal pass-through, optimal indexing to exchange rates, and optimal choice of currency for export pricing.* NBER Working Paper No. 11209, NBER, Cambridge, MA.

Faruqee, H. (2004). *Exchange rate pass-through in the Euro Area: The role of asymmetric pricing behavior*. IMF Working Paper 04/14, Washington D. C.: International Monetary Fund.

Feenstra, R. (1989). Symmetric pass-through of tariffs and exchange rates under imperfect competition: An empirical test. *Journal of International Economics, 27,* 25–45.

Feenstra, R., Gagnon, J., & Knetter, M. M. (1996). Market share and exchange rate pass-through in world automobile trade. *Journal of International Economics*, *40*(1–2), 187–207.

Feenstra, R. C., Gagnon, J. E., & Knetter, M. M. (1993, July). *Market share and exchange rate pass-through in world of automobile trade.* NBER Working Paper No. 4399, NBER, Cambridge, MA.

Gagnon, J., & Knetter, M. M. (1995). Markup adjustment and exchange rate fluctuations: Evidence from panel data on automobile exports. *Journal of International Money and Finance, 14*(2), 289–310.

Ghosh, A., & Rajan, R. S. (2007). How high exchange rate pass-through in India? Has it changed over time? *Journal of International Trade and Economic Development, 16*(3), 373–382.

Ghosh, A., & Rajan, R. S. (2009). What is the extent of exchange rate pass-through in Singapore? Has it changed over time? *Journal of the Asia Pacific Economy, 14*(1), 61–72.

Ghosh, M., & Sinha Roy, S. (2018). Foreign direct investment, firm heterogeneity, and exports: An analysis of Indian manufacturing. *Asian Development Review, 35*(1), 27–51.

Goldberg, P. K., & Knetter, M. M. (1999). Measuring the intensity of competition in export markets. *Journal of International Economics*, *47*(1), 27–60.

Goldberg, P. K., & Knetter, M. M. (1997). Goods prices and exchange rate: What have we learned? *Journal of Economic Literature, 35*(3), 1243–1272.

Goldstein, M., & Khan, M. S. (1985). Income and price effects in foreign trade. In R. W. Jones & P. B. Kenen (Eds.), *Handbook of international economics* (Vol. II, pp. 1041–1105). Amsterdam: North-Holland.

Halpern, L., & Koren, M. (2007). Pricing to firm: An analysis of firm and product-level import prices. *Review of International Economics, 15*(3), 574–591.

Hooper, P., & Mann, C. L. (1989). Exchange rate pass-through in the 1980s: The case of U.S. imports of manufactures. *Brookings Papers on Economic Activity, 1*, 297–337.

Jašová, M., Moessner, R., & Takáts, E. (2016). *Exchange rate pass-through: What has changed since the crisis?* BIS Working Papers No. 583, Basel: Bank for International Settlements.

Khundrakpam, J. K. (2008). Have economic reforms affected exchange rate pass-through to prices in India? *Economic & Political Weekly, XLIII*(16), 71–79.

Kim, Y. (1990). Exchange rates and import prices in the United States: A varying-parameter estimation of exchange rate pass-through. *Journal of Business & Economic Statistics, 8*(3), 305–315.

Kreinin, M., Martin, S., & Sheehey, E. J. (1987). Differential response of U.S. import prices and quantities to exchange-rate adjustments. *Review of World Economics*, *123*(3), 449–462.

Krugman, P. (1987). Pricing to market when the exchange rate changes. In J. D. Richardson (Ed.), *Real financial linkages among open economies* (pp. 49–70). Cambridge, MA: MIT press.

Lopez-Villavicencio, A., & Mignon, V. (2017). *On the seemingly incompleteness of exchange rate pass-through to import prices: Do globalization and/or regional trade matter?* CEPII Working Paper No. 2017–08, Paris: CEPII.

Mallick, S., & Marques, H. (2006). Sectoral exchange rate pass-through: Testing the impact of policy reforms in India. *Scottish Journal of Political Economy, 53*(2), 280–303.

Mallick, S., & Marques, H. (2008). Pass-through of exchange rate and tariffs into import prices of India: Currency depreciation versus import liberalisation. *Review of International Economics, 16*(4), 765–782.

Marazzi, M., et al. (2005). *Exchange rate pass-through to US import prices: Some new evidence*. International Finance Discussion Paper 2005-833. Board of Governors of the Federal Reserve System.

McCarthy, J. (1999). *Pass-through of exchange rates and import prices to domestic inflation in some industrialised economies*. BIS Working Papers No 79, Bank for International Settlements.

Panagariya, A. (2005). *India's trade reform* (Vol. 1). The Policy Reform Forum 2004. Brookings and NCAER, India.

Patra, M. D., Khundrakpam, J. K., & John, J. (2018). *Non-linear, asymmetric and time-varying exchange rate pass-through: Recent evidence from India.* RBI Working Paper Series (DEPR): 02/2018, Mumbai: Reserve Bank of India.

Pollard, P. S., & Coughlin, C. C. (2006). Pass-through estimates and the choice of an exchange rate index. *Review of International Economics, 14*(4), 535–553.

Pollard, P. S. & Coughlin, C. C. (2004). *Size matters: asymmetric exchange rate passthrough at the industry level*, FRB Working Paper Series 2003–029C, The Federal Reserve Bank of St Louis.

Pyne, P. K. (2008). *Exchange rate pass-through and import prices: A sector-wise analysis for India*. Unpublished M. Phil. Dissertation, Kolkata: Jadavpur University.

Qin, D. (2011). Rise of VAR modelling approach. *Journal of Economic Surveys, 25*(1), 156–174.

Sengupta, A., & Sinha Roy, S. (2018). India's trade policy: Which way now? *Decision—Official Journal of Indian Institute of Management Calcutta, 45*(2), 129–145.

Sharma, A., Hatekar, N., & Kulkarni, S. (2011). What drives inflation in India: Overheating or input costs? *Economic & Political Weekly, 46*(34), 46–51.

Sinha Roy, S. (2005). Post-reform export growth in India: An explanatory analysis. In N. Banerjee & S. Marjit (Eds.), *Development, displacement and disparity: India in the last quarter of the twentieth century, Hyderabad*. India: Orient Longman Limited.

Sinha Roy, S., & Pyne, P. K. (2014). Reforms, exchange rate pass-through and India's export prices. In R. Acharyya & Margjit, S. (Eds.), *Trade, globalization and development: Essays in honour of Kalyan K. Sanyal*. New Delhi: Springer.

Taylor, J. B. (2000). Low Inflation, pass-through, and the pricing power of firms. *European Economic Review, 44,* 1389–1408.

Zorzi, M., Hahn, E., & Sanchez, M. (2007). *Exchange rate pass-through in emerging markets.* ECB Working Paper no. 739. Brussels: European Central Bank.

Pricing of Energy Goods in India: How Is It Done?

Amrita Ganguly and Ranajoy Bhattacharyya

Abstract The retail price of petrol, diesel, liquid petroleum gas (LPG) and other energy products in India is a complex mixture of crude prices, taxes and subsidies by the central as well as state governments. In this paper, we discuss the break-up of these components of pricing. The discussion is set in the background of international pricing mechanisms and recommendations of international agencies such as the International Institute of Sustainable Development and the International Monetary Fund. The main point that emerges from the discussion is that since these products are necessities with very low elasticity of demand, incidence of taxes and/or subsidies on these products are almost 100% making them ideal tools for the international cartels (seeking profits) and local governments (seeking revenues) to maximize their pecuniary benefits from the manipulation of such prices. The complex formula leading to the retail price is a corollary of these interests.

Keywords Petroleum pricing · Energy subsidy · Electricity cross-subsidy · Diesel subsidy · LPG subsidy · Petrol subsidy

1 Introduction

India has historically subsidized energy with the objective of protecting its consumers from international price volatility and providing energy access for its citizens, especially the poor. However, energy subsidies place a heavy burden on government exchequer, while often failing to reach their targeted beneficiaries. Energy subsidies also have significant ecological ramifications due to overuse of fossil fuels. Increases

A. Ganguly
KPMG Advisory Services Private Limited, Godrej Waterside, Unit No. 604, 6th Floor, Tower 1, Plot No. 5, Block – DP, Sector V, Salt Lake, Kolkata 700091, India
e-mail: amritag@kpmg.com

R. Bhattacharyya (✉)
Indian Institute of Foreign Trade, Kolkata Campus, 1583 Madurdaha Chowbaga Road, Kolkata 700107, India
e-mail: ganguly.amrita@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_14

in emissions due to overuse of petroleum products and inefficient consumption of cheap fuel are genuine and urgent concerns. The petroleum sector is one of the most heavily subsidized energy sources in India. India's fuel subsidies reached an estimated INR 1.4 trillion in the financial year ending March 2014,[1] representing about 1.9% of total GDP.[2] The pricing of petroleum products has changed over time, alternating between free market and regulated regimes. Until 2010, the central government controlled the prices of petrol, diesel, kerosene and liquid petroleum gas (LPG). In June 2010, the Indian government deregulated the price of petrol, and in October 2014, price of diesel was deregulated, taking advantage of the falling crude oil prices. All this has meant that the actual retail price of energy goods is somewhat of a mystery to the users and even to all but the industry insiders.

This paper is intended to provide information on the break-up of prices of energy goods in India. We consider only four of the several energy goods in use today in India: petrol, diesel, LPG and electricity. Since the first three goods are derivatives of crude petroleum, the international crude oil market and its price variations are also briefly discussed. The rest of the paper is arranged as follows: the next section briefly studies some of the ways in which international crude oil prices are decided upon. Sections 3 and 4 analyse the price break-up of petroleum products and electricity in India.

2 International Crude Oil Pricing

Until the late 1950s, the international oil industry outside the USA, Canada, the erstwhile USSR and China was characterized by the dominant position of large multinational oil companies known as the Seven Sisters or the majors. In 1950, these majors controlled 85% of the crude oil production in the world outside Canada, USA, Soviet Russia and China (Danielsen 1982). The host governments did not participate in production or pricing of crude oil and acted only as competing sellers of licences or oil concessions. In return, host governments received a stream of income through royalties and income taxes based on posted prices. Being a fiscal parameter, the posted price did not respond to the usual market forces of supply and demand and thus did not play any allocation function. The multinational oil companies were comfortable with the system of posted prices because it maintained their oligopolistic position, and until the late 1960s, OPEC countries were too weak to change the existing pricing system.

By the late 1950s, this dominance of the vertically integrated companies was challenged by the arrival of independent oil companies who were able to invest in upstream operations and obtain access to crude oil outside the Seven Sisters' control. Several factors like the US decision to impose mandatory import quotas

[1] http://www.thenational.ae/business/economy/energy-subsidies-prove-drain-on-indian-economy.
[2] https://www.imf.org/external/country/ind/rr/2013/050613a.pdf.

which increased competition for outlets outside the USA were an additional factor that placed downward pressure on oil prices. The formation of the Organization of the Petroleum Exporting Countries (OPEC) in 1960 was an attempt by member countries to prevent the decline in the posted price, and thus for most of the 1960s, OPEC acted as a trade union whose main objective was to prevent the income of its member countries from declining.

The OPEC-administered pricing system emerged around 1965–1973. The oil industry witnessed a major transformation in the early 1970s when some OPEC governments stopped granting new concessions and started to claim equity participation in their existing concessions, with some of them opting for full nationalization. Equity participation gave OPEC governments a share of the oil produced which they had to sell to third-party buyers. It led to the introduction of new pricing concepts to deal with this reality (Mabro 1984). As owners of crude oil, governments had to set a price for third-party buyers. The concept of official selling price (OSP) or government selling price (GSP) entered at this point and is still currently used by some oil exporters. However, for reasons of convenience, lack of marketing experience and inability to integrate downwards into refining and marketing in oil-importing countries, most of the governments' share was sold back to the companies that held the concession and produced the crude oil. These sales were made compulsory as part of equity participation agreements and used to be transacted at buyback prices. The administered oil pricing regime that emerged in 1974–75 after the short-lived episode of the buyback system was radical in many aspects, as it represented a complete shift in the power of setting the oil price from multinational companies to OPEC. The new system was centred on the concept of reference or marker price with Saudi Arabia's Arabian Light being the chosen marker crude. In this administered pricing system, individual members retained the OSPs for their crudes, but these were now set in relation to the reference price. The differential relative to the marker price used to be adjusted periodically depending on a variety of factors such as the relative supply and demand for each crude variety and the relative price of petroleum products among other things.

In the late 1970s, national oil companies in OPEC started to increase the number of their non-concessionaire customers. The appearance of independent oil companies, Japanese and independent refineries, state oil companies, trading houses and oil traders permitted such a development. The pace accelerated during and in the aftermath of the 1979 Iranian crisis. The new regime in Iran cancelled any previous agreements with the oil majors in marketing Iranian oil—they became mere purchasers as with any other oil companies. Faced with this virtual disruption of traditional supply channels, multinational oil companies were forced to enter the market. This had a profound effect on oil markets as de-integration and the emergence of new players expanded the external market where buyers and sellers engaged in arm's length transactions. The crude market became more competitive and the majority of oil used to move through short-term contracts or the spot market.

The decline in oil demand in the mid-1980s caused by a worldwide economic recession and the growth in non-OPEC crude oil production responding to higher oil prices and taking advantage of new technologies represented major challenges to

OPEC's administered pricing system and were ultimately responsible for its demise. New discoveries in non-OPEC countries meant that significant amounts of oil began to reach the international market from outside OPEC. This increase in supply also meant an increase in the number and diversity of crude oil producers who were setting their prices in line with market conditions and hence proved to be more competitive. The new suppliers who ended up having more crude oil than required by contract buyers secured the sale of all their production by undercutting OPEC prices in the spot market. Buyers who became more diverse were attracted to these offers of competitive prices. With the continued decline in demand for its oil, OPEC saw its own market share in the world's oil production fall from 51% in 1973 to 28% in 1985. It became clear by the mid-1980s that the OPEC-administered oil pricing system was unlikely to hold for long and OPEC's or more precisely Saudi Arabia's, attempts to defend the marker price would only result in loss of market share as other producers could offer to sell their oil at a discount to the administered price of Arabian Light. OPEC abandoned the administered pricing system and transferred the pricing power of crude oil to the so-called market.

The collapse of the OPEC-administered pricing system in 1986–1988 ushered in a new era in oil pricing in which the power to set oil prices shifted from OPEC to the so-called market. First adopted by the Mexican national oil company PEMEX in 1986, the market-related pricing system received wide acceptance among most oil-exporting countries, and by 1988, it became and still is the main method for pricing crude oil in international trade. The oil market was ready for such a transition. The end of the concession system and the waves of nationalizations, which disrupted oil supplies to multinational oil companies, established the basis of arm's length deals and exchange outside the vertically and horizontally integrated multinational companies. The emergence of many suppliers outside OPEC and more buyers further increased the prevalence of such arm's length deals. This led to the development of a complex structure of interlinked oil markets which consists of spot and also physical forwards, futures, options and other derivative markets referred to as paper markets. Technological innovations that made electronic trading possible revolutionized these markets by allowing 24-h trading from any place in the world.

Given the large variety of crude oils, the price of a particular crude oil is usually set at a discount or at a premium to a marker or reference price. These references prices are often referred to as benchmarks. The formula used in pricing oil in long-term contracts is straightforward. Specifically, for crude oil of variety x, the formula pricing can be written as

$$Px = PR \pm D$$

where Px is the price of crude x; PR is the benchmark crude price; and D is the value of the price differential. The differential is often agreed at the time when the deal is concluded and could be set by an oil-exporting country or assessed by price reporting agencies. It is important to note that formula pricing may apply to all types of contractual arrangements, be they spot, forward or long term. For instance, a spot transaction in the crude oil market is—pricing wise—an agreement on a spot value of the differential between the physical oil traded and the price of an agreed oil benchmark, which fixes the absolute price level for such trade, normally around the time of delivery or the loading date.

Differences in crude oil quality are not the only determinant of crude oil price differentials however. The movements in differentials also reflect movements in the gross products' worth (GPW) obtained from refining the reference crude *R* and the crude *x*.[3] Thus, price differentials between the different varieties of crude oil are not constant and change continuously according to the relative demand and supply of the various crudes which in turn depend on the relative prices of petroleum products. For instance, at the beginning of 2008, the differentials between Arab Super Light and Arab Heavy widened sharply to reach more than US$15 per barrel; fuel oil, a product of heavy crude, was in surplus while the demand for diesel, a product of lighter crudes, was high. In the first months of 2009, the price differential between heavy and light crude oil narrowed to very low levels as the implementation of OPEC cuts reduced the supply of heavy crude and increased the relative value of heavy sour crudes.

The 'equivalence to the buyer' principle, which means that in practice prices of crudes have equivalent prices at destination, adds another dimension to the pricing formulae. The location in which prices should be compared is not the point of origin but must be closer to the destination where the buyer receives the cargo. Since the freight costs vary depending on the export destination, some formulae also take into account the relative freight costs between destinations. Specifically, they allow for the difference between the freight costs involved in moving the reference crude from its location to a certain destination (e.g. Brent from Sullom Voe to Rotterdam) and the costs involved in moving crude *x* from the oil country's terminal to that certain destination (e.g. Arabian Light from Ras Tanura to Rotterdam). In such cases, the sale contract is close to a cost, insurance and freight (CIF) contract. This is in contrast to a free on board (FOB) contract which refers to a situation in which the seller fulfils his obligations to deliver when the goods have passed over the ship's rail. The buyer bears all the risks of loss of or damage to the goods from that point as well as all other costs such as freight and insurance.

A major advantage of formula pricing is that the price of an oil shipment can be linked to the price at the time of delivery which reflects the market conditions prevailing. When there is a lag between the date at which a cargo is bought and the date of arrival at its destination, there is a big price risk. Transacting parties usually share this risk through the pricing formula.

At the heart of formulae, pricing is the identification of the price of key 'physical' benchmarks, such as West Texas Intermediate (WTI), the ASCI price, Dated Brent (also called Dated North Sea Light, North Sea Dated, Dated BFOE) and Dubai (see Fig. 1). The prices of these benchmark crudes, often referred to as 'spot' market prices, are central to the oil pricing system. The prices of these benchmarks are used by oil companies and traders to price cargoes under long-term contracts or in

[3]Individual crudes have a particular yield of products with a gross product worth (GPW). GPW depends both on the refining process and the prices at which these products are sold.

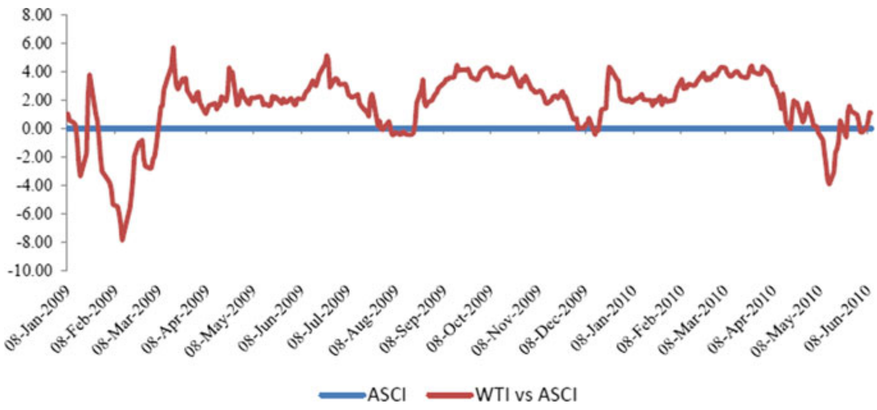


Fig. 1 Price differential between WTI and ASCI ($/Barrel) (ASCI Price = 0). *Source* Argus

spot market transactions, by futures exchanges for the settlement of their financial contracts; by banks and companies for the settlement of derivative instruments such as swap contracts; and by governments for taxation purposes.

Given the central role that benchmarking plays in the current oil pricing system, it is important to highlight some of the main features of the most widely used benchmarks. These are summarized in Table 1. First, unlike the futures market where prices

Table 1 Some basic features of benchmark crudes

First-quarter 2010 averages by Argus	ASCI	WTI CMA + WTI P-Plus	Forties	BFOE	Dubai	Oman
Production (MBPD)	736	300–400	562	1220	70–80	710
Volume spot traded (MBPD)	579	939	514	635	86	246
Number of spot trades per cal month	260	330	18	98	3.5	10
Number of spot trades per day	13	16	<1	5	<1	<1
Number of different spot buyers per cal month	26	27	7	10	3	5
Number of different spot sellers per cal month	24	36	6	9	3	6
Largest 3 buyers % of total spot volume (%)	43	38	63	72	100	50
Largest 3 sellers % of total spot volume (%)	38	51	76	56	100	80

Source Argus

are observable in real time, the reported prices of physical benchmarks are identified or assessed prices. These assessments are carried out by oil pricing reporting agencies, the two most important of which are Platts and Argus. Assessments are needed in opaque markets such as oil where physical transactions concluded between parties cannot be directly observed by market participants. After all, parties are under no obligation to report their deals. Assessments are also needed in illiquid markets where not enough representative deals or where no transactions take place. Oil reporting agencies assess their prices based on information on concluded deals which they observe, or bids and offers, and failing that on market talk, other private and public information gathered by reporters, and information from financial markets. It is important to note that PRAs do not use in all markets a hierarchy of information cascading down from deals to bids and offers, which would imply that deals are the best price discovery and bids/offers are a poorer alternative. The methodology may vary from market to market in accordance with the published methodology for that market. In some markets, bid/offer information takes precedence over deals in identifying the published price—e.g. if the deal is either not representative of the market as defined in the methodology, or was done earlier or later in the day to the prevailing depth of market. In other markets, price identification relies on observed deals. For instance, Argus' main benchmark ASCI is entirely deal based. Most, however, accept that a done deal does represent the highest form of proof of value, unless there is a supervening issue with the trade's conduct. If assessments are intended to represent an end-of-day price, analogous to a futures settlement however, a fully evidenced bid/offer spread at a later point when markets have clearly moved in value is an acceptable proxy in the absence of a trade.

Finally, in the last two decades or so, many financial layers (paper markets) have emerged around these benchmarks. These include the forward market (in Brent), swaps, futures and options. Some of the instruments such as futures and options are traded on regulated exchanges such as Intercontinental Exchange (ICE) and Chicago Mercantile Exchange (CME) Group, while other instruments, such as swaps and forward contracts, are traded bilaterally OTC. Nevertheless, these financial layers are highly interlinked through the process of arbitrage and the development of instruments that link the various markets together such as the Exchange of Futures for Swaps (EFS) which allow traders to roll positions from futures to swaps and vice versa. Over the years, these markets have grown in terms of size, liquidity, sophistication and have attracted a diverse set of players, both physical and financial. These markets have become central for market participants wishing to hedge their risk and to bet (or speculate) on oil price movements. The final price trend appearing out of these is reported in Fig. 2. Note that though the different prices are close, they remain distinct throughout the period.

Since India imports 85% of her crude from OPEC countries, we here consider only the OPEC crude oil price. Introduced on 16 June 2005, the OPEC Reference Basket is currently made up of the following: Saharan Blend (Algeria), Girassol (Angola), Oriente (Ecuador), Minas (Indonesia), Iran Heavy (Islamic Republic of Iran), Basra Light (Iraq), Kuwait Export (Kuwait), Es Sider (Libya), Bonny Light (Nigeria), Qatar

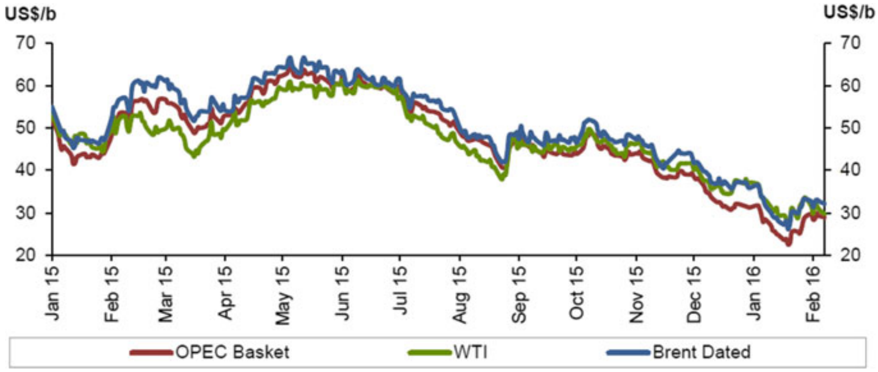


Fig. 2 Crude oil price movement, 2015–2016 (http://www.opec.org/opec_web/static_files_project/media/downloads/publications/MOMR%20February%202016.pdf). *Source* OPEC Secretariat

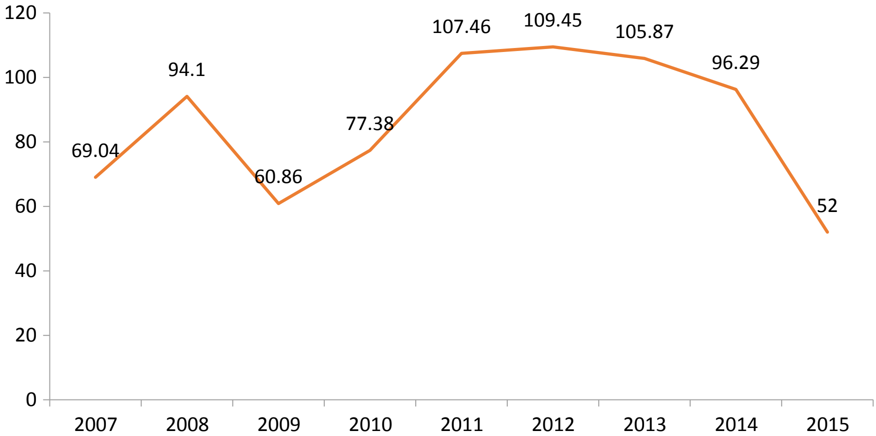


Fig. 3 International crude oil price—average OPEC crude price (in USD/bbl). *Source* http://www.ft.com/cms/s/0/517b2aa4-0acd-11e5-9df4-00144feabdc0.html#axzz3sCckeV74

Marine (Qatar), Arab Light (Saudi Arabia), Murban (UAE) and Merey (Venezuela).[4] International crude price fluctuations are presented in the following figure (Fig. 3).

During of period of study, crude oil prices fell from US$69.04 (2007) per barrel to US$52 per barrel (2015), a decrease of 25%. However, there have been great fluctuations (standard deviation is 22). The difference between the peak rates of US$109.45 to the barrel (2012) to US$52 to the barrel (2015) is a decline of 52%.

[4] http://www.opec.org/opec_web/en/data_graphs/40.htm.

3 Pricing of Petroleum Products in India

The pricing of petroleum products in India can be clearly classified into the (i) administered pricing mechanism (APM) era and (ii) the free market pricing era.

3.1 APM and Regulated Market—Salient Features

The earlier system of pricing of petroleum products was based upon the recommendations of Oil Price Committee 1976 and as amended by the Oil Cost Review Committee in 1984. Under APM regime the pricing of the petroleum products was based on 'retention' concept by which refiners as well as marketing & distribution companies were allowed to retain out of the sales proceeds- Cost of crude, Refining cost and Reasonable return on investment.

The basic objectives for the introduction of government-controlled price mechanism included optimization of the utilization of refining and marketing infrastructure by treating the facilities of all the oil companies as common industry infrastructure, the access of which would be available to all the oil companies by hospitality arrangements, thus avoiding wasteful duplication of investment. All products should be made available at uniform price ex- all refiners so as to minimize cross haulage of products and associated energy costs. It ensured continuous availability of products/crude to refiners by recognizing import needs wherever there are deficits in indigenous production and returns to the refiners were reasonable. Stable prices were ensured by insulating domestic market from the volatility of prices in international market. Socio-economic objectives of the government were achieved by ensuring availability of certain products at subsidized rates for weaker sections of the society and priority sector sin the industry through cross subsidization of products.

The basic principles on which the edifice of APM was built are summarized as follows. Raw materials were made available at a predetermined fixed price at the manufacturing point (delivery cost of crude) on a sustained and continuous basis to the refiners. Finished products were also made available at the marketing companies at predetermined prices (ex-refinery prices). Refining/conversion/marketing costs were reimbursed and post-tax 12% assured return on net worth, and compensation for investments in fixed assets was ensured. Rewards and penalties were built into the system to encourage efficiency.

From 1 April 2002, the APM was dismantled and India moved towards the free market pricing regime.

3.2 Free Market Pricing

With the ushering in the age of liberalization, the petroleum sector has undergone some very fundamental changes and desperately felt the need to be opened up to the global market. In the APM regime, right from crude procurement stringent controls was present till the final distribution of finished products. To be fair to the policy-makers, the current controls did succeed in achieving the macro-economic objectives of the government for more than two decades. However, the experts and the policy-makers felt that APM might no longer sustain the growth of oil industry in the changing environment of liberalized economy as the energy security of the country would be under serious threat if a robust industry along with competitive playing field had not been created. The factors that were fundamentally instrumental behind the gradual transformation of Indian oil sector from regulated to deregulated regime through a nebulous phase of 'pseudo-deregulation' have been discussed here. Inefficient running of the refineries due to an assured return of the capital employed, there was no incentive for minimizing cost and improving productivity. Cross subsidization has led to distortion in demand and prices; heavy subsidization of some of the products has resulted in excessive wastage and adulteration of petro-fuels. Gradual increase in the import of crude oil was due to the stagnation of domestic crude production and slowing down of exploration and production activities due to dearth of adequate investment. Increasing rate of crude import produced hazardous effect on macro-economic stabilization, especially on exchange rate and inflation.

The concept of APM has outlived utility as it had created large oligopolies with no guarantee that the assets are being put to the most optimal use. The cost-plus formula became inoperative with the new private sector entrants in the country over whom the government has no effective control. Capital generation in the upstream and downstream sector had been affected by the artificially low crude prices paid to the exploration and production companies. Key differences between APM and non APM regime are illustrated in Table 2.

3.3 Rangarajan Committee Suggestions

The Government on 26 October 2005 had set up a committee to look into the various aspects of pricing and taxation of petroleum products with a view to stabilizing/rationalizing their prices, keeping in view the financial position of the oil companies, conserving petroleum products and establishing a transparent mechanism for autonomous adjustment of prices by the oil companies. Based on the deliberations in the meetings, three areas were identified by the committee for detailed study in order to meet the objectives set out in the terms of reference. These areas included alternative models for pricing of petroleum products; taxes and duties on crude oil and petroleum products; and subsidies on Public Distribution System (PDS) kerosene and domestic LPG.

Table 2 Key differences between APM and non-APM regime

APM regime	Non-APM regime
Administration of oil pricing mechanism through oil pool account	Decontrol of all oil products except diesel, LPG, petrol and ATF initially
Planning and recommending crude and product import	Market-based price determination of decontrolled products. Rationalization of import duties on crude and petroleum products. Payment of higher percentage of price of imported crude to local crude producers
Planning and recommending indigenous and imported crude oil distribution to different refineries	Abolishment of cost-plus formula and assured return to the oil companies
Fixing the market price and retention price of various products	Decanalization of imports/exports of all products except crude, aviation turbine fuel, motor spirit and high-speed diesel
Fixing profit margins to be earned by exploration and production, refining and marketing companies	Sourcing of crude permitted for private/joint sector refineries under user licenses

Some important recommendations made in this report including shifting to a trade parity pricing formula for determining refinery gate as well as retail prices. Government was to keep herself at arms' length from price determination and to allow flexibility to oil companies to fix the retail price under the proposed formula. Effective protection was reduced by lowering the customs duty on petrol and diesel to 7.5%. This set of recommendations should be implemented as an integrated package as selective implementation would create more distortions.

The second set of recommendations relates to pricing of domestic LPG and PDS kerosene, such as restricting subsidized kerosene to BPL families only; increasing the price of domestic LPG by INR 75/cylinder; discontinuing the practice of asking ONGC/GAIL/OIL to provide upstream assistance, but instead collecting their contribution by raising the Oil Industry Development Board (OIDB) cess from the present level of INR 1800/MT to INR 4800/MT and Government meeting the balance cost of subsidy from the budget. The 'PDS Kerosene and Domestic LPG Scheme 2002' had to be suitably amended for this purpose. Similarly, this set of recommendations too should also be implemented as an integrated package as partial implementation will not yield sustainable results.

The third set of recommendations relates to restructuring excise duties from the present mix of specific and ad valorem to a pure specific levy and calibrating the levies at INR 5.00/l of diesel and INR 14.75/l of petrol.

Based on all these regulatory changes, the prices of LPG, petrol and diesel are determined in India as discussed in the following subsections.

Table 3 Price build-up of LPG (as of 2014–15) (INR/14.2 kg cylinder)

Price head	Delhi	Kolkata	Mumbai	Chennai
Current ex-storage point price (ESPP)	340.56	340.56	340.56	340.56
Increase over last year	50	50	50	50
Basic price	340.56	340.56	340.56	340.56
State surcharge	0	0	25.45	0
Freight up to bottling plant	1.42	2.27	0	0
Price adjustment factor (PAF)[a]	21.43	22.27	21.88	10.16
Assessable value	**363.41**	**365.10**	**387.89**	**350.72**
Excise duty	0	0	0	0
Octroi at bottling plant	0	0	0	0
Price at bottling plant	**363.41**	**365.10**	**387.89**	**350.72**
State subsidy	0	0	0	0
Freight/local delivery charges	10	10	10	10
Octroi/transit tax	0	0.17	0	0
Price including octroy	**373.410**	**375.27**	**397.89**	**360.72**
Sales tax (ST)	0	0	11.94	0
Surcharge on ST	0	0	0	0
Billable rate (or Sub total)	**373.41**	**375.27**	**409.83**	**360.72**
Price to dealer	373.41	375.27	409.83	360.72
Distributor commission	40.71	40.71	40.71	40.71
Price after distributor commission	**414.12**	**415.98**	**450.54**	**401.43**
VAT (at applicable rates in respective states)	0	0	1.22	0
Retail selling price (RSP) of 14.2 kg cylinder	414.12	415.98	451.76	401.43
RSP rounded off	414	416	451.5	401

[a]PAF is included to adjust prices closer to international prices

(a) **Pricing of LPG**

LPG prices vary in India from state to state, based on the taxation structure of each state government. For the purpose of this study, LPG price build-up in Delhi, Kolkata, Mumbai and Chennai (as of 2014) has been illustrated. However for the sake of analysis, the price of Delhi has been considered (Table 3).

3.4 *Pricing of Petrol and Diesel*

In case of petrol, the price build-up for Delhi has been illustrated in Table 4. In case of diesel, the price build-up for Delhi has been illustrated in Table 5.

Table 4 Price build-up of petrol (as of 2014–15)

No.	Elements	Unit	Figure
1	Cost and freight price of gasoline (petrol) BS III equivalent	$/bbl	61.91
2	Average exchange rate	Rs/$	66.08
3	Refinery transfer price (RTP) on landed cost basis for BS-IV petrol (price paid by the oil marketing companies to refineries)	Rs/l	26.28
4	Price charged to dealers (excluding excise duty and VAT)	Rs/l	29.27
5	Add: Specific excise duty @ Rs 17.46/l	Rs/l	17.46
6	Add: Dealer commission	Rs/l	2.27
7	Add: VAT (including VAT on dealer commission) applicable for Delhi @ 25%	Rs/l	12.25
8	Retail selling price at Delhi (rounded off)	Rs/l	61.25

Source http://www.hpretail.in/documents/pdf/pb/pricebuildup_Petrol.pdf

Table 5 Price build-up of diesel (as of 2014–15)

No.	Elements	Unit	Figure
1	Cost and freight price of gasoil (diesel) BS III equivalent	$/bbl	60.30
2	Average exchange rate	Rs/$	66.08
3	Refinery transfer price (RTP) on landed cost basis for BS-IV diesel (price paid by the oil marketing companies to refineries)	Rs/l	25.75
4	Price charged to dealers (excluding excise duty and VAT)	Rs/l	26.63
5	Add: Specific excise duty @ Rs 10.26/l	Rs/l	10.26
6	Add: Dealer commission	Rs/l	1.43
7	Add: VAT (including VAT on dealer commission) applicable for Delhi @ 16.6% and air ambience charges @ Rs 0.25/l	Rs/l	6.65
8	Retail selling price at Delhi (rounded off)	Rs/l	44.97

Source http://www.hpretail.in/documents/pdf/pb/pricebuildup_Diesel.pdf

3.5 Subsidies on the Three Fossil Fuels

The non-subsidized and subsidized prices of the three fuels, i.e. LPG, petrol and diesel, have been analysed for the last 10 years. From these two price sets, the subsidy trend of these fuels has been computed. The subsidy on petrol was completely removed in 2010 while the subsidy on diesel was removed in 2013. LPG subsidy has reduced from as high as 76–26% in these 10 years. A comparison of subsidized price and non-subsidized price of these fuels over the last 10 years has been presented in Table 6.

Table 6 Subsidy rates of selected fossil fuels for the last 10 years

Year	Subsidized retail price			Non-subsidized price			Net subsidy %		
	LPG (INR/14.2 kg cylinder)	Petrol (INR/l)	Diesel (INR/l)	LPG (INR/14.2 kg cylinder)	Petrol (INR/l)	Diesel (INR/l)	LPG	Petrol	Diesel
2004	294.75	33.7	21.7	1250	38	32	76	11	32
2005	294.75	37.8	26.5	1250	40	35	76	6	24
2006	294.75	43.5	30.5	1250	45	37	76	3	18
2007	306.3	44.5	31.3	1079	46	38	72	3	18
2008	279.7	43.5	30.8	994	45	35	72	3	12
2009	281.2	40.6	30.9	905	41	35	69	1	12
2010	399.26	44.7	32.9	880	44.7	35	55	0	6
2011	399	58.4	37.8	865	58.4	40	54	0	6
2012	410.5	65.5	40.9	626	65.5	43	34	0	5
2013	414	67.3	47.7	559.5	67.3	47.7	26	0	0

4 Electricity Pricing

A major catalyst for growth in developing countries is electricity or more specifically the availability and accessibility of reliable electricity. Any country's growth is driven by the availability of reliable and cheap electricity. It also allows industries to take advantage of new and innovative productivity-enhancing technologies, the bulk of which are reliant on electricity. This in turn improves efficiency of industries and improves their cost competitiveness. In the developed world, industrial users generally pay lower prices for electricity compared to other users (IEA 2012). These lower prices indicate the lower cost of supplying power to industrial users which is an outcome of their more stable power demand patterns and their ability to use electricity at higher voltages without the power utility incurring extra cost of stepping down the voltage. However, in developing countries including India the existence of cross-subsidies reduces electricity prices for domestic and agricultural users at the expense of industrial users.

Cross-subsidies in electricity tariff can be defined as 'a mechanism whereby some consumer groups are charged a higher tariff as compared to the cost of supplying power to them. The additional revenue generated from them is used to tide over the revenue shortfall from other consumer groups, who are charged lesser tariff as compared to the cost of supplying power to them'. In India, cross-subsidies are targeted at consumer groups who either do not have enough paying capacity or need to be supported for undertaking economic activities such as agriculture, which in some way benefit the larger sections of society. In case of cross-subsidies, subsidization is inbuilt in the electricity tariff, unlike any external support which is provided in the case of direct fossil fuel subsidy. Cross-subsidy is a matter of tariff design, which can be adjusted depending on the intended level of cross subsidization. There are many detrimental effects of cross-subsidies. It leads to wastage of economic resources. In India, lower electricity tariffs in the agriculture sector lead to indiscriminate pumping of groundwater by farmers, which wastes not only power but also water. On the other hand, higher power tariffs (than the cost of power supply) charged to industries increase their cost of production rendering them uncompetitive in today's globalized world. Cross-subsidies also lead to revenue loss for state utilities as industries scale up their 'captive power generation' to avoid grid power.

The industrial sector contributes to about a third of India's GDP and constitutes a critical pillar of the Indian economy and society, and will continue to be so in the foreseeable future. As discussed above, electricity prices are one of the major factors determining the profitability of industries as well as the cost competitiveness of both domestic goods and export goods. Rise in electricity tariffs in India in the past few years has been of key concern to most industrial consumers. Since 2007, electricity prices increased from an average of INR 4.16/kWh to an average of INR 7.64/kWh in 2015—almost 47% change in 8 years. Typically countries where electricity prices have been very low have emerged as preferred export hubs. In 2011–12, in Asia, such low-cost exporters included Bangladesh with average electricity tariff of USD 0.063/kWh followed by Vietnam (USD 0.063/kWh), India (USD 0.064/kWh), Thailand (USD 0.07/kWh) and Indonesia (USD 0.08/kWh).[5] Compared to the industrial consumers, agricultural consumers in India paid a tariff of INR 0.77/kWh in 2007 and INR 1.83/kWh in 2015. On an average, Indian industries pay about 12% higher cost while agricultural consumers pay about 55% lower than the average cost of supply of power.[6] As a result of the high cost of electricity, firms may reduce their electricity usage, which has implications for the types of technologies they are able to use and ultimately, their growth.

There are several factors which drive electricity tariffs in India depending on the mechanism of interaction between selling and buying power. The Indian electricity market predominantly follows a wholesale decentralized model. In this model, the generation, transmission and distribution companies are unbundled. Multiple generators, including Independent power producers and public sector generation companies, are allowed to participate in the supply of electricity. This ensures supply security and removes monopoly on the prices. The generators are allowed to openly compete, which enables system operators to schedule and dispatch the power based on the different contracted prices and also the distribution companies to procure power at competitive prices. However, in this model, the choice available to all the retailers and consumers to procure power in the open market besides their distribution companies is restricted. To ensure more choices to consumers, the Electricity Act, 2003, provides for open access for private generators and bulk consumers. Open access is available for consumers above 1 MW of requirement. Open access is a mechanism that allows generators to sell power to the highest bidders while consumers can source their needs from the most economic seller. There is also a power exchange available in the country, which allows the consumers to bypass distribution companies and procure power at the spot market. Power exchange has been introduced to offer a nationwide voluntary access, e-trading, no counter-party risk, robust platforms and delivery-based contracts. However, due to volatility/uncertainty in prices and requirement of higher liquidity; the percentage of power traded in the exchange is very low in the country. The above discussion summarizes broadly a market model followed in the country. The power can either be directly sold by the generation companies to the distribution companies or through an intermediary, i.e. an independent body who can purchase power in bulk. However, there can be some change in orientation of the above model from state to state depending on state-specific regulations.

The electricity tariff-setting mechanism in India is formulated by Central Electricity Regulatory Commission (CERC) for generating companies owned or controlled by the Central Government or for generating companies who enter into or otherwise have a composite scheme for generation and sale of electricity in more than one state. The State Electricity Regulatory Commissions (SERCs) determines tariff for generation, supply transmission and wheeling of electricity within the individual states. The Tariff Policy, 2006, is the main basis for pricing of electricity while the National Electricity Policy, 2005, lays down that the amount of cross-subsidy surcharge and

[5] http://talkenergy.files.wordpress.com/2011/02/asean-electricity-tariff-2011.pdf.

[6] Except the northeastern states (with the exception of Assam) where industrial power tariff is also significantly subsidized.

the additional surcharge to be levied from consumers. The extent of subsidy for different categories of consumers can be decided by the state government keeping in view various relevant aspects.

The electricity tariff in India essentially has three components. Firstly, there is a fixed charge payable every month by each beneficiary to the generator for making capacity available for use. The fixed charge is not the same for each beneficiary. It varies with the share of a beneficiary in a generators' capacity. The fixed charge, payable by each beneficiary, will also vary with the level of availability achieved by a generator. Secondly the energy charge (defined as per the prevailing operational cost norms) is payable per kWh of energy supplied as per a pre-committed schedule of supply drawn upon a daily basis. Thirdly, there is a charge for Unscheduled Interchange (UI) charge for the supply and consumption of energy in variation from the pre-committed daily schedule. This charge varies inversely with the system frequency prevailing at the time of supply/consumption. Hence, it reflects the marginal value of energy at the time of supply.

5 Concluding Comments

It should be mentioned that there was a secular decline in subsidies in all kinds of fuels like LPG, petrol and diesel over the last decade in India. The only difference was the pace of reduction of subsidies and whether subsidies were reduced to zero or not. The pace was highest for petrol, followed by diesel. In both these cases, subsidies were reduced to zero. The pace was slowest for LPG and subsidies until now has only been reduced by about 50%. The nature of removal of subsidies was completely different for electricity. Here, the agricultural sector was cross-subsidized by the industries. Thus, the question of removal of subsidies does not arise here. Rather the question was whether cross-subsidies were reduced and if so to what extent.

Since these three types of fuels together account for about 6% of India's GDP and since fuel subsidies were removed at a time when the cost of crude was on the decline, it is expected that the overall impact of fuel subsidy removal on the macroeconomic parameters will be marginal. Also, the effect of the decline in subsidy was thwarted to a large extent by the falling crude price. For this reason, it is important to distinguish clearly between the three following cases: (i) change in crude price only, (ii) change in subsidies only and (iii) the two together. For the years 2008–09, 2012–13, 2013–14 and 2014–15, both subsidies and crude oil prices fell while for the rest of the years, subsidies fell and crude oil prices rose.

Duties on fuel have been a big source of revenue for both the centre and states. Taking advantage of the inelastic nature of their demand, energy products, especially petroleum, have been leveraged by the government as a cash cow in financing India's budget deficit with scant regard for consumers.

References

Abouleinein, S., El-Laithy, H., & Kheir-El-Din, H. (2009). *The impact of phasing out subsidies of petroleum energy products in Egypt*. Egyptian Center for Economic Studies, Working Paper No. 145.

Argus Media Website. *Various reports*. Retrieved from http://www.argusmedia.com/. Accessed on July 15, 2015.

Census of India's. (2011). *Provisional population totals of rural-urban distribution in the country, released by Union Home Secretary R.K. Singh*. Retrieved from http://censusindia.gov.in/2011-prov-results/paper2/data_files/india/Rural_Urban_2011.pdf. Accessed on December 02, 2014.

Danielsen, A. (1982). The Evolution of OPEC. New York: Harcourt Brace Jovanovich.

Essama-Nssah, B., Delfin, S. G., Kearney, M., Korman, V., Robinson, S., & Thierfelder, K. (2008). *Economy wide and distributional impact of an oil price shock on the South African economy, The world bank and institute of development studies*. Retrieved from http://siteresources.worldbank.org/INTPSIA/Resources/490023-1171551075650/3457768-1241469484668/AFRI_485-530_ch12.pdf.

Financial Times. Retrieved from http://www.ft.com/cms/s/0/517b2aa4-0acd-11e5-9df4-00144feabdc0.html#axzz3sCckeV74. Accessed on July 05, 2014.

Grave, K., Hazrat, M., Boeve, S., von Blücher, F., Bourgault, C., Breitschopf, et al. (2015). *Electricity costs of energy intensive industries—An international comparison, for the German ministry of economic affairs and energy*. Retrieved from http://www.ecofys.com/files/files/ecofys-fraunhoferisi-2015-electricity-costs-of-energy-intensive-industries.pdf. Accessed on December 02, 2014.

Handbook of Statistics on Indian Economy. *RBI*. Retrieved from https://www.rbi.org.in/scripts/annualPublications.aspx?head=Handbook+of+Statistics+on+Indian+Economy. Accessed on December 22, 2015.

Handbook on Direct Benefit Transfer for LPG Consumer (DBTL). *Ministry of petroleum and natural gas, government of India*. Retrieved from http://petroleum.nic.in/dbt/DBTL_Handbook.pdf. Accessed on December 02, 2014.

Hindustan Petroleum Corporation Limited Website. Retrieved from http://www.hpretail.in/documents/pdf/pb/pricebuildup_Diesel.pdf and http://www.hpretail.in/documents/pdf/pb/pricebuildup_Petrol.pdf. Accessed on December 02, 2014.

IEA. *World energy outlook 2011*. Retrieved from http://www.worldenergyoutlook.org/weo2011/. Accessed on November 25, 2013.

IEA. *World energy outlook 2012*. Retrieved from http://www.worldenergyoutlook.org/weo2012/. Accessed on June 20, 2014.

Indexmundi Statistics. Retrieved from http://www.indexmundi.com/. Accessed on March 22, 2016.

India in Figures. (2015). *Government of India, ministry of statistics and programme implementation*. Central Statistics Office, Research & Publication Unit, New Delhi. Retrieved from http://mospi.nic.in/Mospi_New/upload/India_in_figures-2015.pdf. Accessed on December 20, 2015.

Indian Express Articles. Retrieved from http://indianexpress.com/article/business/commodities/fm-time-to-remove-lpg-subsidy-for-people-like-us/. Accessed on December 02, 2015.

Indiastat Website. *Various documents and reports*. Retrieved from www.indiastat.com. Accessed on July 18, 2015.

Inter-governmental Panel on Climate Change. *Various documents and reports*. Retrieved from www.ipcc.ch. Accessed on July 18, 2014.

Labour Bureau. *Government of India*. Retrieved from http://labourbureau.nic.in/indtab.html. Accessed on December 22, 2015.

Mabro, R. (1984). *On oil price concepts, oxford institute for energy studies, WPM 3*. Retrived from https://www.oxfordenergy.org/wpcms/wp-content/uploads/2010/11/WPM3-OnOilPriceConcepts-RMabro-1984.pdf . Accessed on July 18, 2014.

Ministry of Power. *Government of India*. Retrieved from http://powermin.nic.in/upload/loksabhatable/pdf/LS19032015_Eng.pdf. Accessed on July 18, 2014.

Moneycontrol Website. Retrieved from http://m.moneycontrol.com/news/economy/need-to-dealdiscom-losses-onstate-level-expert_1506161.html. Accessed on July 18, 2014.
Moorthy, V., & Kolhar, S. (2011). Rising food inflation and India's monetary policy. *Indian Growth and Development Review, 4*(1), 73–94.
NSS 66th Round. *Report No. 537: Employment and unemployment situation in India, 2009–10*. Retrieved from http://mospi.nic.in/mospi_new/upload/nss_report_552.pdf. Accessed on December 02, 2014.
OPEC Website-various Reports. Retrieved from http://www.opec.org/opec_web/static_files_project/media/downloads/publications/MOMR%20February%202016.pdfand http://www.opec.org/opec_web/en/data_graphs/40.htm. Accessed on July 05, 2015.
Soni, A. (2014). *Global oil markets and India's vulnerability to oil shocks, the energy and resources institute TERI-NFA working paper series No. 18*. Retrieved from http://www.teriin.org/projects/nfa/pdf/working-paper-No18-Oil-volatility.pdf. Accessed on July 18, 2014.
Sriring, O. (2011). *Thailand may remove oil subsidy as fund runs out*. Retrieved from Reuter's website http://in.reuters.com/article/2011/03/04/thailand-economy-subsidy-idINSGE72302F20110304. Accessed on October 23, 2012.
Swain, A., & Charnoz, O. (2012). *In pursuit of energy efficiency in India's agriculture: Fighting 'free power' or working with it? AgenceFrançaise de Développement,* working paper 126. Retrieved from http://www.afd.fr/webdav/shared/PUBLICATIONS/RECHERCHE/Scientifiques/Documents-de-travail/126-document-travail-VA.pdf. Accessed on July 18,2014.
Talkenergy Website of WordPress.com. Retrieved from https://talkenergy.files.wordpress.com/2011/02/asean-electricity-tariff-2011.pdf. Accessed on July 18, 2014.
The Final Report of the Expert Group on Low Carbon Strategies for Inclusive Growth. *Planning Commission, Government of India, April 2014*. Retrieved from http://planningcommission.nic.in/reports/genrep/rep_carbon2005.pdf. Accessed on July 18, 2014.
The National Newspaper Article. Retrieved from http://www.thenational.ae/business/economy/energy-subsidies-prove-drain-on-indian-economy. Accessed on July 20, 2015.
US Environment Protection Agency. *Various documents and reports*. Retrieved from http://www.epa.gov. Accessed on July 18, 2014.
United Nations Conference on Trade & Development, Various reports accessed at http://unctad.org/en/Pages/Statistics.aspx. Accessed on July 18, 2015.
World Economic Outlook 2007 published by IMF. Retrieved from https://www.imf.org/external/pubs/ft/weo/2007/02/pdf/c4.pdf. Accessed on December 02, 2014.

International Joint Ventures in Developing Countries: The Implications of Policy Uncertainty and Information Asymmetry

Tarun Kabiraj and Arijit Mukherjee

Abstract An important way through which many developing countries get foreign investments and foreign technologies is international joint ventures (IJVs). The literature paid significant attention to analyze the rationale and consequences of IJVs. This chapter reviews the literature on IJVs in developing countries and discusses the implications of government policy and information asymmetry.

1 Introduction

Joint venture (JV) as a form of business is observed worldwide. Under JV, two or more legally independent firms agree on a project to share jointly business risk, returns, and control, but the parent firms retain their identities. When in a JV at least one partner has its parent organization headquartered outside the JV's country of operation, it is called international joint venture (IJV).

Our particular interest in this chapter lies in the IJVs formed by the multinational corporations and the firms from developing countries. Since the 1980s, many developing countries started to liberalize their economies to attract foreign direct investments and technologies. Since then, the formation of IJVs between domestic companies and foreign multinational corporations became popular, as IJVs provide flexible mechanisms that allow the domestic and foreign companies to form mutually gainful business entities. As many developing countries have been growing at a faster rate than the industrialized countries and the trend is likely to continue, the interest

T. Kabiraj
Indian Statistical Institute, Kolkata, India

A. Mukherjee (✉)
Nottingham University Business School, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, UK
e-mail: arijit.mukherjee@nottingham.ac.uk

CESifo, Munich, Germany

INFER, Cologne, Germany

GRU, City University of Hong Kong, Hong Kong, China

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_15

of the foreign multinationals in these economies is increasing and direct investment flows have been accelerating. For example, as mentioned in Beamish (1988), IJVs grew by 59% between 1981 and 1983. Korbin (1988) showed that in some less developed countries, nearly 80% of total foreign capital entered through JVs. In China, out of one hundred seventy-five thousand foreign investment projects approved during 1979 and 1993, about 75% took the form of a joint venture between a domestic firm and a foreign multinational.[1] In India, in the immediate post-liberalization period, most of the foreign investments have come through JVs (Gupta and Chawla 1995). Although after 1995 the trend of JV formation slowed down a little bit as the foreign firms find some other modes of entry, such as fully owned subsidiaries and mergers, more attractive, the spree is still going on. A comprehensive study on IJVs for a period 1990–2000 can be found in Moskalev and Swensen (2007).

The importance of IJVs as a business form and its growing trend have encouraged many researchers to look at the rationale for IJVs and on their implications. In this chapter, we consider some analytical issues related to IJVs and focus on two important factors, viz. policy uncertainty and information asymmetry. In particular, we argue that policy uncertainty and incomplete information on either side of the firms can be important factors for the formation of IJVs. At the initial phase of liberalization and opening up, there exist many uncertainties about the stability and commitment of the policies of the developing countries, and similarly the behaviors and the strategies of the JV partners may be unknown to each other.

The remainder of the chapter is organized as follows: Section 2 focuses on the effects of policy uncertainties, and Sect. 3 shows the implications of information asymmetry. Section 4 provides some potential areas for future research.

2 Policy Uncertainties

One dominant character of many developing countries is policy uncertainty, which makes many foreign firms cautious about their investments in developing countries. We discuss in this section how a suitably designed JV helps to reduce the risk of policy uncertainty or policy moral hazard.

Marjit (1990) shows that a public–private JV helps to overcome the threat of expropriation. Assume that a foreign firm wants to invest in a developing country. This project lives for infinite periods, and the domestic government charges an optimal lump-sum tax per period, T^*. Therefore, the discounted lifetime earning of the government from the tax revenue is $\frac{T^*}{(1-\delta)}$, where $\delta \in (0, 1)$ is the discount factor. Assume that the foreign firm's optimal amount of invest is I^0. Further, assume that the salvage value of investment is greater than the earning from tax revenue, i.e., $I^0 > \frac{T^*}{(1-\delta)}$, which gives the domestic government the incentive to expropriate foreign investment. If there are no other costs of expropriation, such as loss of repu-

[1] See Almanac of China's Foreign Relations and Trade (1994).

tation, the foreign firm anticipates that the domestic government will expropriate its investment. This prevents the foreign firm from investing in the domestic market.

Now, consider a JV between the domestic government and the foreign firm where the domestic government shares $(1-\theta)$ fraction of the project's investment. Hence, in the public–private JV, the foreign firm invests θ fraction of the investment, i.e., θI^0. If the government expropriates foreign investment, its gain from expropriation following the JV is θI^0. Thus, one can choose θ suitably to satisfy $\theta I^0 < \frac{T^*}{(1-\delta)}$, which will encourage the foreign firm to invest in the country.

It is important to note that the government cannot reverse the inequality $I^0 > \frac{T^*}{(1-\delta)}$ by charging more as tax revenue because T^* is the maximum amount that the government can charge. The government will set T^* in a way so that the foreign firm gets its reservation payoff only. In particular, $T^* = R(I^0) - I^0(1-\delta) - \bar{\pi}(1-\delta)$, where $R(I^0)$ is the project's gross payoff, $\bar{\pi}$ is the reservation payoff of the foreign firm, and I^0 is the optimal amount of investment. Hence, a higher tax than T^* will induce the foreign firm not to invest in the country. Thus, a higher tax than T^* cannot be a solution to the problem.

The public–private JV provides an answer to the problem, although it does not give the foreign firm the incentive to bring its technology embodied in the investment. If the foreign firm brings its technology embodied in the investment, the discounted payoff of the domestic government after expropriation will be $\frac{R(I^0)}{(1-\delta)}$, and the host government will have the incentive to expropriate foreign investment under the JV. Hence, the JV with no transfer of technology acts as an insurance for the foreign firm.

Proposition 1 (Marjit 1990) *A suitably formed public-private JV can reduce the expropriation threat and attract foreign investment. However, foreign firm will not bring its technology embodied in the investment.*

Abdalla (1992) considers joint production by a foreign firm and a domestic firm and shows the implications of the expropriation threat on the choice of inputs in the JV. For example, consider that firms 1 and 2 form a JV providing e_1 and e_2 non-contractible inputs, respectively. Therefore, firm i's ($i = 1, 2$) problem is to determine the optimal amount of input supply to maximize its profit, i.e.,

$$\underset{e_i}{\text{Max}}\, \alpha_i R_i(e_1, e_2) - c_i(e_i), \quad i = 1, 2 \tag{1}$$

where $c_i(e_i)$ is the cost of providing input e_i by firm i and α_i shows the share of the ith firm in the JV with $\sum_1^2 \alpha_i = 1$. In this framework, Abdalla (1992) shows that the threat of expropriation induces the foreign firm to underinvest in the input it supplies, which leads to the following result.

Proposition 2 (Abdalla 1992) *The firm that faces the expropriation threat underinvests in the input it supplies and provides higher lump-sum payment to the domestic firm compared to the situation with no expropriation threat.*

Although the threat of expropriation might be remote in today's world, there might exist other types of policy uncertainties, such as uncertainties in tax and tariff policies.

Broll and Marjit (1995) provide a rationale for a public–private JV when a domestic government finds it difficult to commit to the tax policy. Assume that a foreign firm invests I and gets a gross profit $R(I)$, $R' > 0$, $R'' < 0$. The government charges a tax on $R(I)$ at a rate $t < 1$. Consider that I^t is the optimal amount of investment in the non-cooperative regime; i.e., I^t maximizes $(1 - t)R(I) - I$.

In the case of a public–private JV, assume that the government shares $(1 - \theta_2)$ and $(1 - \theta_1)$ fraction of investment and gross profit, respectively. Hence, the foreign firm invests θ_2 fraction of investment and gets θ_1 fraction of the gross profit. Consider $\theta_1 = \frac{\theta_2}{(1-t)}$. Since $t < 1$, the foreign firm's investment share is less than its share in the gross profit. Further, $\theta_1 = \frac{\theta_2}{(1-t)}$ shows that the foreign firm's loss in gross profit due to tax is balanced by its gain from lower investment share. Hence, it helps to eliminate the distortionary effect of taxation and induces the foreign firm to invest more than I^t. The foreign firm will actually invest up to I^0 that maximizes $R(I) - I$.

Now, consider a situation where the tax rate follows a distribution $t \in [t_a, 1]$ with a density $f(t)$, where t_a is the announced tax rate but the foreign firm apprehends that the domestic government may raise its tax in future. They show that if the foreign firm is risk-neutral, a public–private JV with $\theta_1 = \frac{\theta_2}{(1-\bar{t})}$ will be preferable to a non-cooperative situation, where $\bar{t}$ is the expected tax rate.

Proposition 3 (Broll and Marjit 1995) *A public-private JV Pareto dominates the non-cooperative situation whether or not the government can commit to the tax rate.*

Das (1998, 1999) explains how the foreign firms can alter their mode of entry when they expect policy reversal. Das (1998) considers a situation where the project needs two inputs—one of them is specific to the foreign firm and cannot be contracted at arm's length, but the other input can be produced by a domestic firm more efficiently than the foreign firm. Without any other problem, there is no reason to form a JV between the foreign firm and the domestic firm. The foreign firm can open its fully owned subsidiary in the host country and make a contract with the domestic firm for the local input.

Now, assume that the government imposes a production tax and the tax rate is determined after the entry decision of the foreign firm, implying that the foreign firm faces policy uncertainty when taking the investment decision. In this situation, the foreign firm has the incentive to form a JV with the domestic firm.

Proposition 4 (Das 1998) *JV between a foreign firm and a domestic firm is a better option than the foreign firm's fully owned subsidiary in the presence of policy uncertainty.*

An increase in the equity share of the domestic firm reduces the incentive for rent extraction by the domestic government and creates the incentive for making a favorable policy toward the venture. Thus, equity participation by the domestic firm helps to reduce the effect of policy moral hazard. Although equity sharing by the domestic firm adversely affects the foreign firm's input choice, the gain from the reduction of policy moral hazard makes the IJV as a better option than the foreign firm's fully owned subsidiary.

Das (1999) focuses on the role played by risk aversion when a foreign firm has the option to license technology, forming JV and opening a fully owned subsidiary. He considers a scenario where the foreign firm has the basic technology to produce a good, and it is deciding to either open a production facility in the domestic country (i.e., JV or fully owned subsidiary) or license the technology to a domestic firm. Production requires a country-specific knowledge-based input (e.g., managerial input) that cannot be traded at arm's length, and the domestic firm can provide this input more efficiently than the foreign firm. Further, assume that the domestic firm is more risk-averse than the foreign firm. While the domestic firm's participation makes the project more cost-efficient, the domestic firm needs more compensation due to higher risk aversion. Hence, the foreign firm prefers a JV, provided the project is not too risky. For a highly risky project, the domestic firm needs a larger compensation and the participation of the domestic firm will not be profitable. However, for a less risky project, the domestic firm does not need larger compensation but it makes the project more cost-efficient. Hence, for a less risky project, the JV becomes an attractive mode of operation. Technology licensing by the foreign firm is dominated by either JV or fully owned subsidiary of the foreign firm.

Now, consider that there is also the possibility of government policy intervention after the entry decision of the foreign firm. The policy moral hazard provides further incentive for the domestic firm's participation in the project. It turns out that if the project has very low risk, the foreign firm prefers no equity holding in the project and licenses its technology to the domestic firm. Licensing with an up-front fixed fee implies that, after the initial payment made by the domestic firm, it will retain the entire profit from the project, thus creating no incentive for government intervention. Hence, if the project has low risk, licensing with an up-front fixed fee is the optimal entry strategy for the foreign firm because it reduces the policy moral hazard to a minimum as well as it does not ask for a larger compensation to the domestic firm. Fully owned subsidiary is preferred when the project has high risk, but a JV is preferred for moderate risk.

Proposition 5 (Das 1999) *Assume that a foreign firm is less risk averse than a domestic firm.*

(a) *With no policy moral hazard, JV (fully owned subsidiary of the foreign firm) is the best option if the market-size adjusted risk is high. Either a JV or a fully owned subsidiary dominates technology licensing.*
(b) *With policy moral hazard, technology licensing may be the best option. Technology licensing fully owned subsidiary is the best choice if the market-size adjusted risk is too low too high. A JV is preferred for moderate market-size adjusted risk.*

Al-Saadon and Das (1996) and Mukherjee (2000) also show that policy uncertainties affect a foreign firm's entry decision, and examine whether a committed or a non-committed domestic policy creates higher domestic welfare.

Consider a JV between a foreign firm and a domestic firm. The domestic government imposes a tax per unit of output. Assume that the firms get only dividend incomes and the profits are distributed to the partners. The technology of the project is given. If the government cannot commit to the tax rate before the equity sharing decision of the firms, i.e., there is policy moral hazard, it induces the domestic firm to hold a larger equity share compared to the situation with no policy moral hazard, i.e., when the government can commit to the tax rate. A higher domestic participation reduces the rent-seeking motive of the domestic government, which, in turn, increases the project's profit. Hence, the policy moral hazard increases the domestic firm's participation in the project, the distortion due to the production tax reduces, and the tax revenue falls compared to the situation where the domestic government can commit to the tax rate before the equity sharing decision of the firms. As a result, the welfare of the domestic country may be higher under policy moral hazard than a committed government regime.

Proposition 6 (Al-Saadon and Das 1996) *Suppose a foreign firm and a domestic firm forms a JV. If there is no possibility of lump-sum transfer and the technology of the project is given, the domestic welfare is higher under a non-committed domestic tax policy compared to a committed domestic tax policy.*

Al-Saadon and Das (1996) do not consider the possibility of lump-sum payment as well as the possibility of technology choice by the foreign firm. Mukherjee (2000) relaxes these assumptions. The impossibility of lump-sum payment reduces the foreign firm's available strategies by eliminating the possibility of technology licensing.

Assume that the foreign firm has the basic technology to produce the good, and it has three options: (1) license its technology to a domestic firm, (2) open its fully owned subsidiary in the domestic country, and (3) form a JV with a domestic firm where the foreign and the domestic firms share equity. Consider that the domestic government imposes a production tax. Further, assume for simplicity that the goods are sold entirely outside the domestic country.[2] If the domestic government can commit to a tax rate before the decision of the foreign firm, the foreign firm has no incentive for licensing or JV. However, if there is policy moral hazard, the domestic government does not commit to its tax rate before the decision of the foreign firm, which creates the incentive for a larger domestic participation in the project. Since the motive for tax revenue comes from rent-seeking by the domestic government, domestic firm's participation in the project unambiguously reduces the incentive for rent-seeking by the domestic government. Thus, it reduces the distortion created by the production tax and increases the project's profit. If the foreign firm holds no equity share in the project, it fully eliminates the rent-seeking incentive of the domestic government and helps to increase the project's profit. The extra profit generated can be shared by the firms according to their bargaining power. If the bargaining power of the foreign firm is sufficiently high, it takes a larger portion of the benefit, thus creating a lower domestic welfare under policy moral hazard compared to a committed government policy. However, with a sufficiently lower bargaining power

[2] The basic result will not change if one assumes that all the goods are sold in the domestic country.

of the foreign firm, the larger portion of the benefit from a lower tax rate remains in the domestic country, which outweighs the loss from tax revenue. Therefore, a sufficiently higher bargaining power of the domestic firm creates a higher domestic welfare under policy moral hazard compared to a committed tax policy.

Often, the foreign firms face important decisions about the quality of the technologies to be brought in the domestic country because the transfer of technologies needs the foreign firms to bear sufficient costs as resource costs or R&D costs (see, e.g., Teece 1977, 1981; Lee and Shy 1992). As discussed above, if the foreign firm has sufficiently higher bargaining power, it gets a higher profit under policy moral hazard compared to a committed tax policy. In this situation, the foreign firm has the incentive to bring a relatively better technology under policy moral hazard compared to a committed tax policy. This efficiency gain creates another incentive for no commitment by the domestic government. It turns out that the range of bargaining power of the foreign firm over which a committed tax policy creates a higher domestic welfare compared to a non-committed tax policy reduces with foreign firm's endogenous technology choice.

Proposition 7 (Mukherjee 2000) *If a foreign firm does not have any restriction on profit extraction, the domestic welfare is higher (lower) under a committed domestic tax policy compared to a non-committed domestic tax policy provided the bargaining power of the foreign firm is sufficiently high (low). If the foreign firm also has the technology as a choice variable, it reduces the range of bargaining power of the foreign firm over which the domestic welfare is higher under a committed domestic tax policy compared to a non-committed domestic tax policy.*

Considering a two-period game, Marjit et al. (2004) explain how uncertainties in the reform process may encourage the formation of an IJV even if the JV gives future benefits that are distinct from current benefits. Consider that there is a foreign firm and a domestic firm in period one. The domestic firm can either license its technology or form a JV with the domestic firm. However, there is an element of uncertainty in the continuation of the domestic government policy in the sense that domestic government may not allow any new JV in the second period. Strong opposition from various quarters or a newly elected government may be the reason for this type of policy uncertainty. In this scenario, an IJV between the foreign firm and the domestic firm may occur at the beginning of the first period even if the JV provides no benefit in period one.

It is assumed that, at the beginning of period two, the foreign firm may innovate a better technology and there is a domestic entrant in the market. Thus, at the beginning of period two, there are two buyers of the new foreign technology. If the foreign firm has the only option of licensing its new technology to the domestic firm in period two, the existence of asymmetric information about the new foreign technology and the threat of imitation from the domestic licensee may not allow the foreign firm to extract the entire surplus generated from technology licensing. However, if there is an IJV, it may reduce the threat of imitation from the domestic JV partner and allows the domestic JV partner to outbid the other domestic firm's maximum offer for the technology.

More formally, assume that the foreign firm and the domestic incumbent (who was in the domestic market at the beginning) makes a JV in period one and α shows the domestic firm's equity share in the JV. Consider that the market is characterized by price competition and the domestic entrant can produce the product with a marginal cost of production that is equal to that of the existing JV. In the case of a policy reversal, the foreign firm can either license its technology to the domestic entrant (the domestic firm who enters in the second period) or give the technology to the existing JV. Assume that the foreign firm knows that its new technology will generate a surplus V_2 in the domestic market, while the domestic firms believe that the return from the new foreign technology will be either $V_1 < V_2$ or V_2 with probability p and $(1-p)$, respectively. Further, consider that the domestic entrant can imitate the new foreign technology after getting it by incurring a cost I whereas the possibility of JV reduces the threat of imitation from the domestic JV partner. The more vertically integrated structure or some ownership advantage of the foreign firm in the JV may reduce the imitation possibility of the domestic JV partner by allowing better monitoring by the foreign firm. Hence, if the foreign firm licenses its technology to the domestic entrant, it gets the minimum of (V_1+I) and V_2.[3] A binding imitation threat implies that $(V_1+I) < V_2$. Therefore, the foreign firm cannot extract the entire surplus V_2 generated from its technology. However, if the foreign firm brings its new technology to its existing JV, it will able to extract more than (V_1+I) as the JV has reduced the imitation threat (or, increased the imitation cost). This possibility gives the foreign firm the incentive to give the technology to the existing JV. Therefore, the domestic incumbent firm faces the trade-off of sacrificing period one profit by sharing equity with the foreign firm and gaining in period two by attracting the new foreign technology to the JV. Hence, to the domestic incumbent firm, the JV acts as a hedge against the risk of not getting the surplus left to the licensee, in the case of a policy reversal.

If the domestic entrant, however, comes with a relatively inefficient technology compared to the domestic incumbent firm's (or, the JV's) technology, the JV in the first period may not occur if the domestic incumbent firm can outbid the domestic entrant under the policy reversal.

Proposition 8 (Marjit et al. 2004) *Even if a domestic firm does not get any immediate benefit from a JV with a foreign firm, it may be interested to form an IJV if the IJV gives the domestic firm an expected future benefit.*

Foreign firms often increase their equity stakes in the JVs over time and buy out the JVs completely. This is often referred as JV instability. Mukherjee and Sengupta (2001) discuss this issue in a two-period model. There is a foreign firm, called incumbent, in period one. The incumbent has the technology to produce a good, and it needs to form a JV if it wants to enter the domestic country, since the domestic regulation does not allow fully owned foreign subsidiary. Hence, the foreign firm forms a JV with a domestic firm and the firms share the profit of the JV firm according to their shares in the JV. The domestic government liberalizes its policy in the next period and allows fully owned foreign subsidiaries. However, the firms do not have any idea

[3]For the details of this solution, one may see Gallini and Wright (1990).

about future liberalization in period one. This liberalized policy of the domestic government encourages the foreign incumbent firm and other n foreign firms[4] to open respective fully owned subsidiaries. However, the entry decision of the incumbent foreign firm is different from that of the other foreign entrants since it has already established its JV in the domestic country. The decision of the incumbent foreign firm depends on several factors such as its shares in the existing JV, the controlling structure in the JV and the objective of the management of the incumbent foreign firm, sunk investment, and competition in the market characterized by the number of other foreign entrants. Assume that the firms behave like Cournot oligopolists with the same marginal cost of production, and production requires investment in capacity that is reversible.

If the managements of the firms maximize profits of the respective subsidiaries, i.e., if the firms act like 'holding companies' as defined in Kwoka (1992), the incumbent foreign firm will open its fully owned subsidiary following the liberalization if:

$$\pi_i(n+2) + (1-s)\pi_j(n+2) > (1-s)\pi_j(n+1) \tag{2}$$

where π_i, π_j and s are the profit of the incumbent foreign firm's fully owned subsidiary, profit of the JV firm, and the equity share of the domestic firm in the JV. If condition (2) is satisfied, i.e., if the initial equity shareholding of the incumbent foreign firm in the JV is not sufficiently high, the incumbent foreign firm's threat of entry is credible. Hence, it may create the incentive for share redistribution in the JV toward the incumbent foreign firm. However, the domestic partner of the JV will have an incentive for such redistribution if it makes the domestic firm better-off compared to no-redistribution, i.e., if

$$\underline{s}\pi_j(n+1) > s\pi_j(n+2) \tag{3}$$

where $\underline{s}$ equates the two sides of the condition (2).

If the total profits of two firms in a Cournot oligopoly is lower than the profit of a single firm when the number of firms in the industry is reduced by one, following Salant et al. (1983), the domestic partner will not adjust shares in the JV if the domestic firm has sufficiently high equity stake in the JV. In this situation, the loss from a lower share in the JV dominates the gain from less product market competition. However, if the domestic partner's equity share in the JV is not very high, it encourages the domestic firm to reduce its equity stake up to $\underline{s}$ that prevents the entry of the incumbent foreign firm. Since the managers of these firms are interested to maximize the profits of their subsidiaries only, the controlling structure of the JV does not play any role here.

The controlling structure may be important if the management of the incumbent foreign firm maximizes its total profit, i.e., the sum of the profits of its fully owned

[4]It may be that the patent of the incumbent foreign firm has expired or R&D by other foreign firms allows them to produce the good.

subsidiary and its profit from the JV firm, while choosing the output of its fully owned subsidiary. If the JV is controlled by the domestic firm, whether the threat of entry from the foreign incumbent firm is more or less compared to the case of a 'holding company' depends on the product market competition.

If the foreign partner controls the JV, it takes the output decision in the JV and in its fully owned subsidiary, if it opens the fully owned subsidiary. Therefore, it is optimal for the foreign incumbent to produce the colluding output of its two subsidiaries. Hence, the foreign partner may want to buy out the existing JV completely.

If the investment in capacity is sunk, it creates an asymmetry among the firms and allows the JV to behave more aggressively in the product market compared to other firms in the post-entry game. As a result, the threat of entry from the foreign incumbent is less credible compared to reversible investment. Hence, even if the foreign partner controls the JV and the management of the foreign incumbent firm's fully owned subsidiary takes into account the foreign incumbent firm's equity share in the JV, unlike reversible investment, the foreign incumbent may not want to close the JV.

Proposition 9 (Mukherjee and Sengupta 2001) *Suppose there is an existing IJV in a domestic country.*

(a) *Assume that investments are reversible.*

  (i) *There will be complete buyout of the JV by the foreign partner when the foreign partner controls the JV and the manager of the foreign firm's own subsidiary takes into account the foreign firm's equity interest in the JV.*
  (ii) *If the domestic partner controls the JV, the entry threat from the foreign JV partner may be less when the management of the own subsidiary of the foreign JV partner takes into account the foreign partner's equity interest in the JV compared to the case of a 'holding company,' provided there is a large number of foreign entrants.*

(b) *Assume that the investments are irreversible.*

  (i) *Even if the JV is controlled by the foreign firm and the own subsidiary of the foreign JV partner takes into account of the foreign firm's equity interest in the JV, the JV will not face any threat of entry form the foreign JV partner if it has sufficiently large equity stake in the JV.*
  (ii) *If the domestic partner controls the JV, the entry threat from the foreign JV partner is less when the fully owned subsidiary of the foreign partner maximizes total profit compared to the situation where it maximizes the firm profit, if there is a large number of foreign entrants.*

There are some other papers explaining JV instability. Banerjee and Mukherjee (2010) extend Mukherjee and Sengupta (2001) and show how the differences in the marginal costs of the firms affect share adjustment in the JV. They also provide a rationale for the co-existence of an IJV and the fully owned subsidiary of the foreign JV partner in the same market. Other papers show JV instability due to sequential economic liberalization, innovation and imitation (Sinha 2001a, b), private information about the market demand (Sinha 2008), demand uncertainty (de Hek and Mukherjee 2011), synergy and learning (Kabiraj 1999; Kabiraj and Sengupta 2018), product market competition (Marjit and Roy Chowdhury 2004), cultural differences between the JV partners (Kabiraj et al. 2005), and technology adoption (Kabiraj and Roy Chowdhury 2008).

Zhao (1997) provides a rationale for a JV based on a political-economy reason. He considers a market with a technologically advanced foreign firm and a technologically backward domestic firm. The domestic firm has significant power to influence the decisions of the domestic policymakers. In particular, Zhao (1997) considers that the domestic firm can lobby for a quota on the output of the foreign firm and there is a cost of lobbying. This threat of lobbying by the domestic firm may encourage the foreign firm to form an IJV with the domestic firm where the JV competes with the fully owned subsidiaries of the domestic firm and the foreign firm.

Consider a simultaneous move game where the domestic firm and the foreign firm form a JV (fixing their share in the JV and the JV output through Nash bargaining) and the domestic firm decides the lobbying amount and the output of its own subsidiary. Such a JV formation does not necessarily imply less lobbying by the domestic firm. The presence of the JV reduces the market share of the domestic firm's own subsidiary, but the domestic firm gets the benefit of cost efficiency in the JV, since the JV firm uses the foreign firm's better technology. If the foreign firm's equilibrium output under no JV is large compared to the domestic firm's gain from cost efficiency in the JV, the domestic firm will lobby more in the presence of the JV. The domestic firm does it to make up for its loss of output under JV. Only if the foreign firm's equilibrium output under no JV is small compared to the domestic firm's gain from cost efficiency in the JV, the domestic firm will lobby less under the JV.

For sufficiently large cost efficiency of the foreign firm, the domestic firm can be compensated for the loss of market share of its own subsidiary. In this situation, less lobbying from the domestic firm helps the foreign firm's own subsidiary to get a large market share and makes the foreign firm also better-off. Thus, the benefit from lower lobbying can outweigh the loss from higher competition following a JV formation.

The above argument is based on the assumption that the decision on JV formation, lobbying effort by the domestic firm, and the output of the domestic firm's own subsidiary are taken simultaneously. Realizing the threat of domestic firm's lobbying, the foreign firm can form a JV prior to lobbying and the domestic firm's output decision. This possibility will give the foreign firm the benefit of a first mover. Since the JV is formed prior to the domestic firm's lobbying, it reduces domestic lobbying by reducing the market share of the domestic firm's own subsidiary. Thus, it creates the incentive for such a JV formation.

Proposition 10 (Zhao 1997) *If the domestic firm has significant power to influence the trade policies, an international JV may benefit the firms.*

Singh and Bardhan (1991) show how rivalry among the firms as well as the decision structure in the JV may affect the policies of the domestic countries. They consider a JV between a foreign firm and a domestic firm and focus on two situations: one, where the JV produces an intermediate good and, the other, where the JV produces a final good.

First, consider the case where the JV produces an intermediate good with a constant marginal cost. The foreign partner of the JV buys this intermediate product and produces a final good with it. Consider a fixed coefficient technology in the production of the final good. The foreign JV partner competes with another firm in the rest of the world. The final products of the firms are imperfect substitutes. First, the domestic government determines the foreign equity participation in the JV and price of the indigenous input, say labor, with which the intermediate good is produced. Second, the price of the intermediate good is determined. Finally, the firms take the decisions on outputs or prices of the final goods.

The profit of the foreign JV partner, called firm A, is

$$\pi^{\mathrm{A}} = (P_{\mathrm{A}} - t) f^{\mathrm{A}}(P^{\mathrm{A}}, P^{\mathrm{B}}) + \lambda(t - c) f^{\mathrm{A}}(P^{\mathrm{A},} P^{\mathrm{B}}) \tag{4}$$

where P^{A}, P^{B} are the respective prices of the final output of firm A and its competitor in the product market, called firm B. The terms λ, *t,* and *c* stand for firm A's share in the JV, price of the intermediate good, and the cost of the indigenous input required to produce the intermediate good.

The welfare of the domestic country is

$$W = (1 - \lambda)(t - c) f^{\mathrm{A}} + \phi c l \tag{5}$$

where l is the level of employment in the JV and ϕ is the weight given to wages and employment relative to the domestic firm's share of profits in the JV.

If firm A can set the price of the intermediate good, it follows from (4) that it will set $t = c$ as firm A is interested to maximize its global profit. In this situation, it does not matter from the point of view of the domestic country's welfare that how much share is owned by the foreign partner. Hence, the domestic government can maximize the welfare of the country by choosing the price of the indigenous input only. The domestic government tries to maximize the second term in (5), and the nature of competition prevailing in the final goods market may affect the domestic government's decision. Competition is severe under price competition compared to quantity competition. Hence, for the same amount of increase in price of the indigenous input, firm A's output and, therefore, JV employment suffer more under price competition compared to quantity competition. It induces the domestic government to charge a lower price for the indigenous input (i.e., lower c) under price competition compared to quantity competition in the product market.

If the domestic partner of the JV or the JV firm chooses the price of the intermediate product, it will set the price of the intermediate good to maximize the profit of the JV. However, the foreign partner of the JV gets its share according to its equity participation in the JV. Hence, it follows from (5) that the domestic government will opt for complete indigenization (i.e., $\lambda = 0$). Without any moral hazard problem on the part of the foreign JV partner, the welfare of the domestic country is a strictly monotonic function of the domestic firm's equity share, as shown in (5). Further, for a sufficiently a low weight on the wages and employment, the domestic government has no incentive to distort production in the JV and, the domestic government prefers lowest possible price for the indigenous input. Even if there is a sufficiently higher weight on the wage and employment, the domestic government chooses an interior solution for the input price.[5]

Proposition 11 (Singh and Bardhan 1991) *Suppose the JV in the domestic country produces an intermediate input, the foreign JV-partner consumes this input and the foreign partner of the JV competes with another firm in the world final goods market.*

(i) *If the domestic firm or the JV firm sets the price of the intermediate product, the domestic government prefers the minimum possible JV share of the foreign partner. The domestic government prefers minimum possible price of the indigenous input required to produce the intermediate good, provided the weight on the wage and employment in the JV is sufficiently low.*
(ii) *If the foreign partner chooses the price of the intermediate product, the distribution of equity share will have no effect on the domestic welfare. The domestic government's decision on the indigenous input price will depend on the nature of competition in the final goods market, and it prefers a lower price under price competition compared to quantity competition in the final goods market.*

Now, consider the situation where the JV produces a final good. For simplicity, they assume that the JV sells its product in the rest of the world. However, the JV uses two substitutable intermediate inputs, supplied by the foreign JV partner (firm A) and a rival foreign firm (firm B), and a domestic input (say, labor). Suppose the JV firm, as an 'autonomous subsidiary,' takes the decision on input utilization in the JV. If firms A and B compete in prices, a higher share of the foreign partner in the JV reduces the prices of the input supplied by firm A. If the inputs act as strategic complements, it will also reduce the price of the inputs of firm B. However, if the firms compete in quantities, a higher share of the foreign partner in the JV increases the usage of the inputs of firm A, and if the inputs act as strategic substitutes, it reduces the usage of the inputs of firm B.[6] While a higher share of the foreign firm in the JV reduces the welfare of the domestic country by reducing the domestic firm's share of profit in the JV, it helps to increase the domestic welfare by affecting input supply. This may encourage the domestic government to allow some foreign shareholding. If the foreign JV partner takes the decision on the use of input in the JV, it is better for the domestic government not to allow foreign shareholding or complete indigenization. In this situation, the foreign JV partner chooses the inputs in a way that maximizes its global profit. Therefore, the foreign JV partner has the incentive to increase the price of the inputs as transfer pricing without affecting the input choice. However, this will increase the price of the substitute inputs. Further, the incentive for a higher transfer price increases with a higher foreign shareholding as the loss from higher

[5]The domestic government chooses the minimum possible value of c for $\phi < \frac{1}{2}$, but it chooses an interior value for c when $\phi > \frac{1}{2}$.

[6]See Bulow et al. (1985) for the definitions of 'strategic complements' and 'strategic substitutes.'

input prices is lower with more foreign shareholding in the JV. Thus, it reduces the domestic welfare and encourages complete indigenization.

Proposition 12 (Singh and Bardhan 1991) *Suppose the JV produces the final good which is sold in the rest of the world and the JV uses substitutable inputs of the foreign JV-partner and another foreign firm. If the foreign JV-partner takes the decision on the use of inputs in the JV, the domestic government prefers complete indigenization. If the JV firm, as an 'autonomous subsidiary', takes the decision on the use of inputs in the JV, the domestic government is likely to allow some share-holding by the foreign firm in the JV.*

3 Information Asymmetry

Often JVs help to eliminate or reduce the problems due to information asymmetry creating moral hazard or adverse selection.

Chan and Hoy (1991) consider a situation where two firms—a domestic firm and a foreign firm—produce a product jointly, and each of them provides an input to the production process. The quality of the inputs is non-contractible. Thus, it creates a double-sided moral hazard problem and each firm maximizes an objective function similar to the expression in (1). The optimal shares, which will lie between 0 and 1, will depend on the importance of the inputs in the production process. Since providing higher qualities are costly and no firm gets full return from its inputs, each firm has the incentive to lower the quality level of its input from the joint-profit maximizing quality level. Hence, the optimal amounts of inputs, outputs, and profits will be less than the joint-profit maximizing levels.

If the inputs of the foreign firm are verifiable up to a certain level, a minimum standard requirement can be imposed on the foreign firm's inputs. This helps to reduce the moral hazard problem on the part of the foreign firm and may encourage the domestic firm to have a fully owned subsidiary that purchases the foreign firm's input through an arm's length contract. Hence, if the minimum verifiable quality level of the foreign firm is not very low, the fully owned subsidiary of the domestic firm with an arm's length contract for the foreign firm's input creates higher surplus than the JV.

If the foreign firm provides multiple inputs and some of its inputs can be verified up to a level, there could be arm's length contracts for the verifiable inputs. However, the presence of non-verifiable inputs may create the incentive for forming a JV.

Proposition 13 (Chan and Hoy 1991) *If there is a double-sided moral hazard problem, JVs can be optimal. With the verifiability of minimum quality standards for some inputs, the optimal arrangement may be the combination of minimum standard requirements and equity sharing. The higher possibility of verification may increase the incentive for fully owned subsidiaries.*

Often the JV contracts consist of buyback provisions, which specify a fixed division of outputs prior to production. Choi and Maldoom (1992) show that buybacks act as a commitment device to reduce opportunism of the JV partners and generate higher surplus than the market allocated contract.

Consider a problem like the one mentioned above. Two firms provide two non-contractible inputs. Assume that a portion s of the total output is sold in the domestic market and the rest is sold in the world market. The firms share total revenues from both markets by a sharing rule. In the case of a buyback contract, the firms determine the shares of the revenues as well as the portion s of the outputs. These values are determined by maximizing the aggregate profit of the JV. Hence, the buyback contract helps to commit to output sharing before production. While determining the shares of the revenue and the output, the firms consider the corresponding input choice conditional on the sharing rule.

Under a market allocated contract, the share of the revenue is determined before production but the outputs are divided optimally after production. Since the division of outputs is not fixed before production, the output adjustment helps to reduce the negative effect of shirking. Given an input choice, the loss from shirking to a firm is less when the outputs adjust after production compared to the buyback contract. Hence, when deciding the share of the revenue, the market allocated contract imposes an extra incentive constraint due to the possibility of output adjustment and creates higher incentive to shirk under the market allocated contract compared to buyback.

Proposition 14 (Choi and Maldoom 1992) *Under a buyback contract, sharing of outputs are fixed prior to production. Thus, it acts as a commitment and helps to increase profit compared to a situation with no fixation of output sharing prior to production.*

Roy Chowdhury (1997) shows that the moral hazard problem can affect technology transfer in a JV. Consider that a foreign firm, called firm 1, and a domestic firm, called 2, produce a product jointly. If the project is successful, the gross profit is $2R$ which the firms divide equally. This return is assumed to be constant and does not depend on the effort levels of the firms. The payoff is zero in the case of a failure of the product. In the production process, each firm provides non-contractible efforts and each of them faces the costs of effort as $c_i(e_i) = \frac{h_i e_i^2}{2}, i = 1, 2$, where e_i shows the effort level of the ith firm and h_i shows the efficiency of the ith firm in providing the effort level with $h_i > 0$. If $h_i > h_j; i, j = 1, 2; i \neq j$, the firm j is more technologically efficient than firm i. Assume that the project's probability of success depends on the joint effort levels of the firms and is given by $\phi(e_1 + e_2)$. At the beginning, the firms simultaneously decide on the technology levels (h_i) from an interval $[\underline{h_i}, \overline{h_i}]$, and after that, they simultaneously provide the effort levels (e_i). The support levels for the technologies however differ between the firms. The foreign firm is assumed to be technologically efficient with $\overline{h_1} < \underline{h_2}$.

Given the technology levels determined in stage one, the firms choose the effort levels in a way that their technology-adjusted effort levels are equal, i.e., $e_1 h_1 = e_2 h_2$. When determining their technologies in stage one, the firms internalize this fact.

Since the domestic firm is technologically inefficient, it provides less effort in stage two. It encourages the foreign firm to increase its effort level and to provide its efficient technology in stage one.

A sufficiently high return from the project encourages the foreign firm to provide sufficiently high efforts, which, in turn, induces the domestic firm to provide little effort and to provide its inefficient technology.

Proposition 15 (Roy Chowdhury 1997) *Suppose a technologically efficient foreign firm and a technologically inefficient domestic firm engage in a production process. With the possibility of endogenous technology choice, the domestic firm may provide the inferior technology whereas the foreign firm provides the superior technology.*

The above discussion focused on the double-sided moral hazard problem but did not pay much attention on the managerial control. Lee and Shy (1992) address this issue.

Consider an IJV where the foreign firm provides the technology and a better technology helps to produce a higher-quality product. The transfer of technology however imposes a cost, c^f, on the foreign firm, and the cost depends on the quality of the technology and the output of the JV. The domestic firm does not contribute any input in the JV. The foreign firm's share in the JV is fixed ex ante by the domestic government, and the output decision of the JV depends on the managerial control. Consider that majority shareholder in the JV takes the output decision. Further, consider that the cost of production in the JV, c^{JV}, depends on the quality of the transferred technology. A better technology that produces a higher-quality product creates a higher cost of production. Assume that the JV serves only the domestic market and there is a continuum of consumers indexed according to their preferences $\delta \in [\underline{\delta}, \overline{\delta}]$, $\underline{\delta} > 0$. For simplicity, each consumer buys one unit of the good. Hence, the total output of the JV firm is given by the number of people consuming the good. If a consumer buys the good, its utility is $\delta u(x) - p$, where x and p are the quality and price of the product, respectively. Otherwise, the utility of the consumer is 0.

The profit of the foreign firm and the domestic firm are, respectively,

$$(1 - \bar{\alpha})\pi^{\text{JV}} - c^f \text{ and } \bar{\alpha}\pi^{\text{JV}} \tag{9}$$

where the profit of the JV is $\pi^{\text{JV}} = (p - c^{\text{JV}})Q$ and the output of the JV is Q.

First, consider that the domestic firm controls the JV. In this situation, the foreign firm first chooses its technology and then the domestic firm decides the output. If the foreign shareholding increases, it increases the foreign firm's profit for a given output of the JV. Hence, it increases the quality of the transferred technology. However, the effect on the output is ambiguous. The higher-quality good implies a higher price due to a higher cost of production. However, a higher quality increases the utility of the consumers and encourages more people in the market. The trade-off between a higher utility of the consumers and a higher price due to a higher cost of production determines the market size and the output of the JV.

If the foreign firm controls the JV, it chooses the quality of the technology as well as the output of the JV when maximizing its profit. Since the cost of technology transfer depends on the output of the JV, the choice of technology depends on the foreign firm's marginal gain from technology. If the marginal profit of the foreign firm does not fall with a higher quality, a higher shareholding by the foreign firm increases the quality of the transferred technology. However, no relationship between the cost of technology transfer and the output of the JV implies that a higher shareholding by the foreign firm always increases the quality of the transferred technology.

Like the quality of the technology, the effect on the domestic welfare is also ambiguous with respect to the change in foreign shareholding. A higher price due to a higher-quality good makes exit of the consumers with lower willingness to pay. However, the consumers purchasing the product gain from a higher-quality good. Whether the market size, which is equal to the JV's output, will increase or decrease following an increase in the foreign firm's share will depend on the managerial control and the marginal profit of the foreign firm with respect to quality. A higher shareholding by the foreign firm has two effects on the domestic profits. First, it reduces the profit of the domestic firm due to low shares. Secondly, a higher foreign shareholding tends to increase the quality of the transferred technology and helps to increase the profit of the JV and that of the domestic firm. These opposite effects determine the effect on the domestic welfare.

Proposition 16 (Lee and Shy 1992) *The quality of the foreign technology that is transferred to the domestic project and the domestic welfare depend on the managerial control as well as on the cost of technology transfer.*

Although Lee and Shy (1992) show the implications of managerial control, they ignore the possibility of arm's length contract such as technology licensing. Marjit and Mukherjee (2001) take up this issue and consider technology licensing and international JV under asymmetric information about the foreign technology and the threat of imitation from the domestic licensee.

Assume that there is a foreign firm, which has technologies to produce a product but cannot open its fully owned subsidiaries in the domestic country due to government regulations. There is a firm in the domestic country, which can produce the product with a technology that is inferior to the foreign firm's technologies. For simplicity, assume that the foreign firm has two technologies to produce the good and both the technologies are superior to the domestic firm's technology. However, only the foreign firm knows which one is the better technology, and there is no way to verify the qualities of the technologies before technology transfer. One possibility is to disclose all the relevant information to the domestic firm. However, due to the threat of imitation from the domestic firm, the foreign firm has no incentive to disclose information about the technology to the domestic firm before technology licensing.

The foreign firm can either license its technology or form a JV with the domestic firm. Suppose the project's profits from using the domestic firm's technology and the foreign firm's technologies are π_0, π_1 and π_2, respectively, with $\pi_0 < \pi_1 < \pi_2$. Further, assume that the foreign firm has to incur a higher cost for transferring a better technology. Consider v_1 and v_2 as the costs for transferring the foreign firm's relatively inferior and superior technologies, respectively, with $v_1 < v_2$. However, transferring the better technology of the foreign firm is efficient, i.e., $\pi_2 - v_2 > \pi_1 - v_1$.

If the foreign firm licenses its technology, it can charge an up-front fixed fee π_1 and an output-based royalty I, where I is the cost incurred by the domestic firm while imitating. This royalty will be paid, provided the output of the project exceeds certain amount. If the foreign firm charges a royalty higher than the cost of imitation, it is better for the domestic firm to imitate the licensed technology and to avoid the royalty payment by producing with the imitated technology. Hence, the no-imitation constraint requires that the royalty payment should not exceed the imitation cost. Therefore, the maximum amount that the foreign firm can receive from technology licensing is min.$\{\pi_1 + I, \pi_2\}$. Assume that $\pi_1 + I < \pi_2$. Therefore, under licensing, the foreign firm cannot extract the entire surplus generated from its superior technology. Hence, the foreign firm will be interested to transfer the inferior technology if the royalty payment is less than its cost saving from transferring the inferior technology, i.e., $I < v_2 - v_1$. Thus, the threat of imitation along with asymmetric information creates a restriction on rent extraction by the foreign firm and the foreign firm transfers its inferior technology under technology licensing against an up-front fixed fee π_1.

Now, consider a JV between the firms. Assume that, under JV, the firms get the dividend income only according to their shares in investment. If the foreign firm holds sufficiently higher equity share in the JV, it provides the foreign firm the incentive to transfer its better technology in the JV. The threat of imitation from the domestic firm remains even under JV. However, to get the benefit from imitation, now the domestic firm needs to operate its own subsidiary. Thus, the equity share under the JV gives the foreign firm an ownership advantage.

If the foreign firm controls the JV, the domestic firm's fully owned subsidiary needs to compete with the JV. The possibility of competition may not make imitation and opening a fully owned subsidiary by the domestic firm an attractive option for the domestic firm, thus eliminating the threat of imitation by the domestic firm.

If the domestic firm controls the JV, it can operate its fully owned subsidiary and can produce zero output in the JV after imitation. However, a high cost of opening a new subsidiary may not make imitation and opening a fully owned subsidiary by the domestic firm economically viable. Thus, the JV may eliminate the threat of imitation even if the domestic firm controls the JV.

A related problem may arise where the foreign firm has only one technology instead of multiple technologies, and the quality of the technology is known to the foreign firm only. Here, the foreign firm with a better technology faces the challenge to distinguish itself from the foreign firm with an inferior technology. In this situation, the foreign firm with the superior technology signals its type by choosing a licensing contract accordingly. It will charge $\pi_1 + I$ for its technology where I is a lump-sum royalty based on the output of the project. To distinguish itself from the foreign firm with the inferior technology, the foreign firm with the superior technology cannot ask more than π_1 as an up-front fee. However, if the firms form a JV, the ownership advantage may make it easier to separate a foreign firm with the superior technology from a foreign firm with the inferior technology as well as to eliminate the threat of imitation. The argument is similar to the one provided above. However, to make the JV preferable than licensing, an up-front fixed fee and/or output-based royalty may be needed along with equity sharing.[7]

Proposition 17 (Marjit and Mukherjee 2001) *If a foreign firm has private information about its technologies and faces imitation threat from the domestic licensee, a JV between the firms, which attracts a better foreign technology, may be preferable to technology licensing.*

Marjit and Mukherjee (1998) focus on a different type of information problem. They consider that the foreign technology seller does not behave opportunistically but the foreign firm and the domestic firm have different perceptions about a foreign technology. Assume that there is a domestic monopolist who can produce with its technology and can get π_0. There is a foreign firm that wants to enter the domestic market but cannot open its fully owned subsidiary due to domestic regulation. The foreign firm has a better technology than the domestic firm. The foreign firm expects that its technology will yield a profit π_2, while the domestic firm expects that the technology will yield $\pi_1 < \pi_2$. The technology is new, and hence, the domestic firm may worry about the local conditions that might influence the return from the foreign technology. Consider that the (Nash) bargaining process determines the price of the technology.

If the foreign firm wants to license its technology against an up-front fixed fee, the domestic firm wants to pay according to its perceived return from the technology and the price of the technology depends on π_1.

The foreign firm however wants to get a price depending on its own perception about the return, i.e., π_2. One possibility might be to defer payment for the technology and to make the payment contingent on the outcome according to a prespecified bargaining rule. While this does not change the expected return for the domestic firm, it changes the foreign firm's expected payoff as it expects a higher return from the technology. However, under this agreement, the domestic firm gets a larger share than the amount it expects. Hence, the foreign firm is interested to make an arrangement that gives the domestic firm the amount that it gets under the licensing agreement.

Consider a JV between the firms. The firms take financial stake in the project and get dividends. Since there is a difference in perception ex ante investment, the foreign firm may invest in a way so that it guarantees the domestic firm's expected payoff that it will get under the licensing contract.[8] The marginal loss to the domestic firm due to foreign investment is π_1, but the marginal gain to the foreign firm from investment is $\pi_2 > \pi_1$. This potential gain from investment sharing makes the situation Pareto superior compared to the earlier cases. However, the foreign firm believes that ex post

[7] Related empirical findings can be found in Killing (1983), Mansfield (1994), Saqib (1995) and Bhandari (1996).

[8] The assumption is that the domestic regulation allows the foreign firm to hold this amount of equity stake in the project. Otherwise, the foreign firm has to invest up to the maximum allowed foreign equity participation. In this situation, the foreign firm can charge a lump-sum fee in the JV.

investment, the domestic firm gets a dividend income corresponding to π_2. Hence, under the JV, there are some gains not captured by the foreign firm.

The domestic firm acting as an agent of the foreign firm may be another possible agreement. The foreign firm may charge a lump-sum amount that is payable ex post investment so that the domestic firm gets the amount that it gets under the licensing contract. However, this type of arrangement may induce opportunistic behavior (maybe like imitation as mentioned previously) on the part of the domestic firm, which in turn asks the foreign firm to sacrifice some profit. On the other hand, the JV may help to eliminate or to reduce the opportunistic behavior of the domestic firm.

The opportunity cost of investment is often higher for the domestic firm of a developing country compared to a foreign firm from a developed country. A highly protected capital market may be the reason for this. This makes the investible funds more costly to the domestic firms compared to the foreign firms, thus creating another motivation for IJV. In this respect, one may refer to Harvey (1995), which discusses that the negative relationship between the opportunity costs of the investible funds for the domestic and the foreign firms is a major factor behind the equity investment by the foreign firms in many developing countries. If the difference between the opportunity costs of investment to the domestic firm and the foreign firm reflects the risk premium, the portfolio diversification might explain equity investment across the borders. Due to severe restriction on capital market, if there exist some pure profits than only risk premium, the JV will be a preferred arrangement. This avenue would be closed once the domestic firms are allowed to raise capital from the international market.

Proposition 18 (Marjit and Mukherjee 1998) *If a foreign firm perceives a higher return from its technology compared to a domestic firm and the opportunity cost of investment to the foreign firm is less compared to the domestic firm, an IJV may be preferred to technology licensing.*

Marjit et al. (1995) show that the difference in risk perception between the domestic government and a foreign firm may create a public–private IJV. Assume that the gross profit from an investment I is $R(I)$, which is increasing and concave in I. The foreign firm assumes that there are some country-specific risks and it reduces the success of the project. Consider that the foreign firm believes that β is the probability of success due to the existence of a country-specific risk. Hence, the foreign firm's optimal investment under its fully owned subsidiary is given by $\beta R'(I_\beta) = 1$. This amount of investment however is less than the amount if there is no country-specific risk, i.e., for $\beta = 1$.

Assume that there is a difference in perception between the domestic government and the foreign firm about the country-specific risk. The domestic government believes that the foreign firm's belief about the country-specific risk is overestimated. However, the domestic government has no way to convince the foreign firm about it. To the domestic government, the return from the project is $R(I) - I$. This creates a potential gain from collaboration. The domestic government can choose a JV consisting of the foreign firm's investment sharing as τ_1 and the foreign firm's profit-sharing as α_1. In fact, the investment and profit-sharing rule satisfying

$$\tau_1 = \alpha_1 \beta \tag{10}$$

can generate the equilibrium investment level that maximizes $R(I) - I$.

The condition (10) shows that, in the public–private IJV, the invest sharing and the profit-sharing are not the same. The foreign firm's share in investment is less than its share in profit. By this mechanism, the domestic government shares the country-specific risks perceived by the foreign firm, and it helps to insure the foreign firm against its perceived country-specific risk, thus providing the rationale for a JV.[9]

Proposition 19 (Marjit et al. 1995) *If a foreign firm faces some country specific risks, a JV between the domestic government and the foreign firm can achieve the first-best investment level.*

When considering international trade and foreign investment, exchange rate risk is another important concern. Firms participate in futures and forward markets to hedge against the exchange rate fluctuations. However, many less developed countries do not have efficient futures and forward markets, and an IJV between a foreign firm and a domestic firm helps to share the risk from exchange rate fluctuations. Broll and Marjit (2005) explain this.

Consider a situation where a foreign firm can invest I initially and gets a return R afterward with $R > I$. The values of R and I are measured in terms of the domestic currency. Suppose the spot exchange rate at the time of investment is $e_0 (= 1, \text{say})$ and the uncertain spot exchange rate at the time of return is $\tilde{e}_1$. Therefore, the net income of the foreign firm in its own currency is $\tilde{y} = R\tilde{e}_1 - I$. Assume that the foreign firm is risk-averse with a von Neumann–Morgenstern utility function and maximizes the expected utility of its own currency income. With a mean-variance preference, the foreign firm's expected utility is $E(\tilde{y}) - \frac{\eta}{2}\text{var}(\tilde{y})$, where $\eta > 0$ is the constant degree of absolute risk aversion. Hence, the critical expected exchange rate that makes the foreign firm indifferent between investment and no-investment is

$$\mu_1^c = \frac{I}{R} + \frac{\eta}{2}\text{var}(\tilde{e}_1)R \tag{11}$$

where $\mu_1^c = E(\tilde{e}_1)$. Therefore, the foreign firm will not invest if its expected exchange rate is lower than the critical level shown in (11). Assume this is the case. Further, assume that $R\mu_1 > I$.

Now, consider that the foreign firm invests α fraction of the total investment and gets the share of the profit accordingly. The domestic firm invests rest of the amount and gets the share of profit accordingly. Hence, the net income of the foreign firm is $\tilde{y} = \alpha(R\tilde{e}_1 - I)$ and the income of the domestic firm is $(1 - \alpha)(R - I) > 0$. Given a sharing rule, the critical expected exchange rate that makes the foreign firm indifferent between investment and no-investment is

[9]It is worth noting that a particular tax–subsidy scheme can replicate the result obtained from the public–private IJV.

$$\mu_1^{\mathrm{JV}} = \frac{I}{R} + \frac{\eta}{2}\alpha \operatorname{var}(\tilde{e}_1)R. \tag{12}$$

It follows from (11) and (12) that $\mu_1^{\mathrm{JV}} < \mu_1^c$. Hence, the JV helps to increase investment. Since the domestic firm is not facing any risk, it creates the possibility of risk sharing between the firms.

In the framework considered here, it is always optimal for the foreign firm to sell the firm to the domestic firm against a lump-sum amount. This will share the risk to the maximum amount. However, the foreign exchange constraint in the developing countries often does not provide such benefits. Further, the difference in the opportunity cost of raising currencies between the foreign and the domestic markets can make the buyout possibility sub-optimal.

Proposition 20 (Broll and Marjit 2005) *If there is no efficient risk sharing arrangement, an IJV may be preferred to the foreign firm's fully owned subsidiary under exchange rate uncertainty.*

4 Conclusion

In this survey, we discuss some analytical issues to show the implications of policy uncertainties and information asymmetry on JV formation. There are other important issues such as transfer pricing and product market competition, which provide further insights. However, those issues are beyond the scope of this survey.

There are some issues, which did not get much attention in the literature. For example, the implications of the intellectual property rights on JV formation might be an important issue to consider. Formation and breakdown of JV are other dynamic phenomena, which did not get sufficient attention from the researchers. Although this survey discusses few papers related to share adjustment in the JVs, this area deserves a thorough coverage. Thirdly, the JV literature did not pay much attention on JVs in intermediate goods.[10] Finally, while considering JV in a developing country, financial institutions might play an important role, which, in turn, may have consequences on the capital structure of the JV.

Acknowledgements We would like to thank Sugata Marjit and an anonymous referee for helpful comments and suggestions. The usual disclaimer applies.

References

Abdalla, A. E. A. (1992). Distortions to capital mobility: The case of joint ventures. *Economics Letters, 38,* 443–449.

[10] See Kabiraj and Chaudhuri (2001) for a paper in this area.

Al-Saadon, Y., & Das, S. P. (1996). Host-country policy, transfer pricing and ownership distribution in international joint venture: A theoretical analysis. *International Journal of Industrial Organization, 14,* 345–364.

Banerjee, S., & Mukherjee, A. (2010). Joint venture instability in developing countries under entry. *International Review of Economics and Finance, 19,* 603–614.

Beamish, P. (1988). *Multinational joint ventures in developing countries*. New York: Routledge.

Bhandari, L. (1996). *International technology transfer: Control, contracts and learning between firms*. (Ph.D. dissertation, Boston University).

Broll, U., & Marjit, S. (1995). Joint venture under uncertainty. *Current Policies and Economics in Europe, 4,* 78–85.

Broll, U., & Marjit, S. (2005). Foreign investment and the role of joint ventures. *South African Journal of Economics, 73,* 474–481.

Bulow, J., Geanakoplos, J., & Klemperer, P. (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy, 93,* 488–511.

Chan, R., & Hoy, M. (1991). East-West joint ventures and buyback contracts. *Journal of International Economics, 30,* 330–343.

Choi, C. J., & Maldoom, D. (1992). A simple model of buybacks. *Economics Letters, 40,* 77–82.

Das, S. P. (1998). On the choice of international joint venture: The role of policy moral hazard. *Journal of Policy Reform, 2,* 135–150.

Das, S. P. (1999). Direct foreign investment versus licensing. *Review of International Economics, 3,* 86–97.

de Hek, P., & Mukherjee, A. (2011). Joint venture buy-outs under uncertainty. *Journal of Industrial Economics*, *LIX*, 155–75.

Gallini, N. T., & Wright, B. D. (1990). Technology transfer under asymmetric information. *Rand Journal of Economics, 21,* 147–160.

Gupta, S. P., & Chawla, R. L. (Eds.). (1995). *Prospects of foreign direct investment in India in the post liberalization era*. New Delhi: Indian Council for Research on International Economic Relations.

Harvey, C. R. (1995). The risk exposure of emerging equity markets. *The World Bank Economic Review, 9,* 19–50.

Kabiraj, T. (1999). A simple theory of joint venture break down. *Keio Economic Studies, 36,* 87–98.

Kabiraj, T., & Chaudhuri, M. (2001). An input joint venture: Some new results. In A. Guha, K. L. Krishna, & A. Lahiri (Eds.), *Trade and industry: Essays* (pp. 65–86). New Delhi: NIPFP—Ford Foundation Fellows, Vikash Publishing House Pvt. Ltd.

Kabiraj, T., Lee, C. C., & Marjit, S. (2005). Cultural compatibility and joint venture instability—A theoretical analysis. *Indian Economic Review, 40,* 209–220.

Kabiraj, T., & Roy Chowdhury, P. (2008). Adoption of new technology and joint venture instability. *Research in International Business and Finance, 22,* 108–123.

Kabiraj, T., & Sengupta, S. (2018). A theory of joint venture instability under inter-partner learning. *Research in International Business and Finance, 46,* 363–372.

Killing, J. P. (1983). *Strategies for joint venture success*. New York: Praeger Publishers.

Korbin, S. J. (1988). Ownership of American manufacturing subsidiaries in developing countries. *Management International Review* (Special Issue), 73–84.

Kwoka, J. E., Jr. 1992. The output and profit effects of horizontal joint ventures. *Journal of Industrial Economics*, *40*, 325–38.

Lee, F. C., & Shy, O. (1992). A welfare evaluation of technology transfer to joint ventures in the developing countries. *The International Trade Journal, 7,* 205–220.

Mansfield, E. (1994). *Intellectual property rights, foreign direct investment and technology transfer*. Discussion Paper No. 19, International Finance Corporation, Washington, D.C: The World Bank.

Marjit, S. (1990). Rationalizing public-private joint venture in an open economy. *Journal of Development Economics, 33,* 377–383.

Marjit, S., Broll, U., & Mallick, I. (1995). A theory of overseas joint venture. *Economics Letters, 47,* 367–370.

Marjit, S., & Chowdhury, P. R. (2004). Asymmetric capacity costs and joint venture buy-outs. *Journal of Economic Behavior & Organization, 54,* 425–438.

Marjit, S., & Mukherjee, A. (1998). Technology collaboration and foreign equity participation: A theoretical analysis. *Review of International Economics, 6,* 142–150.

Marjit, S., & Mukherjee, A. (2001). Technology transfer under asymmetric information: The role of equity participation. *Journal of Institutional and Theoretical Economics, 157,* 282–300.

Marjit, S., Mukherjee, A., & Kabiraj, T. (2004). Future technology, incomplete information and international joint venture. *Research in Economics, 58,* 219–234.

Moskalev, S. A., & Swensen, R. B. (2007). Joint ventures around the globe from 1900–2000: Forms, types, industries and ownership patterns. *Review of Financial Economics, 16,* 29–67.

Mukherjee, A. (2000). Host-country policy—Commitment or no-commitment: A theoretical analysis. *Journal of Policy Reform, 4,* 75–89.

Mukherjee, A., & Sengupta, S. (2001). Joint ventures versus fully owned subsidiaries: Multinational strategies in liberalizing economies. *Review of International Economics, 9,* 163–180.

Roy Chowdhury, P. (1997). Foreign collaborative ventures and quality of technology. *Journal of Quantitative Economics, 13,* 11–28.

Salant, S. W., Switzer, S., & Reynolds, R. (1983). Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium. *Quarterly Journal of Economics, 98,* 185–199.

Saqib, M. (1995). An overview of the economic policy and FDI in India with special reference to the USA. In S. P. Gupta et al. (Eds.), *Prospects of foreign direct investment in India in post liberalisation era* (pp. 81–99). New Dehli: Indian Council For Research On International Economic Relations.

Singh, N., & Bardhan, P. K. (1991). Government policy towards joint ventures with multinational rivalry. *Mimeo*.

Sinha, U. B. (2001a). Imitative innovation and international joint venture: A dynamic analysis. *International Journal of Industrial Organization, 19,* 1527–1562.

Sinha, U. B. (2001b). International joint venture, licensing and buy-out under asymmetric information. *Journal of Development Economics, 66,* 127–151.

Sinha, U. B. (2008). International joint venture: Buy-out and subsidiary. *Journal of Economic Behavior & Organization, 65,* 734–756.

Teece, D. J. (1977). Technology transfer by multinational firms: The resource cost of transferring technological know-how. *The Economic Journal, 87,* 242–261.

Teece, D. J. (1981). The market for know-how and the efficient international transfer of technology. *ANNALS, AAPSS*, November, 81–95.

Zhao, L. (1997). International joint ventures and endogenous protection: A political-economy approach. *Journal of Economic Integration, 12,* 548–560.

Labour and Human Capital

Household Self-Employment Eliminates Child Labour

Bibhas Saha

Abstract This paper shows that if a minimal proportion of poor households is empowered with self-employment opportunities, child labour will not arise in equilibrium. The economy will have a unique 'good' equilibrium generating a sufficiently high wage to support full child schooling.

Keywords Self-employment · Child labour · Empowerment

JEL Classification Nos: J20 · D13 · O17

1 Introduction

In a seminal paper, Basu and Van (1998) (henceforth BV) offered an explanation of child labour, in which a downward sloping household labour supply curve intersects a standard labour demand curve several times generating multiple equilibria. One of these is low-wage equilibrium where children will have to work along with their parents to meet the family's subsistence consumption. Despite several other explanations, such as negative externality (Baland and Robinson 2000), parental selfishness and external bargaining (Gupta 2000) and credit constraints (Ranjan 2001; Jafarey and Lahiri 2002), the BV model has remained an inspiration for empirical investigations in this literature [see for instance, Bhalotra and Heady (2003) and Basu et al. (2010)].

An implicit assumption crucial to the multiple equilibria of BV is that there is no lower bound on the parent's income. Certain policies work by putting such lower bounds, such as minimum wage legislation (Basu 2000) or a ban on child labour (Basu and Van 1998). However, it is well known that enforcement of a ban or minimum wage is not easy and they might cause unemployment (Basu 2000).

In this paper, we study intervention in the form of providing self-employment opportunities. The intervention is self-enforcing and easy to implement, but

B. Saha (✉)
Durham University Business School, Durham DH1 3LB, UK
e-mail: b.c.saha@durham.ac.uk

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_16

surprisingly it has not received sufficient attention in the child labour literature. We show that if only a fraction of the poor households are endowed with such opportunity, the child labour equilibrium will be eliminated.

The idea is that self-employment puts a natural lower bound on income allowing parents to switch to self-employment at low wages and protect their children's education. Their withdrawal from wage employment moves the labour supply curve inward creating excess demand (at low wages) and in turn eliminating the child labour equilibrium. The fraction of the households switching to self-employment should be just enough to cause the necessary displacement of the labour supply curve, crucial for this outcome.

This insight adds a twist to the recommendation of the microfinance literature for the extensive spread of self-employment (Morduch 1999). We argue that it suffices to help only a critical proportion of the households be self-employed, as the labour market will then correct itself.

The literature on child labour is largely concentrated on wage workers or farming households [see Bhalotra and Heady (2003) and Basu et al. (2010)]. There is also significant interest in the effect of economic liberalization or trade reform on child labour (Swaminathan 1998). However, the evidence of parental occupations on the child's likelihood of being out of school is somewhat limited. From a socio-economic point of view, self-employed households running small shops or informal businesses are considered to be better off than the wage worker households. This is so because self-employment provides a minimum assured income, which also justifies microfinance. If so, then the possibility of multiple equilibria in the labour market as argued by Basu and Van (1998) diminishes. This paper investigates this possibility.

Having said that, there is also an issue that the self-employed households may have a perverse incentive to employ their own children, cutting short their education. This particular aspect is being explored in Pal and Saha (2019), where they find that if own children could be employed, then they will be used more when the market wage rises (as a substitute for the market labour), and less when the market wage falls. In equilibrium, the child labour patterns would be different between the wage worker households and self-employed households. The existing literature has not studied how self-employment affects child labour. This paper looks at one of the two above-mentioned implications of self-employment. In their preliminary empirical analysis, Pal and Saha (2019) do find statistically significant and *positive* relationship between parental self-employment status and child's work hours. This is similar to the findings on the child labour impact of microfinance, such as Hazarika and Sarangi (2008) for rural Malawi and Islam and Choe (2013) for rural Bangladesh. But we need to be aware that these results vary across countries; for example, Karlan and Zinman (2009) found the opposite relationship in Manila.[1] Clearly, the theoretical connections between parental self-employment and child labour are intriguing, and there is a need for more empirical investigations.

[1]There is also a literature on the effect of the minimum wage increase in child labour; see Menon and van der Meulen Rodgers (2018) for India. Minimum wage increase helps to reduce child's work within the household sector, but has no impact on his/her outside work.

The rest of the paper is organized as follows. Section 2 presents the main model and the results, followed by some concluding remarks in Sect. 3.

2 The Model

We consider the same set-up as in BV with N households, each of which consists of one parent and one child, each having 1 unit of labour to supply. One unit of child labour is equivalent to $\gamma(< 1)$ unit of adult labour.[2] Households are identical except in one respect: α proportion of them have self-employment (SE) opportunity, and $(1-\alpha)$ proportion is entirely dependent on wage employment (WE). In both households, the child's time is divided between outside work e and schooling $(1-e)$. The child wage is denoted as ω. The WE parents inelastically supply their labour to earn w each. The SE parents, on the other hand, split their time between wage employment (l) and self-employment $(1-l)$. The income from self-employment is given by $R(x)$ with $R'(x) > 0$ and $R''(x) < 0$, where $x = 1-l$. Thus, the total income of an SE parent is $y(w) = wl + R(x) = wl + R(1-l)$.

An important assumption is that the self-employed parents do not engage any other labour input either from home or outside. If other labour inputs were permitted, they could use their own children, and the model would have different implications. This case is being studied in Pal and Saha (2019).[3]

A set of competitive firms employ both child and adult labour as substitutes (the substitution axiom of BV) and their aggregate labour demand (expressed in adult unit) is given by $L^D = L^D(w, \omega)$ which is declining in both (w, ω). For both types of labour to be used, the relation $\omega = \gamma w$ must hold (the ridge-line equilibrium condition of BV).

Parents must meet a subsistence consumption $\underline{c}$ before they value education (the luxury axiom of BV). Thus, for both groups the objective function is same:

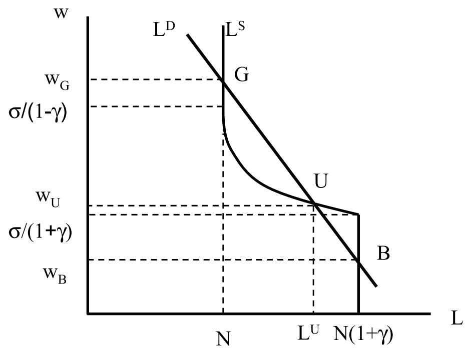
$$\begin{aligned} U &= (c-\underline{c})(1-e) \quad \text{if } c \geq s \\ &= (c-\underline{c}) \qquad \text{if } c < \underline{c}, \end{aligned} \tag{1}$$

where c is parent's consumption, and βc is child's consumption ($\beta < 1$).

[2]The assumption of one child is merely for simplicity. Unless one wishes to introduce other types of activities such as babysitting by an elder sibling, or issues like gender discrimination between children, adding more children to the model is unnecessary; the result will be unaltered.

[3]In their study based on the Indian Human Development Survey data of 2004–05, the authors find that 46% households (from a sample of 33,814 households) were self-employed, and from these households 11% children (aged 10–14) worked, while from other households 7% children worked. The difference in the proportions of child labour is statistically significant.

Fig. 1 Multiple equilibria



The budget constraint for the WE and the SE parents, respectively, is

$$w + e\gamma w \geq c(1+\beta), \tag{2}$$

$$wl + R(1-l) + e\gamma w \geq c(1+\beta). \tag{3}$$

Parents maximize (1) with respect to (e, c) or (e, c, l) (depending on being WE or SE) subject to their respective constraint (2) or (3).

The wage-employed household. A typical WE household's optimal child labour supply is given by

$$\begin{aligned} e_W &= 0 \qquad \text{for} \quad w > \frac{\sigma}{1-\gamma}, \\ &= \frac{1}{2} + \frac{\sigma - w}{2\gamma w}, \qquad \text{for} \quad w \in [\frac{\sigma}{1+\gamma}, \frac{\sigma}{1-\gamma}], \\ &= 1 \qquad \text{for} \quad w < \frac{\sigma}{1+\gamma}, \end{aligned} \tag{4}$$

where $\sigma = \underline{c}(1+\beta)$. If the adult wage exceeds (falls below) a high (low) wage level, child studies (works) full-time. At moderate wages, he works part-time, and it is easy to check that $e'_W(w) < 0$. The household's combined labour supply (in adult unit) is $z_W = 1 + \gamma e_W(w)$, which is also downward sloping at all $w \in (\frac{\sigma}{1+\gamma}, \frac{\sigma}{1-\gamma})$.

Now consider a special case, where all households are wage-employed. The economy's aggregate labour supply curve (expressed in adult labour unit) is $L^S(w) = N[1 + \gamma e_W(w)]$. Suppose $L^D(w)$ $(\equiv L(w, \gamma w))$ intersects $L^S(w)$ at three places as shown in Fig. 1. Two stable equilibria[4] occur at points G and B called 'good' and 'bad' equilibria, respectively. At G, the wage is high and no child works. At B, the wage is low and no child goes to school; they work instead. This is the BV model. Note that for both equilibria to coexist, labour supply must exceed labour demand between w_U and w_B.

[4]The unstable equilibrium U is discarded.

The self-employed household. We show that with a self-employment option, the excess labour supply between w_U and w_B can be eliminated. Consider an SE household's labour supply decision (l, e_S), given by (5)–(6):

$$- R'(1-l) + w = 0, \tag{5}$$

$$\sigma + \gamma w - y(l(w),\ w) - 2\gamma w e = 0. \tag{6}$$

The adult labour supply for wage employment $l(w)$ given by (5) is increasing in w (i.e. $l'(w) = -\frac{1}{R''(.)} > 0$), and for simplification assume $R'(1) = 0$ allowing us to set $l(0) = 0$.

We can now write $y(l(w), w) \equiv y(w) = wl(w) + R(1 - l(w))$. By differentiating $y(.)$ with respect to w, we obtain

$$y'(w) = \underbrace{[l + wl'(w)]}_{>0} - \underbrace{R'(1-l)l'(w)}_{>0}$$

As both terms are positive, as such the sign of $y'(w)$ is ambiguous. However, intuitive reasoning would help us conclude that $y(w)$ would be monotonically increasing as demonstrated in Fig. 2, or at least be flat at $R(1)$.

To understand why $y(w)$ would be non-decreasing, we first note that $l(w)$ maximizes the parent's total income $y(l(w),\ w) = wl + R(1 - l(w))$ with respect to l. Hence, his total income must be *at least* w or $R(1)$ because he can work *full-time* either outside or at self-employment and guarantee himself of w or $R(1)$, whichever is maximum. So given any w, say w_0, if he finds splitting his time between self-employment (say, $1 - l$ unit of time) and outside (l unit of time) a better option than being fully self-employed or being fully wage-employed, it must be that $y(l(w_0), w_0) > \max[R(1), w_0]$.

Now, starting from w_0 if w rises further, the parent can at least maintain the same allocation of time and earn more from wage employment. If indeed $l(w)$ is continuous and strictly increasing, then extra earning from wage employment will

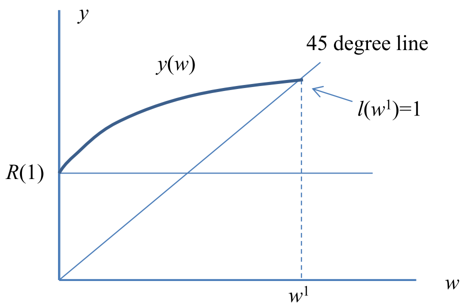


Fig. 2 Shape of $y(w)$

more than compensate for the loss in self-employment. Therefore, $y(w)$ must be increasing or will be just a flat line $R(1)$ at all w. In Fig. 2, we demonstrate the increasing case.

Observation 1 *(i) The self-employed parent's income will have the following lower bound:* $y(l(w), w) \geq \max[w, \ R(1)]$. *(ii) The self-employed parent's total income must be either increasing in w or at least be non-decreasing in w.*

A second implication of the self-employment option is that if $R(1)$—the minimum he can earn under any circumstances—is large enough to support the subsistence consumption σ, the child will not have to be sent to work. Of course, self-employment technology needs to be sufficiently productive. To consider a non-trivial case, assume $\sigma > R(1)$ so that providing schooling still depends on the equilibrium wage.

Now, consider Eq. (6). There are two critical wages. Set $e = 0$ and let $\bar{w}$ implicitly solve $\sigma + \gamma w - y(w) = 0$. At any wage above $\bar{w}$, child labour supply is zero. If the parent's total income covers the subsistence consumption and the foregone child income (γw), the child is kept at school full-time. On the other hand, if $e = 1$ and $\underline{w}$ implicitly solves $\sigma - \gamma w - y(w) = 0$, then we can say that at all wage below $\underline{w}$, the child will work full-time. Here, the combined income of the child and the parent hardly covers the subsistence consumption. More formally, the child labour supply of an SE household is

$$\begin{aligned} e_S &= 0 \qquad \text{for} \quad w > \bar{w} \\ &= \frac{1}{2} + \frac{\sigma - y(l(w), \ w)}{2\gamma w} \qquad \text{for} \quad w \in [\underline{w}, \bar{w}] \\ &= 1 \qquad \text{for} \quad w < \underline{w}. \end{aligned} \tag{7}$$

Child labour is non-increasing in w. In particular at $w \in [\underline{w}, \bar{w}]$,

$$\frac{\partial e_S}{\partial w} = \frac{1}{2\gamma w^2}\left[-\frac{\partial y(.)}{\partial l} l'(w)w - \{\sigma - R(1 - l)\}\right] = -\frac{\sigma - R(1)}{2\gamma w^2} < 0. \tag{8}$$

This is due to our assumption that $\sigma > R(1)$ and the fact that $y(.)$ is maximized by l. The combined labour supply curve of the SE household is $z_S = l(w) + \gamma e_S$. At $w \in [\underline{w}, \ \bar{w}]$, $z'_S(w) = l'(w) + e'_S(w)$ which is ambiguous in sign. But at $w > \bar{w}$ and $w < \underline{w}$, clearly $z'_S(w) = l'(w) > 0$. Then at all w that are close to, but still less than, $\bar{w}$ the curve must bend backward to turn upward sloping again. Figure 3a shows the household supply curve being backward bending. Figure 3b shows that if all households were SE, then we can have a unique market equilibrium where the adult wage is sufficiently high to support full child schooling.

For Fig. 3b to be valid, we must have that the SE households supply *less* labour than the WE households. The following observation confirms this.

Observation 2 *(i) An SE child switches to full schooling earlier and drops out of school later than a WE child. (ii) The total labour supply of an SE household is less, or no greater, than that of a WE household.*

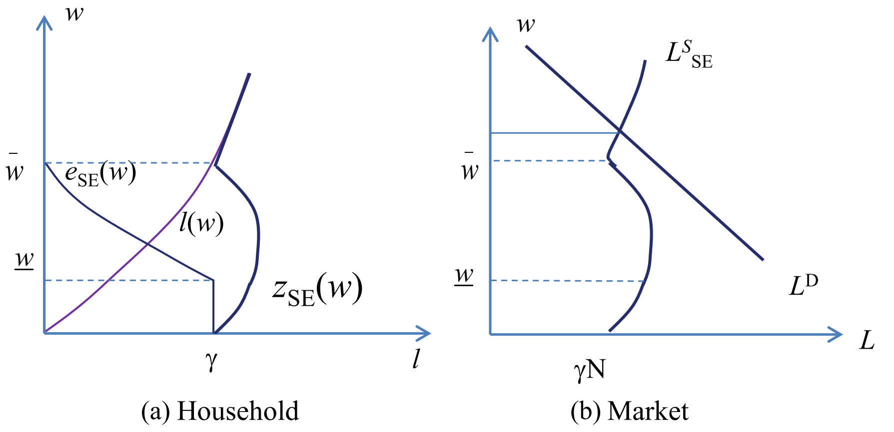


Fig. 3 Equilibrium with only self-employed parents

Part (i) of the above observation requires proving that $\bar{w} < \frac{\sigma}{1-\gamma}$ and $\underline{w} < \frac{\sigma}{1+\gamma}$ [refer to Eq. (4)]. To see this let us consider Eq. (7). At $w = \bar{w}$, we have $\sigma + \gamma\bar{w} - y(\bar{w}) = 0$, or $\sigma + \gamma\bar{w} + \bar{w} - \bar{w} - y(\bar{w}) = 0$, which yields

$$\sigma - [y(\bar{w}) - \bar{w}] = \bar{w}(1-\gamma),$$
$$\text{or}\quad \bar{w} = \frac{\sigma - [y(\bar{w}) - \bar{w}]}{1-\gamma} \le \frac{\sigma}{1-\gamma}.$$

This is because $y(w)$ must be no less than w or $R(1)$ whichever is maximum. Similarly, consider $\underline{w}$ at which we have $\sigma - \gamma\underline{w} - y(\underline{w}) = 0$, obtain

$$\sigma - [y(\underline{w}) - \underline{w}] = \underline{w}(1+\gamma),$$
$$\text{or}\quad \underline{w} = \frac{\sigma - [y(\underline{w}) - \underline{w}]}{1+\gamma} \le \frac{\sigma}{1+\gamma}.$$

For part (ii) of the observation, compare Eq. (4) with Eq. (7) for $w \in [\frac{\sigma}{1+\gamma}, \bar{w}]$. Since $w \le y(w)$, we have $e_W \ge e_S$. The SE parent's labour supply is $l(w) \le 1$. Hence, the SE household's labour supply never exceeds that of the WE household. In particular, $z_S < z_W$ at $w \le \frac{\sigma}{1-\gamma}$.

Elimination of the bad equilibrium. Now, consider the aggregate labour supply with mixed households: $L^S(w;\alpha) = N[\alpha z_S(w) + (1-\alpha)z_W(w)]$. In Fig. 1, we have $\alpha = 0$, and by construction, there are three equilibria; in Fig. 2b, we have $\alpha = 1$, and by construction, there is only one equilibrium. Then for $L^S(w;\alpha)$, we may have one or three equilibria depending on the magnitude of α. $L^S(.)$ is a declining function of α.

Let us consider all $w \in [0, \frac{\sigma}{1-\gamma}]$ and define an excess labour demand function $\phi(w;\alpha) = L^D(w) - L^S(w;\alpha)$. By construction, for $\alpha = 0$ there exists an interval of w such that $\phi(w) < 0$ over the entire interval, and for $\alpha = 1$ at all w ($\le \frac{\sigma}{1-\gamma}$), $\phi(w) >$

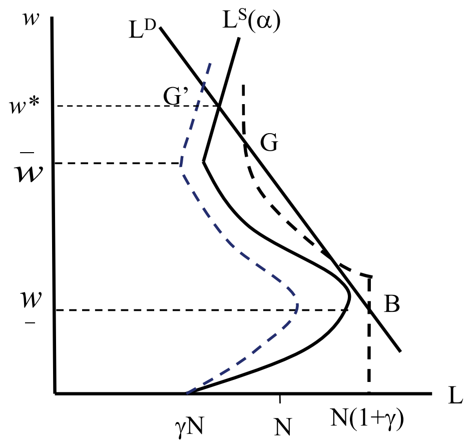


Fig. 4 Elimination of the bad equilibrium

0. Since $\phi(.)$ is a continuous and increasing function of α, by the intermediate value theorem there must exist a critical α, say $\hat{\alpha}$, such that $\phi(w; \hat{\alpha}) = 0$ at some w and strictly positive at all other $w \leq \frac{\sigma}{1-\gamma}$.

Figure 4 shows this critical α at which the labour supply curve swings back sufficiently to be just tangent to the labour demand curve at some wage between w_U and w_B where previously there was an excess supply. The result is that there is only one equilibrium which not only supports full schooling for all children but also improves the equilibrium wage to w^*, as the new equilibrium point G' is to the north-west of the old equilibrium point G.

Proposition 1 *If the self-employment opportunity $R(.)$ is made available to the $\hat{\alpha}$ fraction (or more) households, then the economy will have only the good equilibrium with child labour eliminated.*

3 Concluding Remarks

The policy implication of the result of this paper is dramatic. To eliminate child labour, all households need not be targeted for intervention. All we need is to intervene up to a critical proportion. A related question is how productive self-employment should be to have the required impact. Our model suggests that if $R(1) \geq w_B$ (i.e. the bad equilibrium wage), SE parents will withdraw from the market, which in turn would push the wage upward. From the literature on microfinance, it is seen that extensive provision of small loans can have a significant impact on household welfare including child schooling (see Morduch 1999) and access to microfinance makes self-employment sustainable (see Crepon et al. 2014). We argue that from the child labour perspective, a minimal intervention might just be sufficient.

There is also a caveat in order. A number of studies, such as Islam and Choe (2013) for Malawi and Hazarika and Sarangi (2008) for Bangladesh, and also the ongoing work of Pal and Saha (2019) for India have shown that children from the self-employed households are also more likely to work as child labour (and receive less education) primarily because of a substitution between home labour and outside labour triggered by an outside wage increase. This substitution has not been allowed in the present paper, which is clearly a limitation; but in reality, that possibility has to be taken into account. Therefore, our model has more relevance to those contexts where the self-employment activity is not scalable, i.e. it cannot be expanded by hiring additional labour, especially child labour. For example, a child is unsuitable to man a family-run shop, or cook in a family-run eatery. Our model is also applicable to where the microfinance intervention augments mainly the human capital of the beneficiary and the business is predominantly his/her skill based. It is in these contexts, credit interventions will likely have a positive effect of the kind that we established.

Acknowledgements I would like to thank an anonymous referee for helpful comments that led to improvements in the presentation of the model and discussion of empirical relevance. The remaining errors are mine.

References

Baland, J.-M., & Robinson, J. A. (2000). Is child labor inefficient? *Journal of Political Economy*, *108*, 663–79.

Basu, K. (2000). The intriguing relation between adult minimum wage and child labour. *Economic Journal*, *110*(462), C50–61.

Basu, K., & Van, P. H. (1998). The Economics of child labor. *American Economic Review*, *88*, 412–27.

Basu, K., Das, S., & Dutta, B. (2010). Child labour and household wealth: Theory and empirical evidence of an inverted-U. *Journal of Development Economics*, *91*, 8–14.

Bhalotra, S., & Heady, C. (2003). Child farm labor: The wealth paradox. *World Bank Economic Review*, *17*, 197–227.

Crepon, B., Devoto, F., Duflo, E., & Pariente, W. (2014). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. Massachusetts Institute of Technology: Discussion paper. http://economics.mit.edu/files/6659.

Gupta, M. R. (2000). Wage determination of a child worker: A theoretical analysis. *Review of Development Economics*, *4*, 219–28.

Hazarika, G., & Sarangi, S. (2008). Household access to microcredit and child laobur in rural Malawi. *World Development*, *36*(5), 843–59.

Islam, A., & Choe, C. (2013). Child labor and schooling responses to access to microcredit in rural Bangladesh. *Economic Inquiry*, *51*(1), 46–61.

Jafarey, S., & Lahiri, S. (2002). Will trade sanctions reduce child labour? The role of credit markets. *Journal of Development Economics*, *68*, 137–56.

Karlan, D., & Zinman, J. (2009). Expanding microentreprise credit access: Using randomised supply decisions to estimate the impacts in Manila. *Review of Financial Studies*, *23*(1), 433–64.

Menon, N., & van der Meulen Rodgers, Y. (2018). Child labor and the minimum wage: Evidence from India. *Journal of Comparative Economics*, *46*, 480–494.

Morduch, J. (1999). The microfinance promise. *Journal of Economic Literature*, *37*(4), 1569–1614.

Pal, S., & Saha, B. (2019). *Parental self-employment and child labour*. UK: Mimeo, Durham University Business School.

Ranjan, P. (2001). Credit constraints and the phenomenon of child labor. *Journal of Development Economics*, *64*, 81–102.

Swaminathan, M. (1998). Economic growth and the persistence of child labor: Evidence from an Indian city. *World Development*, *26*(8), 1513–1528.

Handloom in West Bengal: Decline or Dynamism?

Rajesh Bhattacharya and Sarmishtha Sen

Abstract In India, the relative importance of the handloom sector, one of the largest employers following agriculture, has been declining for the last few decades. The All India Handloom Census data for the year 2009–10, however, showed a rather modest decline in the number of weavers in West Bengal, in contrast to a 33% decline at the national level between 1995–96 and 2009–10. Data from the same Handloom Censuses also point to considerable occupational diversification among weaver households in West Bengal. Based on analysis of the same data, we find that recent history of handloom in West Bengal is not only of exit, but also of entry of labourers into the sector along with acquisition of skills by new entrants. Inter- and intra-sectoral labour mobilities call into question the dominant view of 'skills' in this sector, 'skills' being the singular lens through which the handloom sector has traditionally been viewed. We argue for a more labour-focused rather than a 'tradition'-centred approach to the handloom sector.

Keywords Handloom · West Bengal · Income diversification · Unorganized sector

1 Introduction

The handloom sector in India—with labour-intensive production using traditional technologies in tiny, often unregistered units —has long been the largest source

R. Bhattacharya (✉)
Public Policy and Management Group, Indian Institute of Management Calcutta, Kolkata, India
e-mail: rb@iimcal.ac.in

S. Sen
Department of Economics, Syamsundar College, University of Burdwan, Shyamsundar, West Bengal, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_17

of employment after agriculture (Government of India 2015).[1] Increased domestic and international demand for these products in the post-1991 period signified the market potential of these products (Roy 1998b; Mukund and Syamasundari 2001). But available evidence also points to meagre earnings for weavers and allied workers in the handloom sector as well as the increasing vulnerability of the craft producers, including handloom workers, in recent decades (Srinivasulu 1996; Liebl and Roy 2003; CCI 2015).

Handloom, the largest component of the craft-based sector, with traditional technology and institutions and having roots in the pre-colonial artisanal tradition, saw a sharp decline between the second and third Handloom Censuses, conducted in 1995–96 and 2009–10, respectively. Handloom sector employed 43.3 lakh workers on about 23.77 lakh handlooms in 2009–10, down from 65.5 lakh workers and 34.87 looms in 1995–96 (Government of India 2014b). Competition from powerlooms, outdated skills of the artisans and inability to adapt to the changing market environment are some of the commonly cited reasons for this decline (Mukund and Syamasundari 2001; Subrahmanya 2011).[2]

In independent India, policy-making on handlooms has often oscillated between two positions—a 'protectionist' stance that sought to preserve and nurture the sector for its traditional skills or employment-generating potential or both, and a 'transformist' stance that viewed the presence of handloom sector on a large scale as an anomaly and argued for its modernization, including transformation into powerlooms (Mamidipudi 2016). The received wisdom is that most of the textile policies in India—in a bid to protect handloom production from the vagaries of market forces—have helped accumulate inefficiency in the entire structure of production and distribution. Starting with the textile policy of 1985, there was a distinct shift in policy emphasis from protection and employment creation to output growth and export promotion. But the post-liberalization period has also seen the largest decline in handlooms, either because of transition from handloom to powerloom or exit from handloom (NCAER 2010).

However, the simple story of secular decline in handloom and inexorable rise of powerloom in the post-liberalization phase is not uniform across regions and even over time.[3] For example, the total volume of cloth production in the powerloom sector

[1]The number of handloom and all handicrafts workers in 2014 were 4.33 million and 6.89 million, respectively, in 2009–10 (Government of India 2014b). In the Indian context, hand loom means any loom, other than power loom and includes any hybrid loom on which at least one process for weaving requires manual intervention or human energy for production. See https://economictimes.indiatimes.com/industry/cons-products/garments-/-textiles/no-change-in-definition-of-handloom-textile-ministry/articleshow/28600164.cms.

[2]The export market for India's handloom products has not been stable. India's exports of handloom products increased sharply from Rs. 1252.80 crores in 2009–10 to Rs. 28,111.97 crores in 2012–13 after which it has come down to Rs. 2392.21 crores in 2016–17 (Government of India 2014b, 2017b).

[3]Even though the number of looms and workers in the handloom sector declined between 1995–96 and 2009–10, cloth production increased from 3120 to 6930 million square metres over the same period. The number of man days worked, the proportion of full-time weavers and the proportion of weavers who derive more than 60% of income from weaving increased and the proportion of idle

as well as its share in total cloth production has fallen since 2012–13 (Government of India 2017b), while the same in the handloom sector has registered a rise over the same period. West Bengal and north-eastern states of India show an increase in the number of weaver households between the 1995 and 2010 (NCAER 2004, 2010). There has been very little growth of powerlooms in handloom-major states like Orissa and West Bengal till recent times. The Handloom Census of 2009–10 also brings out some of these differences sharply.

West Bengal, the focus of this essay, stands out among Indian states with the highest number of looms and highest number of weaver households according to the third handloom census. The persistence of the handloom industry in West Bengal seems puzzling given the dominant narrative of a steady and secular decline of this sector over a long period of time in India. This calls for a closer look at the handloom sector in West Bengal—particularly with respect to the condition of handloom weavers. Specifically, we need to ask whether the nearly constant size of the handloom sector in West Bengal hides significant changes in the composition of the weavers brought about by simultaneous entry and exit of weavers, as observed, for example, in the handloom industry of Andhra Pradesh (Mukund and Syamasundari 2001; Annapurna et al. 2012). Further, do the intensity of engagement of weavers in handloom production and their degree of dependence on the sector remain constant over the same period, during which the overall size appears to remain unchanged? We seek to examine these aspects with the help of aggregate data on the handloom sector obtained from the All India Handloom Censuses of 1995–96 and 2009–10.

Following this introductory section, we search the literature for insights on the interrelation between survival of handloom as a livelihood and weavers' ability to cope up with market shocks and stresses based on skill development (Sect. 2). In the subsequent two sections (Sects. 3 and 4), we present findings on selected aspects of the handloom sector in West Bengal, based on the second and third Handloom Censuses, and compare them with those for other handloom-major states in India. Section 5 discusses those findings on handloom in West Bengal in the context of its overall economic structure. Section 6 concludes by revisiting the idea of handloom in dominant discourses and argues for a fresh approach in research.

2 Existing Literature and Scope of Present Study

Significant, even if diminishing, presence of handloom has belied all past predictions of steady and inevitable decline of this craft with the growth of the modern sector aided by incessant technological progress. The narrative of the destruction of handicraft industry, especially handlooms, during the colonial period was subsequently

looms decreased over the same period (Government of India 2014b). The total cloth production on handlooms increased to 7638 million m^2 in 2015–16 accounting for 15.3% of India's total cloth production. Around 95% of world's hand-woven fabric comes from India.

revised as a story of resilience exhibited by the handloom producers[4] (Mukund 1992). This strand of literature, while questioning the uniformity of decline supposedly experienced under deindustrialization,[5] privileged structural reorganization and adaptation by innovation, rather than increased self-exploitation by Indian weavers, as responsible for continued survival of handlooms (Roy 2009).

This revised understanding of the history of handloom in India radically changed the way in which handloom is viewed. Handloom has been often represented as an archaic system of traditional production with a conservative workforce reluctant to introduce technological change and to learn. In contrast, in the new understanding, the underlying system of handloom production is portrayed as more resilient, having at its disposal a skilled workforce capable of responding to market competition by offering a far diverse product basket—in comparison with those offered by mills—due to their vastly differentiated techniques of production (Chaudhury 2014; Mukund 1992; Srinivasulu 1996).[6] In recent decades, handloom has exhibited adaptation of a different nature by embracing production of items from outside respective regional specialties. Market demand rather than supply conditions dictate production in this scenario—'…what were regionally specific products are now being produced in a number of distant centres across regions' (Niranjana and Vinayan 2001). Two factors could be identified from this literature as crucial to the sustenance of the sector: (i) the ability to cope with the shocks and stresses, aided by a dynamic repository of skills capable of adjusting continuously and (ii) an enabling institutional structure.

Responsibility for creating a favourable institutional mechanism was with policymakers in the immediate post-independence years, while in the post-reforms period, market actors were expected to play a greater role in imparting dynamism to the sector. The spirit of optimism emanating from limited success of protective handloom policies[7] of the immediate post-independence period gave way to doubts on the effectiveness of, and thus, rationale for continuing such interventions in later years, expressed particularly in the 1985 Textile Policy. This policy was marked by an underlying acceptance of an inevitable death of the sector. Volume of output, productivity/efficiency and profitability of the producing units became the issues of utmost concern. It relied particularly on the mechanized segments for growth of domestic textile production. Characterization of handloom as outdated and per-

[4]This aspect of colonial history of India, known as the 'deindustrialization thesis', concentrated on the distress of artisans resulting mainly from trade-related policy discriminations and internal measures of repressions by the colonial government in India.

[5]The revisionist history argued that (i) it is the eastern part of the country, especially undivided Bengal and Odisha, that heavily bore the brunt of this downfall, (ii) the shift of operations to *karkhana* system and to the employment of full time handloom workers led to perceptible rise in productivity and income in other parts of the country and (iii) hand-spinning, rather than hand-weaving was affected most.

[6]Some have argued, however, that the technological changes are not absent in the handloom sector but, 'remain scarce in relation to the scale of the industry' (Liebel and Roy 2003), and also that this dynamism in the handloom includes transition from handloom to powerloom (Roy 1998a; Liebel and Roy 2003).

[7]Policy measures attempted to raise production efficiency, through interventions related to production and distribution networks and some organizational reforms (e.g. cooperativization).

petually in a state of decline in successive textile policies of later years resulted in starvation of resources pumped into the sector. Allocation of paltry budgetary resources to KVIC or to the programme of cooperativization reinforced the tendency towards the sector's fall (Das 2001).

Hope for a market-driven revival of Indian handloom, however, came under doubt, as stories of suffering and out-migration by the handloom artisans accumulated in the post-reforms period. Differences in the experience at individual and industry levels called for more micro-level analyses to capture heterogeneity of the sector. Need for empirical elaboration of this heterogeneity was acknowledged by scholars—'there is no such thing as "the" weaver, but rather a diversity of conditions that characterize weavers and weaving' (Niranjana and Vinayan 2001). Thus, recent scholarship has shifted the focus of studies on handloom to the handloom weavers and the reality they lived in. What figures most in contemporary analyses is the understanding of (i) handloom weavers as innovators and (ii) inward as well as outward movements of weavers as part of their household strategies for achieving sustainable livelihood.

Whether or not the process of skill generation and innovations can expand or at the least preserve the market for handloom products would depend on the match between the set of skills required for market success on the one side and the aspirations of the emerging class of weavers and expected benefits from such skilling on the other. According to the currently dominant view on the sustainability of the craft in India, handloom should be understood as an 'ensemble of knowledge, skills, technology and social relations' and the weaver as a socio-technologist (Annapurna et al. 2012).

The changing composition of weavers' population, their socio-economic locus, position in the production–distribution network—all of which influences this process of skilling and the craft producers' capability to earn a living—are relatively understudied. In this paper, we seek to attempt a macro-level exploration of such changes in composition and intensity of engagement of weavers based on available secondary data. The present study addresses the following two research objectives.

(1) What are the changes in the composition of the weaving community and in the nature of their engagement with the handloom in West Bengal, compared to other handloom-major states?
(2) What are the trends in handloom production in West Bengal, compared to other handloom-major states?

Specifically, we attempt to understand the changes in the socio-economic profile of the weavers and in the extent and nature of their involvement with handloom work at present. We also try to examine the trends in production—type of products, raw materials, etc.—to form an idea of the vibrancy in the handloom sector in West Bengal. This work departs from other such studies mainly in two respects: (i) handloom work is viewed from the perspective of the producers, who may have approached the work as a livelihood option rather than a continuation of tradition and (ii) the analysis of handloom sector is located in the story of broader structural transformation of the West Bengal economy.

In this essay, we have used mainly two rounds of Handloom Censuses conducted in India: All India Joint Census of Handlooms and Powerlooms, 1995–96 and the

Third All India Handloom Census 2009–10.[8] In addition, we have also used the first Handloom Census, 1987–88 for changes in the overall size of the sector in selected Indian states. The census reports provide data on (i) socio-economic characteristics of the handloom worker households and individual weavers, (ii) the extent of engagement of handloom workers in the sector and (iii) aggregate data on production and productivity. We undertook a comparative analysis—by placing the changes in West Bengal against those in other handloom-major states over the time period between the two Censuses.

3 Profile of Handloom Weavers in West Bengal

The specificities of the handloom sector in West Bengal (WB hereafter) is best appreciated in a comparative account of the handloom workers' socio-economic profiles and various aspects of production across major handloom producing Indian states and the average picture at all-India (AI) level. We compare WB with Andhra Pradesh (AP), Odisha (OD), Tamil Nadu (TN) and Uttar Pradesh (UP) as these are states with relatively large presence of handloom households.[9]

3.1 Composition of the Weaving Community

The handloom sector in West Bengal exhibited a pattern that is markedly different from that of its comparators. Table 1 shows the changes in a number of looms, across the first three Handloom Censuses (1987–88, 1995–96 and 2009–10), in selected states of India with predominantly commercial production. West Bengal not only had the highest number of handlooms in India in 2010, but also exhibited little decline between 1988 and 2010; in fact, the number of looms increased between 1988 and 1996. For other selected states and India as a whole, the decline in number of looms is substantial and linear over the same period. The number of handloom weaver households in WB has gone up from 2.2 lakh in 1995 to 2.4 lakh in 2010, while the AI total has gone down from 25.3 lakhs to 22.7 lakhs over the same period—WB's

---

[8]We did not use the National Sample Survey reports on the unorganized manufacturing sector for this purpose since the sample of handloom units producing fabric would be very small at the state level.

[9]The percentage shares of these five states—WB, TN, AP, UP and OD—in total number of handloom weaver households in the country are 10.75, 7.2, 5.85, 3.72 and 1.47, respectively. Seven north-eastern states, despite accounting for more than 55% of these households are excluded from this analysis because a very significant proportion of looms in these states are operated non-commercially as a 'part of the culture across all social groups' and primarily to meet domestic fabric requirements. Weaving and other handloom activities in the selected states, on the other side, are guided primarily by commercial considerations. Alternatively, we can use the proportion of all adult handloom workers and full-time weavers as criteria of selection. The selection of states remains the same under the alternative criterion.

Table 1 Number of looms in selected states of India: 1988–2010

States/AI	Looms owned (#)			Percentage change	
	1988	1995	2010	1988–1996	1996–2010
AP	219,715 (5.81)	202,100 (5.8)	124,714 (5.25)	−8.02	−38.29
OD	119,005 (3.15)	92,869 (2.66)	43,652 (1.84)	−21.96	−53.00
TN	428,545 (11.34)	413,174 (11.85)	154,509 (6.5)	−3.59	−62.60
UP	260,714 (6.9)	189,570 (5.44)	80,295 (3.38)	−27.29	−57.64
WB	338,499 (8.96)	350,654 (10.06)	307,829 (12.95)	3.59	−12.21
AI	3,778,583 (100)	3,486,308 (100)	2,377,331 (100)	−7.73	−31.81

Note Figures in parenthesis are the percentage shares of the state in total looms in the country
Source NCAER (2010)

Table 2 Handloom worker households by social groups (in percentage): selected states of India, 1995 and 2010

States/AI	1995				2010			
	SC	ST	OBC	Others	SC	ST	OBC	Others
AP	1.59	0.38	93.61	4.42	0.96	0.53	90.86	7.64
OR	28.11	3.35	63.5	5.04	13.55	3.16	82.28	1.01
TN	4.64	1.28	91.24	2.84	2.35	1.48	77.97	18.21
UP	16.15	3.35	76.39	4.11	9.93	1.22	78.68	10.16
WB	23.37	1.5	37.04	38.07	25.22	0.78	11.41	62.59
AI	10.76	25.5	42.65	21.1	9.81	22.11	40.92	27.16

Source NCAER (2004, 2010)

share of handloom weaver households in the country thus increasing from 8.7 to 10.57% (NCAER 2010).

Handloom continues to be predominantly a rural activity even in 2010.[10] However, the caste composition of handloom workers has been undergoing changes all over India for quite some time (Mukund 1992; Das 2001; Khasnabis and Nag 2001). Table 2 shows that handloom worker households (i.e. weaver households plus households with members engaged in allied work in handloom) were predominantly from the OBC castes in all the selected states in 1995. In WB, handloom worker households belonging to OBC caste groups constituted a significant percentage (37.04) of the total, but Scheduled Castes and Other Castes were significant too, accounting for 23.37 and 38.07%, respectively, in 1995. However, in 2010, while in other states, handloom continued to be dominated by OBC castes, in West Bengal, Other Castes accounted for 62.59% of handloom worker households, with the share of OBC castes sharply falling to 11.41% and that of SC castes marginally increasing to 25.22%. While the increase in the share of Other Castes is an all-India trend and is also evident

[10]For India as a whole, 87.5% of handloom weaver households were rural in 2010. The figure drops to 70.1% if we exclude the seven north-eastern states.

in all the selected states, except OD, in no other state has the change been so dramatic as in WB. Therefore, arguably, WB has seen the entry of members from traditionally non-weaving castes in the handloom sector on a significant scale between 1995 and 2010.

The third Handloom Census of 2010 was the first to gather data on religion of handloom worker households—therefore, a comparison across census years is not possible. One field-based study—in nine districts of WB with highest concentrations of handloom, around the same time as the 1995–96 census—reported that handloom weaving was predominantly undertaken by Hindu households (Das 2001). However, census data for 2010 revealed that more than a third of handloom worker households in West Bengal were Muslim households. In this respect, WB was second only to UP, where Muslims formed the traditional skill base of the industry (Table 3). Thus, the increased engagement of Muslim handloom worker households in West Bengal can be inferred to be a recent phenomenon.

A third trend is the feminization of handloom weaving on a significant scale in West Bengal. This trend of greater participation of women in handloom work is observable at the all-India level as well as in all the selected states—as (a) increasing share in full-time employment, (b) declining share in part-time employment and (c) increasing share in total employment (except for UP) of women in handloom work (see Table 4). But this trend is most pronounced in WB, with only TN coming close. Women traditionally did play a significant role in preparatory and allied works related to handloom weaving, but the current trend seems to suggest a deeper engagement of women with handloom production.

Thus, there appears to be a more significant churn in the social composition—in terms of gender, caste and religion—of the handloom worker households in West Bengal between 1995 and 2010, compared to other handloom-major states, even as the overall size of the sector did not change in WB as drastically as in other states.

There is no clear evidence that households persist in the handloom sector in WB only out of economic distress. From Table 5, we can see that the proportion of handloom worker households with Antyodaya Anna Yojana (AAY) or Below Poverty Level (BPL) ration cards in WB was lower than AI average and lowest among all selected states, except UP. Similarly, proportion of handloom worker households

Table 3 Handloom worker households by religion: selected states of India, 2010

States/AI	Religion		
	Hindus	Muslims	Others
AP	94.75	4.11	1.1
OR	97.64	0.05	2.3
TN	97.29	0.85	1.9
UP	14.5	85.37	0.1
WB	63.27	36.6	0.1
AI	77.62	15.04	7.3

Source NCAER (2010)

Table 4 Female employment (as per cent of total employment) in handloom weaving in selected states of India, 1995 and 2010

States/AI	1995			2010		
	Full time	Part time	Total	Full time	Part time	Total
AP	23	55.36	28.94	27.56	48.22	29.07
OR	8.67	32.83	14.04	13.88	25.78	15.41
TN	24.18	77.54	28.69	37.49	67.02	40.29
UP	17.98	51.98	23.38	20.13	42.18	21.79
WB	22.07	59.78	28.5	41.78	46.53	42.5
AI	37.67	87.02	63.25	65.57	93.42	75.74

Source NCAER (2004, 2010)

Table 5 Handloom worker households by type of ration card and average income of weaver households, selected states of India, 2010

States/AI	Handloom worker households				Handloom weaver households
	AAY	BPL	APL	No ration card	Average annual earnings (Rs.)
AP	3.8	87.93	1.31	6.96	30,054
OD	6.98	45.02	26.9	21.1	30,313
TN	1.97	91.65	5.02	1.37	24,707
UP	5.64	13.88	72.86	7.62	23,218
WB	8.87	32.3	54.1	4.73	26,934
AI	9.73	36.89	34.47	18.91	37,704

Source NCAER (2010)

with Above Poverty Level (APL) ration card in WB was higher than the all-India average and higher than all selected states, except UP (NCAER 2010). Handloom weaver households in West Bengal on average earned less annually than weaver households in AP and OD, but more than their counterparts in TN and UP.

Thus, there is no clear evidence that weaver households are particularly worse off in WB compared to other handloom-major states with commercial production. We argue that the pronounced changes in social composition of handloom workers in WB may reflect (a) mobility of labour across sectors as well as (b) income diversification strategies of handloom worker households.

3.2 Intensity of Engagement of Households in Handloom and Household Income from Handloom

It is possible from the third Handloom Census of 2010 to form an idea of the intensity of engagement of weavers and allied workers in handloom production and possession of looms by weaver households. Table 6 shows that there was a sharp fall in idle looms between 1995 and 2010 at AI level and in all the selected states—except in WB where idle looms increased by 28.9%. On the other hand, active looms too declined significantly in all the selected states over the same period, but the decline in WB was smaller compared to AI average and other selected states. Table 6 also shows that percentage of loom-possessing weaver households decreased sharply in all selected states between 1995 and 2010, with the sharpest falls registered in AP and WB. Thus, loomlessness among weaver households in WB has been increasing. This could be due to labour mobility across sector—e.g. due to the entry of weavers from traditionally non-weaving communities as well as due to employment of loomless weavers on household looms, while members of loom-owning households themselves move to non-handloom occupations or to trading operations within the handloom sector.

The average number of person days worked by a weaver in a year was lowest in WB in 2010 among the selected handloom-major states in 2010 (Table 7). The average number of person days of handloom work (weaving as well as allied work) per household was also lowest in WB. Table 7 also shows that the proportion of adult weavers engaged part-time in handloom in WB is highest among the selected handloom-major states with commercial production. This implies that unlike other states, WB does not exhibit clear exit from handloom, but calibrated and partial engagement with the sector, possibly as a result of household diversification of income and reallocation of household labour, including feminization of handloom work. Income diversifica-

Table 6 Change in looms by status during 1995–2010 and possession of looms by weaver households, 1995 and 2010: selected states of India

States/AI	Active		Idle		Percentage change in looms during 1995–2010		Percentage of handloom weaver households possessing looms	
	1995	2010	1995	2010	Active	Idle	1995	2010
AP	183,093	117,745	19,007	6969	−35.69	−63.33	94.02	56.64
OD	78,667	38,637	14,202	5015	−50.89	−64.69	98.22	94.24
TN	360,474	149,442	52,700	5067	−58.54	−90.39	87.24	64.38
UP	157,666	76,174	31,904	4121	−51.69	−87.08	92.24	68.60
WB	319,807	268,068	30,847	39,761	−16.18	28.9	96.00	71.33
AI	3,137,138	2,146,432	349,170	230,899	−31.58	−33.87	95.19	81.66

Source NCAER (2004, 2010)

Table 7 Intensity of engagement in handloom work and share of income from handloom in household income: selected states of India

States/AI	Average number of person days worked per weaver per year	Average days of handloom work per household	Adult weavers (18 years and above years) with part-time engagement in handloom (per cent of total)	Share of handloom income in total household income (in per cent)	
	2010	2010	2010	1996	2010
AP	274	462	7.3	82.89	61.73
OD	249	595	12.9	70.94	75.22
TN	259	427	9.5	80.49	79.42
UP	263	506	7.5	80.73	85.13
WB	241	377	15.2	72.18	42.05
AI	183	264	36.5	38.96	30.18

Source NCAER (2004, 2010)

tion of handloom households in WB is clear from the last two columns of Table 7. Share of handloom income to total household income in WB has sharply fallen from 72.18% in 1995 to 42.05% in 2010. While the figure for WB was comparable to other selected states in 1995, it is lower than other selected states by a wide margin in 2010.

4 Decline or Dynamism: Handloom Production in West Bengal

The evidence presented in the preceding section points to the possibilities of (a) significant changes in the social composition of the handloom weavers, due to entry and exit of labour in the handloom sector, (b) income diversification of handloom households leading to a reduced share of handloom income in total income and consequently and (c) a decline in intensity of engagement of weavers in handloom work as evident from the number of idle looms and number of person days worker by a weaver in a year on average. What do these trends imply for volume and quality of products?

Table 8 shows that WB does not seem to lag behind selected handloom-major states in terms of reported average daily production per weaver of major fabric produced by the household in 2010. Proportion of households with average daily production of less than 2 m of major fabric per weaver is lowest in WB compared to other handloom-major States. WB has the highest proportion of households producing 2–4 m of fabric per day per weaver among the selected households and has the

Table 8 Some aspects of handloom weaving in selected states of India, 2010

States/AI	Proportion of households (in per cent) by average production of major fabric per weaver per day					Proportion of households (in per cent) using cotton by Yarn count		
	<1 m	1–2 m	2–3 m	3–4 m	4 m and above	1–40	41–80	Above 80
AP	51.44	22.32	12.62	9.59	4.03	20.21	58.41	21.38
OD	22.82	25.88	26.79	7.67	16.83	66.21	23.02	10.77
TN	35.36	22.60	19.16	11.28	11.61	41.34	31.02	27.64
UP	50.68	7.58	1.90	2.36	37.13	91.92	6.32	1.76
WB	17.40	14.54	30.42	22.73	14.91	26.6	13.56	59.84
A0I	45.92	32.29	10.59	4.73	6.48	41.16	30.59	28.24

Note Length in metres
Source NCAER (2010)

third highest proportion of households with production above 4 m per weaver per day, ahead of TN and AP, but behind UP and OD.

The last three columns give distribution by yarn count of households using cotton yarn in 2010. In WB, a little more than a quarter of households using cotton used yarn up to 40 counts for production, whereas almost 60% of weaver households used yarn of higher counts (above 80). Other states appear to specialize in yarn count up to 40 (OD and UP) or in yarn count of 41–80 (AP) or have significant household concentration in yarn counts of both 1–40 and 41–80, without any pronounced specialization in a single range of yarn count (TN). Among the selected states, TN has the second highest proportion of households producing with yarn count above 80, second to, but much below the figure for WB.

This is a significant development, since only 12.35% of household units in West Bengal used yarn of over 80 counts in 1996 (NCAER 2004); these figures were comparable to other handloom-major states in 1996. Since the usage of higher count cotton implies finer fabric, it is possible that weaver households have tried to survive, in the face of competition, by moving on to the production of finer cloth that the mill and powerlooms could not imitate easily—a competitive strategy highlighted by Roy (1998b) and Das (2001) among others. Thus, occupational diversification of handloom weaver households and increased labour mobility in the handloom sector in WB does not mean that average daily production per weaver and quality of handloom products is lower other selected states with much smaller handloom sector. Thus, persistence of handloom on a large scale in WB cannot be attributed to survivalist objectives of weavers, but appear to be driven by competitive spirit as well. It is possible that those specializing in lower counts are producing for the local markets and driven mainly by survivalist objectives, but it is equally true that a larger proportion of households specializing in higher yarn count are producing a different set of products for distant markets.

How do the weavers look at the future of handlooms? Table 9 shows the distribution of households by their perception of 'threat' from mills and powerlooms and their perception of children willing or not to continue with handloom production, as reported in 2010. Among the handloom-major states with commercial production of handloom, WB had the lowest proportion of households (16.17%) who perceived the 'threat' from mills and powerlooms to be high and a remarkably high proportion of households (62.92%) who said there was either no 'threat' or they were unaware of any 'threat' from mills and powerlooms. In this respect, West Bengal is strikingly different from other selected states. It could be due to the low penetration of powerlooms in West Bengal, but it could also be due to their specialization in higher yarn count and finer variety product, which enables them to compete with powerlooms. From Table 9, we can also form an idea of whether households believed their children will continue with handloom work or not. Except OD, all other selected handloom-major states showed a lower share of households who believed their children will continue with handloom work compared to the national average—OD which had a sharp decline in handloom sector over the period 1988 to 2010 had the highest proportion of households (70.33%) who believed their children will continue with handloom work. WB, however, had the second lowest proportion of households (21.24%) who believed their children will not continue with handloom work, second to OD and much lower than other selected states. Almost 40% of households in WB did not know whether their children will continue with handloom work and a strikingly high percentage of households (23.2%) found the survey question inapplicable.

These findings regarding household perceptions resonate well with other findings related to the handloom sector in WB presented before. The facts that (a) WB has a relatively high percentage of looms that are idle, (b) there is evidence of decline in intensity of engagement of weavers in handloom activity and (c) handloom's contribution to total household income is declining over time paint a picture of decline of the sector in WB. On the other hand, increasing specialization in higher yarn count, lower perception of threat from mills and powerlooms and significant changes in social composition of handloom households—by caste, religion and gender—involving both labour mobility across sectors as well as occupational diversification and reallocation of household labour, point to dynamic aspects of handloom sector in WB. In such a scenario, perceptions about the future of handloom are likely to be less clear-cut than in those states where there are unmistakable signs of either decline or dynamism. The evidence in WB, as we have seen, is mixed.

5 Handloom in the Context of Economic Trends in West Bengal

We argue that for a nuanced understanding of the state of handloom sector in West Bengal, it must be studied in the overall economic context of West Bengal. While growth rates of gross state domestic product and per capita gross state domestic

Table 9 Some perceptions of handloom households: selected states, 2010

States/AI	Household perception about threat from mill/powerloom sector (per cent of households)					Household belief about children (per cent of households)				
	High	Moderate	No threat	Don't know	Total	Continue with handloom	Not continue with handloom	Don't know	Not applicable	Total
AP	84.6	8.37	3.53	3.5	100	10.88	47.39	32.19	9.54	100
OD	59.13	31.27	4.76	4.84	100	70.33	9.35	17.44	2.88	100
TN	46.5	14.24	27.44	11.82	100	9.17	56.96	25.59	8.28	100
UP	33.89	37.44	19.16	9.51	100	23.68	31.73	38.27	6.32	100
WB	16.17	20.91	20.24	42.68	100	16.37	21.24	39.19	23.2	100
AI	16.77	16.54	44.04	22.65	100	25.28	24.49	35.67	14.56	100

Source NCAER (2010)

product have been lower in West Bengal compared to other states in recent times, they are not unsatisfactory in absolute terms. It is the nature of economic growth in West Bengal that is problematic—since it is driven by the unorganized sector. Agriculture's share is declining in both state domestic product as well as total workforce of the state. But this structural transformation is not accompanied by rising agricultural incomes. While the service sector has expanded its share in state domestic product, it has not absorbed the workers released from agriculture, which have mainly sought employment in the manufacturing sector. Manufacturing sector in West Bengal has, thus, increased its share of workers, while at the same time, its share of output has gone down. Overall, therefore, there is a decline in manufacturing output per worker, while the opposite is true for services (Chowdhury 2016).

Both services and manufacturing in West Bengal are dominated by unorganized sector enterprises. The organized sector in West Bengal has shrunk relative to other states in India and by absolute number of workers as well. Labour productivity in both organized and unorganized sectors is lower in West Bengal compared to India. From Table 10, we can see that, for unorganized manufacturing enterprises in WB, the annual gross value-added per worker was Rs. 3354 per month, while the corresponding figure for all unorganized enterprises, both in manufacturing and services, was Rs. 5275 per month, in 2015–16. The figures for WB are much below the national averages of Rs. 6198 and Rs. 8645 per month.

The average size of unorganized non-agricultural enterprises—as evident from annual gross value-added per enterprise in WB—is also very low. For all enterprises in unorganized manufacturing and services sector, the monthly gross value-added per enterprise was less than Rs. 8000, while for unorganized manufacturing enterprises it was Rs. 5587 per enterprise, in 2015–16. The corresponding figures at all-India level were Rs. 15,159 and Rs. 11,359 per enterprise per month in the same year—WB's unorganized sector enterprises were on average less than half the size of the average firm at the national level. WB's economy is thus dominated by vast numbers of tiny, uneconomic enterprises with low levels of labour productivity and low incomes.

Table 10 Unorganized non-agricultural enterprises (excluding construction) in 2015–16: West Bengal and All-India

Sectors	West Bengal's share of total enterprises (%)	West Bengal's share of total workers (%)	Annual GVA per enterprise (Rs.)		Annual GVA per worker (Rs.)	
			West Bengal	India	West Bengal	India
Manufacturing	21.25	19.3	67,055	136,317	40,258	74,379
Trading	9.81	8.46	137,340	194,877	94,712	115,885
Other services	11.75	9.09	109,914	210,860	80,775	119,947
All	13.99	12.18	96,686	181,908	63,299	103,744

Source Government of India (2017a)

The tiny size and low productivity of these unorganized sector firms in WB point to subsistence-level income for the workers engaged in them.

In terms of per capita income, West Bengal has steadily fallen behind other states in India. Comparison of West Bengal with other states in terms of average monthly per capita consumption expenditure throws up a similar picture. West Bengal's performance in agriculture is unimpressive compared to other states, particularly since 2002–03. Average monthly income per agricultural household in West Bengal was Rs. 2079 in 2002–03 when the all-India figure was Rs. 2115. In 2012–13, the figures were Rs. 3940 for West Bengal and Rs. 6426 for the whole of India (Government of India 2014a). In West Bengal, small and marginal farms make up 95.92% of agricultural landholdings by number (third highest among larger states in India in 2010–11, behind Bihar and Kerala) and 80.72% of agricultural landholdings by area (highest in India). The uneconomic size of landholdings and lack of growth of organized sector employment have depressed per capita income in West Bengal and fuelled explosive growth of the unorganized sector. Table 11 shows that wages in both unorganized manufacturing and unorganized services in West Bengal are among the lowest among all states.

The story of handloom in West Bengal, is, therefore, a part of the overall story of economic transformation in West Bengal—where large units have given way to small units, the organized sector to the unorganized sector. This shows up in productivity,

Table 11 Annual wage rates in major Indian States, 2015

States	Annual wage rate in unorganized sector (2015 Rupees)		
	Manufacturing	Services	Organized manufacturing
Andhra Pradesh	59,319.6	87,016.3	152,125.1
Assam	44,950.8	53,267.5	82,833.1
Bihar	56,640.1	64,853.3	75,517.8
Delhi	98,961.2	100,399.0	148,287.4
Gujarat	87,748.3	76,366.0	144,954.4
Haryana	74,105.2	125,297.7	139,763.6
Himachal Pradesh	77,022.6	78,322.9	142,090.9
Karnataka	87,200.8	99,083.2	158,743.6
Kerala	112,357.0	124,101.7	141,810.7
Madhya Pradesh	53,772.8	67,695.2	139,900.1
Maharashtra	91,083.2	99,048.0	177,206.9
Orissa	54,836.5	63,771.9	175,238.2
Punjab	73,818.8	70,069.7	115,693.0
Rajasthan	83,958.6	80,308.0	139,168.3
Tamil Nadu	79,195.3	85,861.6	119,742.8
West Bengal	58,738.6	56,960.2	138,379.2

Source Azim Premji University (2018)

income and wage figures across all sectors of the income. The story of decline or dynamism of any sector, including handloom, cannot be understood unless it is placed in the overall economic context of West Bengal. This brings us to the crucial question—is the handloom sector in West Bengal a repository of artisanal skills or is it a livelihood sector like other unorganized manufacturing sectors that appear to be expanding in West Bengal?

6 Conclusion

The foregoing analysis highlighted some distinct trends in the handloom sector in WB when compared to other handloom-major states with commercial production in India. First, there is a significant change in the composition of handloom workers, by caste and religion, reflecting entry of new members from traditional non-weaving communities. There is also evidence of greater engagement of women in household handloom work pointing to a feminization of handloom work. Second, there appears to be a trend towards a calibrated and a more partial engagement of workers in handloom work—as evident in lower average number of days worked per year per weaver compared to other states as well as a sharp fall in the share of handloom income in total household income in WB. Third, handloom sector in WB does not appear to be particularly worse off in terms of average output per weaver per day, quality of fabric as captured by cotton yarn count and even earnings, compared to other states where the handloom sector has been in rapid decline between 1995 and 2010. Thus, the anomalous persistence of handloom in WB on a large scale between 1995 and 2010, against the backdrop of rapid decline at AI level, does not point to the survival of a craft under deteriorating conditions. A comparison of household perceptions of the future of handloom in selected states seems to suggest that handloom households in WB are more agnostic of their future than in other states.

In recent times, there is anecdotal evidence and journalistic accounts of sharp decline of the handloom sector in WB, as a result of penetration of powerlooms.[11] It is true that WB had historically lagged behind all the handloom-major states, barring OD, in development of powerlooms.[12] This is certainly one of the factors behind survival of handloom on a large scale in WB between 1995 and 2010 and the recent expansion of powerlooms might be responsible for the reported plight of handloom weavers in journalistic accounts. But one should not conclude from this that expansion of powerloom will necessarily lead to a terminal decline of handloom in WB, as has been observed in other states of India. The evidence presented in the preceding sections shows a picture of labour mobility across sectors in search of

[11](1) https://www.livemint.com/Leisure/aCt3UOff9OSLrvV4SOyejJ/How-the-Phulia-Tangail-went-from-boom-to-bust.html.

(2) https://thewire.in/culture/threadbare-in-santipur.

[12]https://www.pdexcil.org/files/138/NoPowerLooms.pdf.

income (such as new entrants in weaving), reallocation of household labour (greater engagement of women in handloom worker), income diversification by households (falling share of handloom income in household income), while maintaining output and quality in WB at comparable or higher levels than other states. What appears to be happening in WB is adoption of multiple strategies by households in search of maintaining, stabilizing and possibly, even increasing their incomes. In this sense, the handloom sector looks increasingly like the unorganized sector in general, where households tend to have multiple and shifting occupations across its members in order to avoid market risks that specialization brings. This is the new source of resilience of the handloom workers.

The resilience of handloom has been understood in terms of its embeddedness in a socio-technical system.

> Weavers are not individual units of production; they are linked together in a production network that forms a socio-technical system. During times of external crisis, a small core of the system remains stable, as strongly embedded weavers continue to weave, while the peripheries expand and contract with the to-and-fro movement of individual entrepreneurial weavers who jump in and out when the carrying capacity of the system increases or depletes. What then emerges is an image of a system that grows or shrinks in response to opportunity or threat in the environment, constantly striving towards its own sustainability. (Annapurna et al. 2012: 47).

At first blush, the aspects of change and mobility our analysis highlights in the preceding sections can lend themselves to an analysis in terms of socio-technology. However, we are cautious in treading that path. It has been noted that the peasant economy has been changing for several decades now—calling for a different theoretical understanding of what the peasant and the rural society means today (Chatterjee 2008). There is no reason why we should not similarly revisit our understanding of rural craftsmen. While a section of the skilled weavers and master weavers will indeed keep handloom alive and well, producing fabrics that will have steady and possibly expanding demand in domestic and international markets, for a large number of weavers, the competition from powerlooms will present a threat of loss of livelihood.

For many current handloom weavers, the skill of weaving was not handed down within the family, but was acquired outside the family—within or outside the community—as workers made occupational movements from agriculture or even some non-agricultural occupations to handloom. This seems to be happening now and this has also happened in the past (Bhattachaya and Sen 2018). Handloom was an occupational choice in the past and it continues to be so in the present. It is not simply what one was born into—as some romanticized accounts of handloom seem to suggest. The handloom sector shrunk and grew in the past and can do so in the future as a result of labour mobility and skill acquisition. Caste, religion and gender are no longer barriers to such movements, and the requisite skills are much more widely held and diffused than what traditionalist views of handloom production may make us believe. It is increasingly difficult to think of it as a 'system', let alone identify any 'core' of that system. The highly skilled handloom weavers, who face a relatively stable market demand for their products, can hardly be understood as the 'core' of a 'system', even as they impart continuity to the craft skills and act as its repository. What these skilled and master weavers anchor is a sector that expands or shrinks as markets change, essentially due to easy mobility of labour across sectors with very different skill requirements—artisanal as well as non-artisanal. In this paper, we point to the necessity of a more labour-focused approach in place of currently dominant tradition-focused understanding of the sector.

References

Annapurna, M., et al. (2012). Mobilising discourses. *Economic and Political Weekly, 47*(25), 41–51.

Azim Premji University. (2018). *State of working India 2018*. Centre for Sustainable Employment, Azim Premji University, Bengaluru. Available at https://cse.azimpremjiuniversity.edu.in/wp-content/uploads/2019/02/State_of_Working_India_2018-1.pdf. Accessed June 15, 2019.

Bhattacharya, R., & Sen, S. (2018). *Pride and prejudice: The condition of handloom weavers in West Bengal*. CSE Working Paper 2018-16, Centre for Sustainable Employment, Azim Premji University. Available at https://cse.azimpremjiuniversity.edu.in/wp-content/uploads/2018/05/Bhattacharya_Sen_Handlooms_WB.pdf. Accessed June 15, 2019.

CCI (Craft Council of India). (2015, July). *Newsletter*. Available at https://www.craftscouncilofindia.org/wp-content/uploads/2016/03/CCI-Newsletter-July-2015.pdf. Accessed December 11, 2017.

Chatterjee, P. (2008). Democracy and economic transformation in India. *Economic and Political Weekly, 43*(16), 53–62.

Chaudhury, S. (2014). *PritihibirTantghar: Banglar Bastrasilpa O Banijya 1600–1800*. Calcutta: Ananda Publishers Private Ltd.

Chowdhury, S. (2016). *Employment growth in West Bengal: An assessment*. IDSK Occasional Paper No. 52, March, 2016.

Das, S. (2001). *The warp and woof: An enquiry into the handloom industry in West Bengal*. Kolkata: K. P. Bagchi & Company.

Government of India. (2014a). *Key indicators of situation of agricultural households in India, 2013*. National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India, New Delhi. Available at http://mospi.nic.in/sites/default/files/publication_reports/KI_70_33_19dec14.pdf. Accessed June 15, 2019.

Government of India. (2014b). *Annual report, 2013–14*. Ministry of textiles. Retrieved from http://texmin.nic.in/sites/default/files/ar_13_14_english.pdf. Last accessed May 1, 2018.

Government of India. (2015). *All India report of sixth economic census 2013–14*. Ministry of Statistics and Programme Implementation, Government of India. Available at www.mospi.gov.in/all-india-report-sixth-economic-census. Accessed June 15, 2019.

Government of India. (2017a). *Key indicators of unincorporated non-agricultural enterprises (Excluding construction) in India, 2015–16*. National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India, New Delhi. Available at http://www.mospi.gov.in/sites/default/files/publication_reports/NSS_KI_73_2.34.pdf. Accessed June 15, 2019.

Government of India. (2017b). *Annual report, 2016–17*. Ministry of Textiles. Available at http://texmin.nic.in/sites/default/files/AnnualReport2017-18%28English%29.pdf. Last accessed May 1, 2018.

Khasnabis, R., & Nag, P. (2001). Labour process in the informal sector: A handloom industry in Nadia District. *Economic and Political Weekly, 36*(52), 4836–4845.

Liebl, M., & Roy, T. (2003). Handmade in India: Preliminary analysis of crafts producers and crafts production. *Economic and Political Weekly, 38*(51/52), 5366–5376.

Mamidipudi, A. (2016). *Towards a theory of innovation in handloom weaving in India* (Doctoral dissertation, MUSTS, Maastricht University). Available at https://cris.maastrichtuniversity.nl/portal/files/4083287/c5295.pdf. Last accessed May 1, 2018.

Mukund, K. (1992). Indian textile industry in 17th and 18th Centuries: Structure, organisation and responses. *Economic and Political Weekly, 27*(38), 2057–2065.

Mukund, K., & Syamasundari, B. (2001). *Traditional industry in the new market economy: The cotton handlooms of Andhra Pradesh*. New Delhi: Sage Publications.

NCAER. (2004). *Joint census of handlooms & powerlooms 1995–96: Handloom Sector*. New Delhi: National Council of Applied Economic Research.

NCAER. (2010). *Handloom census of India 2009–10*. Ministry of Textiles, Government of India, available at http://handlooms.nic.in/Writereaddata/Handloom%20report.pdf. Accessed June 15, 2019.

Niranjana, S., & Vinayan, S. (2001). *Report on growth and prospects of the handloom industry*. Planning Commission, Government of India, Available at http://planningcommission.gov.in/reports/sereport/ser/stdy_hndloom.pdf. Accessed June 15, 2019.

Roy, T. (1998a). Development or distortion? 'Powerlooms' in India, 1950–1997. *Economic and Political Weekly, 33*(16), 897–911.

Roy, T. (1998b). Economic reforms and textile industry in India. *Economic and Political Weekly, 33*(32), 2173–2182.

Roy, T. (2009). The long globalization and textile producers in India. In L. H. van Voss, E. Hiemstra-Kuperus, & E. van Nederveen Meerkerk (Eds.), *The Ashgate companion to the history of textile workers, 1650–2000* (253–274). Farnham: Ashgate.

Srinivasulu, K. (1996). 1985 textile policy and handloom industry, policy, promises and performance. *Economic and Political Weekly, 31*(49), 3198–3206.

Subramanya, M. H. B. (2011). Small scale industry development for export promotion: India's experience under liberalisation. In K. Das (Ed.), *Micro and small enterprises in India: The era of reforms* (pp. 46–66), New Delhi: Routledge.

Total Factor Productivity Growth of Elementary Education in India and Its Determinants: Evidence from a Non-parametric Data Envelopment Approach

Arpita Ghose

Abstract The major departure of this paper from the available literature is the approach to estimating total factor productivity growth (TFPG) for elementary education in India, by constructing two frontiers (i) for general category states (GCS) and (ii) for special category states and union territories (SCS&UT), over the period 2005–06 to 2014–15; as these two groups are not homogeneous and operate under different fiscal and economic conditions. Hence, maximum educational output producible from an input bundle by a school within a particular group may not be as high as what could be produced if the school could choose to locate in the other groups. TFPG is measured by Malmquist Productivity Index (MPI) using non-parametric data envelopment analysis for primary and upper primary levels of education and for GCS and SCS&UT in a two-output, four-input framework under variable returns to scale, considering both quantities and qualities of outputs and inputs. The outputs are net enrolment ratio and percentage of students passed with 60% and above in the examination, representing output quality. The inputs used are: (i) number of schools per lakh population, (ii) teacher–pupil ratio in the school, (iii) classroom–student ratio in the school, (iv) percentage of teachers with qualification graduate and above in the schools, indicating quality of the teacher input. The generated value of MPI is decomposed into technical change, efficiency change and scale efficiency change. The decomposition results suggest that on average productivity change is mainly facilitated by technical change and efficiency change. After obtaining MPI, a second-stage panel regression is resorted to find out its determinants, considering the effect of favourable and poor infrastructure, social and policy indicators and also the macro-indicators to see whether TFPG has been facilitated by existence of favourable infrastructure, or, existence of poor infrastructure inhibits TFPG, whether inclusion of the backward classes into the system, the provision of more public facilities can increase TFPG and whether favourable macro-indicators, i.e. favourable general economic environment of the state matters in explaining TFPG. The factors influencing the MPI are explained separately for four groups GCS-primary, GCS–upper primary; SCS&UT–primary, SCS&UT–upper primary. Results of panel regression suggest that infrastructural variables, policy variables, school-specific variables and

A. Ghose (✉)
Department of Economics, Jadavpur University, Kolkata, India
e-mail: arpitaju@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_18

also the state-level macro-aggregates are important in explaining MPI, and the interaction effect between different explanatory variables is also evident. Some policy suggestions for improving TFPG are highlighted.

1 Introduction

The most distinctive feature of our economic system is the growth of human capital. Without it there would be only hard manual work and poverty (Schultz 1961). The importance of education and human capital has also been brought out in many studies of endogenous growth literature (Lucas 1988; Romer 1990 among others). Countries with a high amount of human capital would be expected to have a higher level of steady-state income. Mankiw-Romer-Weil (1992) presented a simple extension to the Solow model by incorporating human capital as a separate input and the production technology in this model which has come to be known as the human capital augmented Solow model. The human capital augmented Solow model predicts that, other things being equal, a country should have a higher level of per capita income if it has a high amount of human capital. It is the education and the creation of human capital was responsible for both the differences in labour productivity and the differences in overall levels of technology (Barro 1991).

The development of human capital in an economy is basically dependent on development of education sector. In India's National Development Programme, a substantial emphasis was given to the education sector. This emphasis can be seen from the amount of government fund allotted to education sector in the budget of the India government and also from the national education policy (NEP) undertaken by the government and revised it from time to time. During the period 2011–12 to 2014–15, the expenditure on education as a percentage of total expenditure by general government (Centre & States) is around 11%; the percentage of total expenditure devoted to total services is highest for education; around 46–47% among all the other services and the expenditure on education as percentage of GDP is more or less 3% of GDP. (Budget Documents of Union and State Government, RBI). Regarding NEP, it can be said that the first NPE was promulgated in 1968 by the government of Prime Minister Indira Gandhi, and the second by Prime Minister Rajiv Gandhi in 1986. In 1986, NPE emerged as a result of a participatory process and in-depth review of the Indian educational system. This policy was aimed at providing comprehensive framework to guide the development of education and impart a sense of urgency to the goal of universalizing elementary education. The NPE accorded priority to universal access to elementary education, universal participation and universal achievement of at least minimum level of learning prescribed for the elementary stage of education. Since then, state policies have closely followed national policies and the states have undertaken their own action. Universal access to elementary education is thus viewed as a composite programme of (i) universal access to elementary education for children up to the age of 14; (ii) universal participation until they complete the elementary stage through formal or informal programme and (iii) universal achievement of at least the minimum level learning prescribed for the elementary stage of education. The NEP (2016) focused on addressing the issue of gender discrimination, the creation of educational tribunals and common curriculum for science, mathematics and English and has mainly given emphasis on: (i) preschool education, (ii) curriculum renewal and examination reforms, (iii) setting a standard norm for learning outcome in school education and to be applied for government and private school, (iv) protection of rights of the child and adolescent education and employability, (v) development of skill in education and employability, (vi) use of ICT in education, teacher development and management, teacher development and management, (vii) self-development through comprehensive education, i.e. physical education, yoga, games and sports, NCC, NSS, art education and making Balsansad covering local art, craft, literature and skill and other co-scholastic activities an integral part of the curriculum and daily routine in the school, (ix) open and distance learning programme among others. Recently in 2019, the Committee for Draft National Education Policy Chaired by Dr. K. Kasturirangan submitted the report. The report proposes an education policy, which seeks to address the challenges of: (i) access, (ii) equity, (iii) quality, (iv) affordability and (v) accountability faced by the current education system. It provides for reforms at all levels of education from school to higher education. It seeks to increase the focus on early childhood care, reform the current examination system, strengthen teacher training and restructure the education regulatory framework. It also seeks to set up a National Education Commission, increase public investment in education and strengthen the use of technology and increased focus on vocational and adult education, among others. The draft policy of Kasturirangan committee recommends extending the ambit of the RTE Act to include early childhood education and secondary school education; this would extend the coverage of the Act to all children between the ages of three and 18 years. Further, in the national development programme, emphasis was given not only to the eradication of illiteracy and upgradation of literacy skills but also the upgradation of the quality of life through the improvement in occupational skills.

For the development of different levels of education crucial is the development of elementary level of education. Different programmes were launched by the Government of India to improve the overall performance of elementary education system in India. Elementary education has been of priority in terms of sub-sector allocations. A number of schemes like (i) District Primary Education Programme have been launched by the Central Government to meet the needs of the educationally disadvantaged groups. A comprehensive programme called the Sarva Shiksha Abhijaan (SSA) was launched in November 2000 in partnership with the States. The programme aims to improve the performance of the school system through a community-based approach and to impart quality elementary education to all children in the age group of 6–14 years. It seeks to bridge gender and social disparities at the elementary level. It subsumes all existing programmes, within its overall framework, with the district as the unit level of implementation. National Programme for Education of Girls at the Elementary Level (NPEGEL) has been launched in 2003–04 for providing additional components for education of girls at the Elementary Level under SSA.

Performance of the Education Sector as represented by literacy rate during the Five Year Plan is quite satisfactory. Literacy rates in India have risen dramatically from 18.3% in 1951 to 74% according to provisional census data of 2011. However, vast disparities are present among the States in Gross Enrolment Ratio at primary/upper primary Level, sates in the North-East performing better than the other States. Ten states and union territories, including Kerala, Lakshadweep, Mizoram, Tripura, Goa, Daman and Diu, Puducherry, Chandigarh, National Capital Territory of Delhi and the Andaman and Nicobar Islands, have attained literacy rate of above 85%. Kerala has the highest literacy rate at 93.91% followed by Lakshadweep at 92.28%. Bihar is at the bottom of the ladder with literacy rate of 63.82 followed by Arunachal Pradesh at 66.95. Literacy rate of Rajasthan rise to 67% from 60% (Census Data 2011).

Given this vast development of the education sector, naturally, a related question arises what is the extent of Total Factor Productivity Growth (TFPG) in Education Sector? For any country, which wants to perform in their education sector, needs to enhance its cost-competitiveness by fostering TFPG. It measures the amount of increase in total output, which is not accounted for the increase in total inputs and thus measures shift in output due to the shift in the production over time, holding all inputs constant.

For attainment of equalization of education facilities across different States of India, an analysis of TFPG at the State level is needed. A related question is what are the factors responsible for the variation of TFPG across different states of India in the elementary education level? Such type of analysis will help us to identify the states that are below the average productivity score and to formulate policies for enhancing productivity of less productive states, so that equalization of educational efficiency across different States of India can be ensured.

The perusal of the literature on TFPG of the education sector reveals that there is a dearth in the literature dealing with the analysis of TFPG of elementary education sector around the globe. The present paper contributes to the literature in this direction taking data from Indian states for the period 2005–06 to 2014–15.

The rest of the paper unfolds as follows. Section 2 presents the literature review and contribution of the paper to the literature. Section 3 gives a short description of TFPG and of the factors used to explain the variation of TFPG of elementary education in India and also describes the data sources. Section 4 presents results of estimation. Summary of the results and some policy suggestions to improve the productivity score of India's elementary education are presented in the concluding Sect. 5.

2 The Literature Review and the Contribution of the Paper

In the literature, total factor productivity growth is estimated by assuming the existence of some kind of production relationship between input and output. There are three major approaches of estimating TFPG: (i) growth accounting approach, (ii) parametric approach, i.e. by estimating parametric production functions or cost function and (iii) non-parametric data envelopment approaches (DEA). The first approach, the growth accounting, is a procedure to measure the contribution of different factors to economic growth and to indirectly compute the rate of technological progress, measured as the residual in an economy. Growth accounting decomposes the growth rate of an economy's total output into which is due to an increase in contributing amount of factor used usually the increase in the capital and labour and that which cannot be accounted for by the observable changes in factor utilization. The unexplained part of growth in GDP is taken to represent increase in production, getting more output with the same amount of inputs or a measure of broadly defined technical progress. In the second approach, i.e. the parametric approach, one possibility is to explicitly specify a production function. The value of this function at the input level under consideration denotes the maximum producible quantity. The more common practice is to estimate the parameters of the specified function empirically from a sample of input–output data. The least square procedure permits observed points to lie above the fitted line and fails to construct a frontier. At the same time, specifying a one-sided distribution of the disturbance term leads to a deterministic frontier and any deviation from this frontier is interpreted as inefficiency. The other alternative is to fit a stochastic frontier production function. In stochastic frontier model, one includes a composite error, which is a sum of one-sided disturbance term representing the shortfalls of the actually produced output from the frontier due to random factors. Both the deterministic as well as stochastic frontier model can be used to estimate total factor productivity growth using the econometrically estimated values parameters of the model. For the econometric procedure, one must select a particular functional form, (e.g. Cobb Douglas), out of a number of alternatives. At any input bundle x_0, the value obtained by $f(x_0)$ will depend on the functional form chosen. Further, the parameter estimates are also sensitive to the choice of the probability distribution specified for the disturbance term. The misspecification in either the functional form or in the disturbance term will affect the estimation process and may lead to wrong conclusions. In the third approach, i.e. non-parametric DEA, the major advantage is that one need not assume an explicit form of the production and construct a benchmark technology from the observed input bundle of the decision-making unit (DMU) by making only a minimum number of fairly general assumption.

At *the international level* measurement of TFPG was done by Johnes (2008) and Bradley et al. (2010). Johnes (2008) has used the distance function to compute Malmquist Productivity Indexes for 112 English Higher Education Institutions (HEIs) over the period 1996–97 to 2004–05. His study showed that the 1% annual average increase in productivity of the HEIs. Two measures of inputs and five measures of outputs are considered in this study. The inputs are staff (combination of full-time and part-time academic staff), admin (expenditure on various factors including staff, student facilities and grant), expenditure on library and learning resource centres, etc., UG students and PG students. The outputs are: UG output, PG output and Research (income received in various research grants). Using DEA, Bradley et al. (2010) revealed that the productivity change was 12% over the period 1999–2003 for the data of 200 further education providers. Parteka and Wolszczak-Derlacz (2013) examined the patterns of productivity change in a large set of 266 public higher education institutions (HEIs) in 7 European countries across the time period 2001–2005. They adopt consistent bootstrap estimation procedures to obtain confidence intervals for Malmquist indices of HEI productivity and their components and are able to assess the statistical significance of changes in HEI productivity, efficiency and technology. Estimated results suggest that, assessed vise-a-visa a common 'European' frontier, HEI productivity rose on average by 4% annually. Statistically significant changes in productivity were registered in 90% of observations on the institutions in the sample, but statistically significant annual improvements in overall productivity took place in only 56% of cases. There are considerable national differences, with German, Italian and Swiss HEIs performing better in terms of productivity change than HEIs from the other countries examined. Arjomandi et al. (2015) analysed efficiency and productivity changes using the Hicks–Moorsteen total factor productivity index, in the context of higher education institutions in Malaysia. It is assumed that the production technology exhibits variable returns to scale, which is more plausible than the constant returns to scale assumption because universities usually operate at suboptimal scales. Three major groupings of Malaysian public universities are used in the case study: research, comprehensive and focused universities. The results show that technical efficiency has improved after the 2007 National Higher Education Strategic Plan within all the three university groupings. The other studies that applied DEA in measuring the productivity in education sector around the globe are Bessent et al. (1982, 1984), among others.

But *in the Indian context,* research work on measuring productivity in education sector is not quite sufficient. There are some studies relating to efficiency of elementary education in the Indian context. Tyagi (2009) assessed the technical efficiency and efficiency differences among 348 elementary schools of Uttar Pradesh state in India by using DEA. Sengupta and Pal (2010) explained the efficiency primary education sector in India using DISE statistics. They identified five basic aspects of education: deprivation aspects, social aspects, policy aspects, traditional performance indicators and Farrell's non-parametric efficiency scores. The country has been divided into five zones: Northern, North-eastern, Eastern, Western and Southern districts, respectively. Using the formula of Human Poverty Index (HPI) suggested by Anand and Sen (1996), they derived Grand Poverty Index comprising of various poverty indicators of education system in India and tried to relate these with the efficiency score in DEA. Dutta (2012) assesses the technical efficiency and efficiency differences in elementary education system across Indian states by using DEA and regression model. It considers different educational outcome variables such as enrolment rate in primary schools and upper primary schools, and completion rate in primary schools and upper primary schools only for the year 2007–2008. Results show that Delhi, Kerala, Tamil Nadu and Nagaland with relatively high achievement values in education sector are also operating at the efficient frontier. On the other hand, Bihar, Madhya Pradesh, Odisha and Uttar Pradesh have low outcomes, but the problem with these states is that given their current input levels, there is limited scope for improving their outcomes and catching up with the better-off states. The study also makes an attempt to identify the determinants of efficiency by using simple regression model. Ghose (2017) estimated technical-efficiency scores (TE) for primary and upper-primary level of education in India separately by constructing *two separate frontiers* for *two groups of states* (i) General Category States(GCS) and (ii) Special Category States (SCS) and the union territories (UT) using non-parametric Data-Envelopment-Analysis (assuming variable-returns-to-scale), obtains technology closeness ratio (TCR) measuring the proximity of the group frontiers to the Meta-frontier (representing the frontiers comprising of all states). Estimation is carried out for 2005–06 to 2008–09. A second-stage panel regression explains the determinants of TE of both primary and upper primary level separately for GCS and SCS.

The perusal of the literature suggests that there is a dearth of studies dealing with TFPG of the education sector in general in the case of India. The present study addresses this issue and measures TFPG of elementary education in India using non-parametric DEA. The major advantage of non-parametric DEA is that one need not assume an explicit form of the production. The problem with the parametric approach is that misspecification in either the functional form or in the disturbance term will affect may lead to biased estimator. The other reason for choosing DEA is that Charnes et al. (CCR) (1978, 1981) introduced the method of DEA to address the problem of efficiency measurement for decision-making units (DMUs) with multiple inputs and multiple outputs in the absence of market prices. They coined the phrase decision-making units to include nonmarket agencies like schools, hospitals and courts, which produce identifiable and measurable outputs from measurable inputs but generally lack market prices of outputs (and often of some inputs as well). Since no prices are available, what we would need in this situation is to use vectors of ‘shadow’ prices of inputs and outputs, which are calculated using some programming method. As the paper attempts to estimate TFPG of elementary education for which no price data are available for some of the inputs and outputs, DEA method is adopted to construct the frontier. ***The contributions of the paper to the literature are the following***.

First of all, the TFPG in different states of India is estimated by using Malmquist Productivity Index (MPI) introduced by Caves et al. (CCD) (1982), applying non-parametric DEA to measure MPI. While estimating TFPG, the present paper departs from the assumption that not all the states of India operate under same fiscal environment. Some of the states, namely Special Category States (SCS) like Arunachal Pradesh, Assam, Himachal Pradesh, Jammu and Kashmir, Manipur, Meghalaya, Mizoram and Nagaland, are given more benefits by the central government. The main reason behind this categorization is the development of that particular state where there are many problems due to hilly terrains, international borders, etc., as there cannot be good industrial development. These states have also geographical disadvantages in their effort for infrastructural development. The states in the North–East are also late starters in development. The finances of the state are also less. Thus, the central government comes into picture. Public expenditure plays a significant role in the Gross State Domestic Product of the states. 90% of the central assistance is treated as grant and remaining 10% is considered as loan unlike other states which get 30% grant and 70% loan. The most important prescription for SCS is interest-free loan with rationalization of public expenditure based on growth-enhancing sectoral allocation of resources. For SCS unlike other states, there is no hard budget constraint as the central transfer is high. Through the enactment of Fiscal Responsibility and Budget Management, FRBM, these states are also availing themselves of the benefit of debt swapping and debt relief schemes which facilitate reduction of the average annual rate of interest. The union territories (UT) also get more grants as compared to General Category States (GCS). Thus, this paper does not assume a common frontier for all the states and the union territories. Rather, this paper constructs two group frontiers: the first one comprising of the states only with GCS and the second one comprising of SCS and UT. This is because maximum educational output producible from an input bundle by a school within a particular group of state may not be as high as what could be produced if the school could choose to locate in other groups. **Secondly**, the present paper intends to compare the performance of TFPG of primary and secondary level of education. As estimation of productivity score for more than one year always gives more information, it measures TFPG of both primary as well as upper primary level of education for ten successive years from 2005–06 to 2014–15, under the assumption of variable returns to scale. **Thirdly**, the present paper decomposes the productivity change into technical change, efficiency change and scale efficiency change following the method as suggested by Ray and Desli (1997). **Finally**, while finding out the determinants of total factor productivity index, it takes into account (i) some factors from poor infrastructure to see whether poor infrastructure inhibit the achievement of productivity, (ii) some social indicator to investigate whether inclusion of the backward classes into the education system increases productivity, (iii) some policy variable to test whether provision of more public facilities increases productivity and also (iv) some state-level macro-aggregates like per capita net state domestic product, income inequality, number of persons below the poverty line and population density of the state to see to what extent the general environment of the state matters in explaining productivity. Also, rather than using a composite index, the present paper attempts to find out the individual effect of the different explanatory variables. Because for policy purposes, it is necessary to know the individual effect of different explanatory variables in order to boost up TFPG. The major disadvantage of using composite indicator as done by some of the earlier studies is that some of the individual variables may significantly affect TFPG, while the others are not and the use of the composite indicator cannot differentiate between these two. Relevantly, it can be mentioned here that the extensive Monte Carlo Simulation reveals that a DEA-based procedure in the first stage followed by Ordinary Least Squares (OLS) analysis in the second stage yields consistent estimators of the impact of contextual variables if the contextual variables, affecting productivity, are independent of the input variables (although they may be correlated with each other). Additionally, two-stage DEA-based methods with OLS in the second stage significantly outperform the parametric methods (Banker and Natarajan 2008). The present paper also adopts the two-stage process. After estimating the TFPG in the first stage, it uses panel regression framework to determine the factors determining TFPG.

3 Methodology and Data Sources

3.1 Estimation of Total Factor Productivity Growth (TFPG)

TFPG is estimated by applying non-parametric Data Envelopment Analysis (DEA) which basically rests on assumed production relationship between input and outputs. Like other production systems, education also has the production function showing the relationship between school or student inputs and a measure of school output.

Educational achievement can be viewed as a production process, where inputs of resources are applied to the relatively 'unfinished' child, and an output-pupil achievement results. The objective is to generate the maximum achievement using a given amount of school resources.

Following Ray (2004), educational production function can be defined as

$$A = H.f(X_a, \ldots, X_m),$$

where A = some measure of school output—for example, enrolment ratio and or marks obtained in standardized examination system. The second output represents the quality of output.

X_a, …, X_m = variables measuring the school resources. These variables can be treated as inputs to the production process. The variables here would typically include the amount and quality of teaching services, the physical infrastructure or facilities of the school related to teaching and learning, (i) like teacher–pupil ratio in the class, (ii) classroom–student ratio in the school, (iii) qualification of the teacher and the like. *Note here that ordering of the variables is not important.* The technical relationship between input and output is by 'f' function.

In this context, let us note that the extent to which these inputs can produce output will depend on the availability of the other accompanying variables like (i) I_1, …, I_n, the variables representing favourable infrastructure that can promote productivity, (ii) $\mathrm{PI}_1, \ldots, \mathrm{PI}_p$, the variable representing poor infrastructure that may reduce productivity level, (iii) PO_1, …, PO_q, the variable representing policy variables that can promote productivity level, (iv) S_1, …, S_r, the social indicator variable. These variables are school characteristics and represent the composition of the school body. The effect of these variables on productivity is not expected prior and in fact is a testable proposition, (v) E_1, …, E_q, the variables representing environmental influences on learning outside the school. These variables take care of the general environment of learning that the student faces outside the school. The combined influence of these variables (i)–(v) will determine the position of the 'f' function and can be treated as shift factor. This shift factor can be represented by the term H. Here also note that the ordering of the determinant variables does not matter.

The point to be noted is the variables under (i)–(v) are not the inputs in the production process. Although they are different variables from the inputs, they determine the outcome value of the educational output, given the inputs. Hence, these variables may be treated as the determinants of TFPG.

In 1-output-1 input case, the rate of productivity growth is measured by the difference in growth rates of output and input quantities, respectively. When multiple inputs are involved, the rate of multi-factor productivity growth (MFPG) can be measured by the difference in growth rate of output and that of total input where growth rate of total input can be computed by the growth rates of individual inputs weighted by the partial output elasticity. In parametric analysis, the specification of some explicit functional form of a production, cost or profit function is needed. In non-parametric analysis, the exact technological relationship is unspecified. The relevant assumptions made in the paper are: (i) both inputs and output are freely disposable and the production possibility set is convex. (ii) All input–output combinations, actually observed, are by definition feasible. (iii) Variable returns to scale prevails.

Consider, for simplicity, a single-input–single-output sector consists of n decision-making units (DMU). x_k^t and y_k^t represent the input and output quantities of DMU k at time t. The average productivity of this firm at time t is

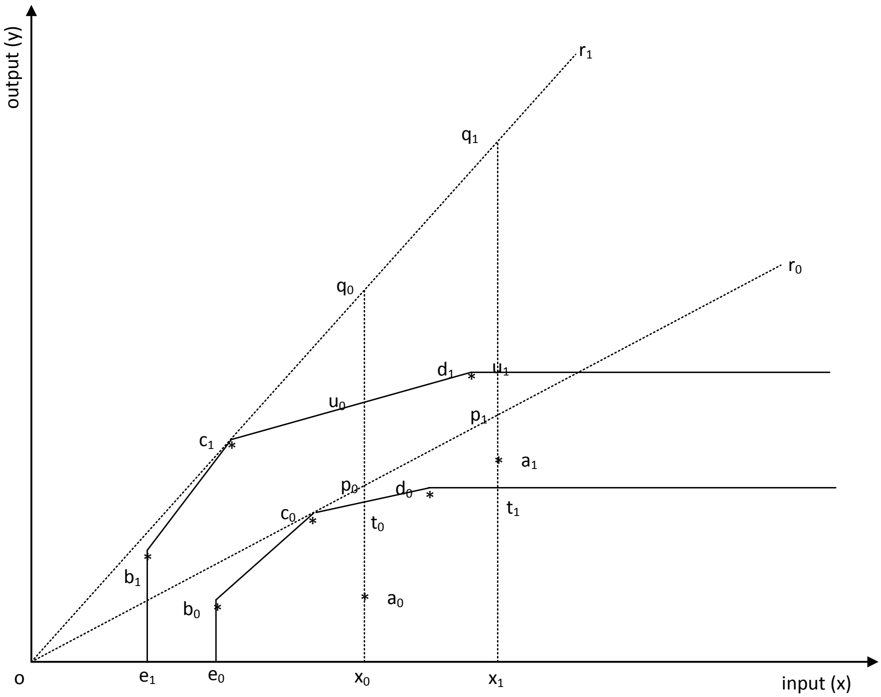
$$\mathrm{AP}_k^t = \frac{y_k^t}{x_k^t} \tag{1}$$

Thus, a productivity index for this DMU at time $t + 1$, with period t treated as the base, will be

$$\prod_{k(t+1)} = \frac{\mathrm{AP}_k^{t+1}}{\mathrm{AP}_k^t} = \frac{\frac{y_k^{t+1}}{x_k^{t+1}}}{\frac{y_k^t}{x_k^t}} \tag{2}$$

which does not in any way depend on the assumptions about returns to scale. In order to identify the sources of productivity change, however, a benchmark technology is needed, where returns to scale assumption becomes important.

According to Varian (1984), the free disposal convex hull of observed input–output vectors provides an inner approximation to the true underlying production possibility set, if the above-mentioned assumptions (i) and (ii) hold good.



Consider a sector consisting of four DMU: a, b, c, and d. Following the figure, a_0, b_0, c_0 and d_0 show the observed input–output levels of the respective DMU in period 0. Similarly, points a_1 through d_1 show their input–output levels in period 1. DMU 'a' uses input ox_0 to produce output a_0x_0 in period 0 and input ox_1 to produce output a_1x_1 in period 1. Thus, the productivity index for DMU 'a' in period 1 is

$$\prod_{a_1} = \frac{\frac{a_1x_1}{ox_1}}{\frac{a_0x_0}{ox_0}} \tag{3}$$

By convexity, all the points in the convex hull of the points a_0, b_0, c_0 and d_0 (i.e. the convex combinations of these points) represent feasible input–output combinations in period 0.

The free disposal convex hull is the set of points bounded by the horizontal axis and the broken line $e_0b_0c_0d_0$—extension. Under VRS, all points in this region represent feasible input–output combinations in period 0, although under constant returns to scale (CRS) all radial expansion and (non-negative) contraction of feasible input–output bundles are also feasible, thus the CRS production possibility set in period 0 is the cone formed by the horizontal axis and the ray or_0 through the point c_0.

The VRS frontier in period 1 is the broken line $e_1b_1c_1d_1$—extension and the CRS frontier is the ray or_1 through the point c_1. Define the production possibility set as

$$S^t = \{(x, y): y \text{ can be produced from } x \in \text{period } t\} \tag{4}$$

The output distance function[1] is

$$D^t(x, y) = \min \theta : \left(x, \frac{1}{\theta} y\right) \in S^t \tag{5}$$

In period 0, the maximum producible output from input ox_0 is t_0x_0 under the VRS assumption. Thus, the distance functions are
$D_V^0(x_0, y_0) = \frac{a_0x_0}{t_0x_0}$ and $D_V^o(x_1, y_1) = \frac{a_1x_1}{t_1x_1}$, in period 0
The productivity index for DMU 'a' is

$$\prod_{a_0} = \frac{\frac{a_1x_1}{ox_1}}{\frac{a_0x_0}{ox_0}} = \frac{\frac{a_1x_1}{p_1x_1}\cdot\frac{p_1x_1}{ox_1}}{\frac{a_0x_0}{p_0x_0}\cdot\frac{p_0x_0}{ox_0}} = \frac{D_c^0(x_1, y_1)}{D_c^0(x_0, y_0)} \tag{6}$$

Analogously,

$$\prod_{a_1} = \frac{\frac{a_1x_1}{ox_1}}{\frac{a_0x_0}{ox_0}} = \frac{D_c^1(x_1, y_1)}{D_c^1(x_0, y_0)} \tag{7}$$

According to Färe et al. (1994) for any reference technology, the distance functions can be calculated. But the problem is that the productivity index is given by the ratio of the CRS distance functions even if the technology was not characterized by CRS. With explicit assumption of VRS, comparing CRS and VRS frontiers in period 0, we get both t_0 and t_1 are points on the production frontier, (both are technically efficient), and the average productivity at t_0 is higher than that of t_1. The point of the highest average productivity along the VRS frontier in period 0 is c_0, whereas along the CRS frontier, that remains constant. The point of the highest average productivity along the VRS frontier is called the most productive scale size (MPSS). At the MPSS, CRS and VRS frontiers coincide (Banker et al. 1984). Notably, the average productivity at the MPSS of the VRS frontier (point c_0) is equal to the constant average productivity at any point on the CRS frontier (say, p_0 or p_1). The scale efficiency at any point on the frontier is measured by the ratio of the average productivity at that point to the average productivity at the MPSS.

Thus,

$$\mathrm{SE}^0(x_0, y_0) = \frac{\mathrm{AP}(t_0)}{\mathrm{AP}(c_0)} = \frac{t_0x_0}{p_0x_0} = \frac{D_c^0(x_0, y_0)}{D_v^0(x_0, y_0)} \tag{8}$$

Also,

[1]Let the production possibility set: $T = \{(x, y): x \text{ can produce } y\}$. Let (x, y) be any input–output bundle, (not necessarily feasible), then the output-oriented distance function is. $D(x, y) = \min \theta : \left(x, \frac{1}{\theta} y\right) \in T$. Thus, $(x, y) \in T$ implies $D(x, y) \leq 1$.

$$\mathrm{SE}^0(x_1, y_1) = \frac{\mathrm{AP}(t_1)}{\mathrm{AP}(c_0)} = \frac{D_c^0(x_1, y_1)}{D_v^0(x_1, y_1)} \tag{9}$$

Now Eq. (6) can be written as

$$\prod_{a_0} = \frac{D_v^0(x_1, y_1) \cdot \frac{D_c^0(x_1, y_1)}{D_v^0(x_1, y_1)}}{D_v^0(x_0, y_0) \cdot \frac{D_c^0(x_0, y_0)}{D_v^0(x_0, y_0)}} = \frac{D_v^0(x_1, y_1)}{D_v^0(x_0, y_0)} \cdot \frac{\mathrm{SE}^0(x_1, y_1)}{\mathrm{SE}^0(x_0, y_0)} \tag{10}$$

In a perfectly analogous manner,

$$\prod_{a_1} = \frac{D_v^1(x_1, y_1)}{D_v^1(x_0, y_0)} \cdot \frac{\mathrm{SE}^1(x_1, y_1)}{\mathrm{SE}^1(x_0, y_0)} \tag{11}$$

Now, the MPI can be decomposed, as done by Ray and Desli (1997), in the following manner. The expression is,

$$\begin{aligned}
\prod_{a} &= \left(\prod_{a_0} \cdot \prod_{a_1}\right)^{\frac{1}{2}} \\
&= \left[\frac{D_v^0(x_1, y_1)}{D_v^0(x_0, y_0)} \cdot \frac{\mathrm{SE}^0(x_1, y_1)}{\mathrm{SE}^0(x_0, y_0)} \times \frac{D_v^1(x_1, y_1)}{D_v^1(x_0, y_0)} \cdot \frac{\mathrm{SE}^1(x_1, y_1)}{\mathrm{SE}^1(x_0, y_0)}\right]^{\frac{1}{2}} \\
&= \left[\frac{D_v^0(x_1, y_1)}{D_v^0(x_0, y_0)} \cdot \frac{D_v^1(x_1, y_1)}{D_v^1(x_0, y_0)}\right]^{\frac{1}{2}} \times \left[\frac{\mathrm{SE}^0(x_1, y_1)}{\mathrm{SE}^0(x_0, y_0)} \cdot \frac{\mathrm{SE}^1(x_1, y_1)}{\mathrm{SE}^1(x_0, y_0)}\right]^{\frac{1}{2}} \\
&= \frac{D_v^1(x_1, y_1)}{D_v^0(x_0, y_0)} \times \left[\frac{D_v^0(x_0, y_0)}{D_v^1(x_0, y_0)} \cdot \frac{D_v^0(x_1, y_1)}{D_v^1(x_1, y_1)}\right]^{\frac{1}{2}} \times \left[\frac{\mathrm{SE}^0(x_1, y_1)}{\mathrm{SE}^0(x_0, y_0)} \cdot \frac{\mathrm{SE}^1(x_1, y_1)}{\mathrm{SE}^1(x_0, y_0)}\right]^{\frac{1}{2}} \\
&= \text{peffch. techch.sch}
\end{aligned} \tag{12}$$

where,
peffch $= \frac{D_v^1(x_1, y_1)}{D_v^0(x_0, y_0)}$ measures pure technical efficiency change,
sch $= \left[\frac{\mathrm{SE}^0(x_1, y_1)}{\mathrm{SE}^0(x_0, y_0)} \cdot \frac{\mathrm{SE}^1(x_1, y_1)}{\mathrm{SE}^1(x_0, y_0)}\right]^{\frac{1}{2}}$ measures change in scale efficiency, and
techch $= \left[\frac{D_v^0(x_0, y_0)}{D_v^1(x_0, y_0)} \cdot \frac{D_v^0(x_1, y_1)}{D_v^1(x_1, y_1)}\right]^{\frac{1}{2}}$ measures technical change,[2] which is the geometric mean of the shift in the production function at x_0 and x_1.

FGNZ 1994 showed a similar decomposition. But there exists some inconsistency in their method of analysis. The technical change factor, according to FGNZ 1994, is the geometric mean of the shift in the pseudo production function[3] and not of actual production function (Ray and Desli 1997).

[2]The terminologies peffch, techchandsch are borrowed from FGNZ 1994.

[3]Let 1-input 1-output technology be represented by the production function $y = f(x)$.

3.2 Non-parametric Methodology

The decomposition of MPI into technical change, technical efficiency change and scale efficiency change can be applied in practical sense if the reference technology set is constructed from sample data in the following way—let, y_j^t and x_j^t represent the output and input vectors, respectively, of firm j (j = 1, 2, 3, …, N) in period t. Following Varian (1984), an inner approximation to the underlying production possibility set in period t will be $S^t = [(x, y) : \sum_{j=1}^{N} \lambda_j x_j^t \leq x; \sum_{j=1}^{N} \lambda_j y_j^t \geq y; \sum \lambda_j = 1; \quad \lambda_j \geq 0 \quad (j = 1, 2, 3 \ldots, N)]$.

To be noted here that, by assumption, any observed input–output bundle $\left(x_t^j, y_t^j\right)$ is feasible in period t. By the convexity assumption, any input–output pair $(\bar{x}, \bar{y})$ satisfying $\bar{x} = \sum_{j=1}^{N} \lambda_j x_j^t, \bar{y} = \sum_{j=1}^{N} \lambda_j y_j^t, \sum_{j=1}^{N} \lambda_j = 1, \quad \lambda_j \geq 0, \quad (j = 1, 2, 3 \ldots, N)$ is also feasible, and by the free disposability assumption, any $x \geq \bar{x}$ corresponds $\bar{y}$.

Hence, x can also produce y if $y \leq \bar{y}$.

The output-oriented distance function under VRS is obtained as $D_v^t\left(x_k^i, y_k^i\right) = \frac{1}{\Phi^*}$, where $\Phi^* = \max \Phi$ subject to $\sum_{j=1}^{N} \lambda_j y_t^j \geq \Phi y_k^t; \sum_{j=1}^{N} \lambda_j x_t^j \leq x_k^t; \sum_{j=1}^{N} \lambda_j = 1; \quad \lambda_j \geq 0, \quad (j = 1, 2, 3 \ldots N)$ The own-period distance functions can be found for $t = k$, while $t \neq k$ will define the cross-period distance functions.

The present paper considers **two outputs,** viz. (i) net enrolment ratio and (ii) percentage of students passed with 60% in the examination. This variable measures achievement of quality output. The **inputs** used are: (i) number of schools per lakh population, (ii) teacher–pupil ratio in the school, (iii) classroom–student ratio in the school, (iv) percentage of teachers with qualification graduate and above in the schools. This variable measures quality of the teacher input.

The included General Category States are Andhra Pradesh (AP), Bihar (BI), Chhattisgarh, Goa, Gujarat (GU), Haryana (HA), Jharkhand, Karnataka (KA), Kerala (KE), Madhya Pradesh, Maharashtra (MH), Orissa (OR), Punjab (PU), Rajasthan (RA), Tamil Nadu (TN), Uttar Pradesh (UP) and West Bengal (WB). The included Special Category and Union Territories are Andaman and Nicobar Islands, Arunachal Pradesh, Assam, Chandigarh, Dadra & Nagar Haveli, Daman & Diu, Delhi, Himachal Pradesh, Jammu & Kashmir, Lakshadweep, Manipur, Meghalaya, Mizoram, Nagaland, Puducherry, Sikkim, Tripura and Uttarakhand. The period of analysis is 2005–06 to 2014–15.

After obtaining productivity score, a second-stage panel regression is resorted to find out the important factors responsible for enhancing growth prospects.

Average productivity of $x = \frac{f(x)}{x}$. Let it be maximized at $x = x^*$ where, $f'(x) = \frac{f(x)}{x}$. Taking $f'(x^*) = w$, the pseudo production function be defined as $R(x) = wx$ which exhibits CRS and is a ray through the origin.

GCS_p.

3.3 The Determinants of Productivity Score

The factors used to explain the variation of productivity score are the following:

(a) **Unfavourable or Poor Infrastructure**: to investigate whether poor infrastructure negatively affects productivity score considering the effects of the following variables, namely (i) proportions of schools without building (SWB), (ii) proportions of schools having no pucca building (NPB), (iii) proportion of classrooms in 'bad' condition in the schools (CBC), (iv) proportion of single teacher in school (SNTS), (v) proportion of para teachers in the school (PTS), proportion of schools with no drinking water (NDW). The variable para teachers are included because due to lack of sufficient number of full-time teachers many schools employ a significant number of para teachers and the basic question is whether the para teachers play any significant role in promoting TFPG?
(b) **Favourable Infrastructure**: to see whether the existence of favourable infrastructure positively affects TFPG comprising of the variables, namely (i) percentage of schools with common toilet (CT) and (ii) percentage of schools with girls toilet (GT).
(c) **Social Indicators**: to determine whether more inclusion of the backward classes into the system increases TFPG taking into account the effects of the following variables, namely (i) proportion of SC teacher to total teacher in the schools (SCT), (ii) proportion of ST teacher to total teacher (STT), (iii) proportion of SC enrolment (SCEN), (iv) proportion of ST enrolment (STEN), (v) proportion of female teachers to male teachers (FT) and (vi) proportion of girls enrolment to boys (GEN).
(d) **Policy Variables**: to investigate whether provision of more public facilities increase productivity score including the effects of the following variables like (i) proportion of students getting free textbooks in the schools (FTB), (ii) proportion of girls getting free textbooks to boys (GFTB), (iii) proportion of girls getting free stationary to boys (GFST), (iv) proportion of schools received School Development Grant (SDG) and (v) proportion of schools received Teaching Learning Material Grant (TLMG).
(e) **Macro-indicators**: to find out whether general economic environment of the state has something to do with its TFPG considering the impacts of inequality in distribution of income, i.e. Gini coefficient (GINI), the number of persons belonging to below poverty line (BPL), density of population (POPDEN) and the per capita net state development product of the districts (PCNSDP).

In the second-stage panel regression generated productivity score will be taken as a dependent variable and the above indicators as explanatory variables. Since the basic interest is to find out the impact of the individual explanatory variables, the panel regression with composite index for each of the used broad indicators like poor infrastructure and social indicators has not been tried out. It is possible that some of the individual variables may be significant while the others are not and the use of a composite index cannot differentiate between these possibilities.

3.4 Data Sources

This study considers secondary data from collected from ‘DISE Statistics (Elementary Education; State Report Cards)’, for the successive years: 2005–06 to 2014–15. This information is based on District Information System for Education (DISE) district raw data which was developed by the National University of Educational Planning and Administration (NUEPA). The data for the net per capita state domestic product has been collected from Ministry of Statistics and Programme Implementation (MOSPI), Government of India, while Gini coefficient data are collected from the Planning Commission and the Office of the Registrar General and Census Commissioner. The data for number of persons belonging to below poverty line has been taken from the statistics released by Ministry of Food and Consumer Affairs, Government of India. The data for per capita net state domestic product is collected from National Accounts Statistics.

4 Empirical Findings

4.1 Results of Estimation on Total Factor Productivity Growth

The mean value of the productivity index as generated for all the years taken together both for GCS and SCS&UT categories of the states separately and for primary and upper primary level is presented in Table 1. The results suggest that for GCS, MPI is higher for primary level education as compared to upper primary level. For SCS&UT, MPI is higher for upper primary level, as compared to primary. Tables 2, 3, 4 and 5 **report the** generated values of productivity index for each of the sample period 2005–06 to 2014–15 for CGS and SCS and for primary and upper primary level separately. Tables 2, 3, 4 and 5 also report the yearwise mean of MPI and the grand mean of MPI for all the years taken together for each of the cases GCS primary and upper primary, SCS&UT primary and upper primary, respectively.

Table 1 Average level of MPI of elementary education in India

Level	Categories	
	GCS	SCS&UT
Primary	1.186	1.086
Upper primary	1.2299	1.2585

Table 2 MPI in GCS primary in different years

States	2005–06	2006–07	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15
AP	0.975	0.976	0.969	1.014	1.003	1.048	1.001	0.97	1.145	0.997
BH	1.246	0.345	0.285	8.115	0.98	1.057	0.859	0.989	1.048	0.91
CHAST	0.979	0.463	1.966	1.231	0.822	1.144	0.83	0.879	0.91	0.922
Goa	1.06	0.93	1.091	0.849	0.972	0.982	1.357	1.232	1.03	0.946
GU	1.238	0.957	1.297	0.763	1.026	1.068	0.986	0.918	1.075	0.887
HA	1.167	1.195	1.068	1.053	0.941	1.079	1.053	0.997	0.993	0.869
JH	0.693	1.176	0.244	8.767	0.9	0.956	0.954	0.955	1.037	0.866
KA	1.051	0.628	1.439	0.818	1.027	0.992	1.847	0.602	1.212	0.828
KE	0.969	1.006	0.984	0.99	0.863	1.031	1.186	0.917	1.113	0.996
MP	0.42	1.022	0.177	8.655	1	0.919	0.941	0.984	0.808	0.857
MH	1.012	0.985	1.043	0.864	0.982	1.019	0.996	0.9452	0.991	0.973
OR	0.77	1.575	0.668	1.324	0.875	1.15	0.724	1.065	1.095	0.949
PU	1.013	1.021	1.021	0.993	1.197	1.116	1.012	0.986	0.979	0.964
RA	1.067	0.968	0.125	6.094	0.959	1.063	0.895	0.916	1.047	0.971
TN	1.085	0.94	1.015	0.96	0.965	1.062	0.918	0.92	1.034	1.028
UP	0.317	1.041	0.272	8.671	0.942	1.022	0.957	0.878	0.959	0.97
WB	0.989	1.017	1.028	0.898	0.895	0.994	0.939	2.727	0.949	1.012
Yearwise mean	0.94412	0.9556	0.8642	3.0622	0.9617	1.0395	1.0267	1.0517	1.025	0.9379

Grand mean of MPI for all the GCS and all the sample year taken together at primary level = 1.1869

Table 3 MPI in GCS upper primary in different years

States	2005–06	2006–07	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15
AP	1.322	0.722	7.286	0.072	1.053	1	2.147	0.550	0.935	1.089
BH	1.099	1.11	1.31	0.91	1.039	0.941	0.959	1.437	1.104	1.086
CHAST	0.856	0.992	2.691	0.458	0.914	1.134	0.949	1.015	0.966	1.016
Goa	0.928	0.416	1.159	0.726	1.027	0.965	1.699	0.972	1.148	1.045
GU	1.05	1.233	1.301	0.681	1.21	0.813	1.284	0.958	1.114	0.936
HA	1.239	1.113	1.266	0.444	1.081	0.986	1.037	0.981	1.169	0.87
JH	0.738	0.859	1.732	1.189	1.018	1.014	0.947	1.163	1.078	1.035
KA	1.003	1.243	0.665	1.027	1.027	1.096	1.85	0.975	1.118	0.724
KE	1.233	0.961	1.125	1.527	1.053	1.055	1.079	0.97	0.945	0.972
MP	0.741	0.85	1.973	0.601	1.032	1.049	1.027	0.951	0.965	0.954
MH	0.854	1.048	1.321	1.69	1.019	0.973	1.104	1.075	1.007	0.962
OR	1.04	1.931	0.965	0.575	1.133	1.154	0.086	1.124	1.141	1.046
PU	1.016	0.895	1.686	1.455	1.499	1.055	0.831	0.98	1.265	1.228
RA	0.991	1.012	1.312	0.666	1.093	1.071	0.989	1.095	1.018	1.065
TN	0.788	1.357	1.619	1.704	1.057	1.02	0.829	1.008	1	1.003
UP	0.943	1.034	2.343	1.443	1.028	0.979	1.113	1.021	1.033	0.905
WB	1.001	0.961	1.173	0.793	1.037	1.129	0.741	1.424	1.059	1.05
Yearwise mean	0.9907	1.0433	1.5861	0.9388	1.0894	1.0255	1.1141	1.012	1.0626	0.9992

Grand mean of MPI for all the GCS and all the sample year taken together at upper primary level = 1.0862

Table 4 MPI in SCS&UT primary in different years

States	2005–06	2006–07	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15
Andaman & Nicobar Islands	0.882	1.116	0.889	0.695	1.014	1.052	1.204	0.812	1.197	0.884
Arunachal Pradesh	0.488	1.094	0.17	7.629	0.983	0.998	0.891	0.909	1.494	0.987
Assam	0.578	1.242	0.168	6.756	1.207	1.063	0.709	0.97	1.217	0.875
Chandigarh	4.228	0.689	0.5	1.184	1.355	0.739	0.789	0.971	0.949	0.899
Dadra & Nagar Haveli	0.99	0.434	0.259	4.688	1.031	0.955	0.874	0.795	1.03	0.918
Daman & Diu	1.067	0.666	1.594	0.921	0.928	0.948	0.935	0.931	1.04	0.908
Delhi	0.548	0.985	1.381	1.011	1.059	1.041	0.973	0.938	0.994	0.982
Himachal Pradesh	0.683	0.949	0.965	0.885	1.016	0.999	0.85	0.959	1.036	0.915
Jammu & Kashmir	1.124	0.978	1.242	0.749	0.837	0.937	0.66	0.813	1.146	0.995
Lakshadweep	0.659	1.168	1.006	0.91	0.867	1.211	0.873	0.647	1.074	1.039
Manipur	0.887	1.144	0.148	7.211	0.837	0.996	0.892	0.845	0.927	0.983
Meghalaya	0.173	1.679	0.214	9.114	0.987	1.044	0.903	0.966	1.218	0.995
Mizoram	0.796	1.381	0.154	9.075	0.962	0.995	0.816	0.88	1.706	0.963
Nagaland	1.126	0.692	0.183	7.033	1.111	0.943	0.926	1.012	0.765	0.898
Puducherry	1.742	0.864	1.225	0.569	0.912	0.941	0.97	0.842	0.917	0.951
Sikkim	0.405	1.129	1.015	0.895	0.849	0.931	0.863	0.932	1.094	0.891
Tripura	1.092	0.909	0.156	7.578	1.06	0.888	0.648	0.949	1.822	0.986
Uttarakhand	0.617	1.133	0.977	0.858	1.02	0.961	0.883	0.938	0.969	0.934
Yearwise mean	1.0047	1.014	0.6833	3.7645	1.0019	0.9801	0.8699	0.8949	1.1441	0.9446

Grand mean of MPI for all the SCS&UT and all the sample year taken together at primary level = 1.229

Table 5 MPI in SCS&UT upper primary in different years

States	2005–06	2006–7	2007–08	2008–09	2009–10	2010–11	2011–12	2012–13	2013–14	2014–15
Andaman & Nicobar Islands	0.842	1.723	1.052	0.589	1.005	1	1.876	0.633	0.443	2.592
Arunachal Pradesh	1.688	0.733	1.412	0.788	1.669	0.934	0.804	1.112	3.355	1.092
Assam	1.356	1.247	1.556	0.621	1.049	0.961	0.853	1.058	1.064	0.897
Chandigarh	1.349	1.168	2.756	0.241	1.075	0.966	1.29	0.891	0.78	1.535
Dadra & Nagar Haveli	1.13	1.131	0.684	1.49	1.954	0.678	0.751	1.014	1.418	1.056
Daman & Diu	0.401	0.998	1.002	0.758	1.156	1.004	1.025	0.852	0.971	1.14
Delhi	0.811	0.978	1.407	0.814	1.021	1.053	1.156	1.112	1.022	0.941
Himachal Pradesh	0.943	1.013	1.441	0.467	1.058	0.944	0.771	1.043	1.27	0.987
Jammu & Kashmir	1.083	0.906	1.0507	0.665	0.714	1.174	0.699	0.651	1.377	1.012
Lakshadweep	0.794	0.931	1.463	0.508	1.394	1.052	1.324	0.692	0.958	0.833
Manipur	1.413	1.065	1.408	0.8	1.116	1.02	0.967	0.926	1.309	0.955
Meghalaya	1.16	1.433	1.651	0.589	1.061	1.09	0.801	1.273	1.581	1.134
Mizoram	0.843	1.272	2.759	0.393	0.813	0.972	0.933	1.016	1.723	1.07
Nagaland	1.238	0.8	1.114	0.489	1.038	0.951	0.789	1.101	1.111	0.936
Puducherry	1.647	0.415	9.237	0.091	0.995	1	1.072	0.836	0.296	6.923
Sikkim	0.556	1.091	19.536	0.055	1.45	0.793	4.479	0.667	1.473	1.153
Tripura	1.054	1.05	1.594	0.436	1.202	0.636	1.04	2.085	0.461	0.98
Uttarakhand	0.914	1.272	1.3	0.588	1.025	0.994	0.886	0.935	1.011	0.984
Yearwise mean	1.0679	1.0681	2.9124	0.5768	1.1522	0.9570	1.1953	0.9943	1.2013	1.4566

Grand mean of MPI for all the SCS&UT and all the sample years taken together at upper primary level = 1.2585

4.2 *Decomposition of the Productivity Score*

The generated value of MPI is decomposed into technical change, technical efficiency change and scale change following the method as discussed above.

The yearwise average values of each of the changes taking into account all GCS, SCS&UT separately and for primary and upper primary level are presented in Tables 6, 7, 8 and 9.

The results of the decomposition suggest that on average productivity changes are mainly driven by technical changes and efficiency changes. The same result holds for primary and upper primary and for GCS and SCS&UT.

Table 6 Yearwise decomposition of total factor productivity change into efficiency change, technical change and scale change: GCS primary (Average of all states)

Year	Efficiency change	Technical change	Scale change	Total factor productivity change
2006–07	0.7948	1.1881	0.8647	0.9442
2007–08	1.0536	0.9075	1.0082	0.9556
2008–09	0.7445	1.2351	0.7709	0.8642
2009–10	4.1778	0.7729	2.0929	3.0623
2010–11	0.9661	0.9952	0.9561	0.9617
2011–12	0.9942	1.0508	1.0095	1.0413
2012–13	1.0482	0.9722	1.0061	1.0268
2013–14	0.9571	1.0838	0.9655	1.0498
2014–15	0.9932	1.0339	1.0035	1.0250

Table 7 Yearwise decomposition of total factor productivity change into efficiency change, technical change and scale change: GCS upper primary (Average of all states)

Year	Efficiency change	Technical change	Scale change	Total factor productivity change
2006–07	0.8560	1.1871	0.9516	1.0026
2007–08	1.0449	1.0201	1.0348	1.0637
2008–09	1.1401	1.5548	1.0475	1.7429
2009–10	1.0294	0.6429	1.0131	0.6534
2010–11	1.0147	0.8343	1.0016	0.8395
2011–12	1.0001	1.0256	0.9902	1.0255
2012–13	1.0325	1.0984	0.9512	1.1406
2013–14	0.9151	1.1331	0.9651	1.0414
2014–15	0.9519	1.1167	0.9778	1.0626

Table 8 Yearwise decomposition of total factor productivity change into efficiency change, technical change and scale change: SCS&UT primary (Average of all states)

Year	Efficiency change	Technical change	Scale change	Total factor productivity change
2006–07	0.9145	1.0259	0.9173	0.9478
2007–08	1.2371	0.8537	1.1792	0.9982
2008–09	0.8269	1.0774	0.9290	0.9018
2009–10	1.3404	1.0931	1.0222	1.5299
2010–11	1.1652	1.0617	1.0063	1.2741
2011–12	0.9899	1.0302	0.9903	1.0183
2012–13	1.0533	0.8691	1.0487	0.9077
2013–14	0.9884	0.9223	1.0119	0.9115
2014–15	0.9669	1.1785	0.9784	1.1411

Table 9 Yearwise decomposition of total factor productivity change into efficiency change, technical change and scale change: SCS&UT upper primary (Average of all states)

Year	Efficiency change	Technical change	Scale change	Total factor productivity change
2006–07	0.8842	1.4163	0.9104	1.1468
2007–08	1.5228	0.5952	1.3279	0.9322
2008–09	0.7720	3.1492	0.7906	3.0177
2009–10	1.5877	0.3577	1.5475	0.5771
2010–11	1.2968	0.6780	1.2758	0.7914
2011–12	1.0058	0.9983	1.0041	1.0057
2012–13	0.9616	1.1220	0.9131	1.0731
2013–14	1.0358	0.9211	1.0092	0.9441
2014–15	0.9891	1.3163	0.9899	1.3239

4.3 Analysis of the Determinants of Total Factor Productivity Growth

To test for appropriateness of the assumption of fixed effect vis. a vis. the random effect model, Haussmann's specification test is performed for each of the regressions which strongly rejects the assumption of fixed effect model in favour of random effect model. The results of estimation are presented in Tables 10, 11, 12 and 13. The tables also report Haussmann test results for each of the cases.

The four panels consist of MPI for each of the 17 GCS states for primary and upper primary levels and also for 18 SCS&UT states for primary and upper primary levels, over the period 2005–06 to 2014–2015. Different regressions are tried out and not all the explanatory variables are significant. The best-fitted model is reported containing

Table 10 Significant variables influencing MPI of GCS primary

Variables	Coefficients	*P*-value	Goodness of fit of model
CBC	−0.0515	0.0000	Adjusted *R*-square 0.7428
SNTS	−0.1316	0.0222	
PTS	0.0085	0.0000	
NDW	−0.0198	0.0000	
GFST	0.0024	0.0000	
TLMG	0.0012	0.0090	
SWB*SNTS	−0.0566	0.0249	
Redundant fixed effects tests			
	Statistic	Prob.	
Cross-section F	0.491079	0.9484	

Table 11 Significant variables influencing MPI of GCS upper primary

Variables	Coefficients	*p*-value	Goodness of fit of model
SWS	−0.0191	0.0000	Adjusted *R*-square 0.3426
CBC	−0.0128	0.0473	
SDG	0.0132	0.0062	
PCNSDP	0.0012	0.0075	
Redundant fixed effects tests			
	Statistic	Prob.	
Cross-section F	0.394094	0.9823	

Table 12 Significant variables influencing MPI of SCS&UT primary

Variables	Coefficients	*p*-value	Goodness of fit of model
GFST	0.0006	0.0070	Adjusted *R*-square 0.4225
NDW	−0.0558	0.0112	
SWB*SNTS	−0.0039	0.0000	
Redundant fixed effects tests			
	Statistic	Prob.	
Cross-section F	1.360164	0.1633	

only the significant explanatory variables. While doing the regression, one needs to check the possibility of the possible correlation among the explanatory variables. The correlation matrix of the significant explanatory variables is reported for each of the panel regression. The correlation matrix of the significant explanatory variables GCS primary, GCS upper primary, SCS&UT primary, SCS&UT upper primary is reported in Tables 14, 15, 16 and 17, respectively. As can be seen from the reported figures, the correlation between the significant explanatory variables is either not significant or

Table 13 Significant variables influencing MPI of SCS&UT upper primary

Variables	Coefficients	p-value	Goodness of fit of model
SDG	0.0062	0.0003	Adjusted R-square 0.3710
SWB	−0.0017	0.0002	
SWB*SNTS	−0.00043	0.0003	
Redundant fixed effects tests			
	Statistic	Prob.	
Cross-section F	0.723219	0.7758	

Table 14 Correlation matrix of the significant variables influencing MPI of GCS primary

		TLMG	GFST	PTS	CBC	SNTC	SWB	NDW
TLMG	Pearson correlation	1	0.326**	−0.108	0.137	0.024	0.133	0.226**
	Sig. (2-tailed)		0.000	0.161	0.075	0.751	0.084	0.003
GFST	Pearson correlation	0.326**	1	−0.068	0.367**	−0.048	0.514**	0.406**
	Sig. (2-tailed)	0.000		0.381	0.000	0.537	0.000	0.000
PTS	Pearson correlation	−0.108	−0.068	1	−0.021	0.035	0.268**	0.222**
	Sig. (2-tailed)	0.161	0.381		0.789	0.646	0.000	0.004
CBC	Pearson correlation	0.137	0.367**	−0.021	1	0.004	0.419**	0.365**
	Sig. (2-tailed)	0.075	0.000	0.789		0.959	0.000	0.000
SNTC	Pearson correlation	0.024	−0.048	0.035	0.004	1	0.015	0.198**
	Sig. (2-tailed)	0.751	0.537	0.646	0.959		0.849	0.010
SWB	Pearson correlation	0.133	0.514**	0.268**	0.419**	0.015	1	0.582**
	Sig. (2-tailed)	0.084	0.000	0.000	0.000	0.849		0.000
NDW	Pearson correlation	0.226**	0.406**	0.222**	0.365**	0.198**	0.582**	1
	Sig. (2-tailed)	0.003	0.000	0.004	0.000	0.010	0.000	

**Correlation is significant at the 0.01 level (2-tailed)

Table 15 Correlation matrix of the significant variables influencing MPI of GCS upper primary

		SWB	CBC	SDG	PCNSDP
SWB	Pearson correlation	1	−0.020	−0.208**	−0.051
	Sig. (2-tailed)		0.797	0.007	0.510
CBC	Pearson correlation	−0.020	1	0.420**	−0.003
	Sig. (2-tailed)	0.797		0.000	0.965
SDG	Pearson correlation	−0.208**	0.420**	1	−0.069
	Sig. (2-tailed)	0.007	0.000		0.369
PCNSDP	Pearson correlation	−0.051	−0.003	−0.069	1
	Sig. (2-tailed)	0.510	0.965	0.369	

**Correlation is significant at the 0.01 level (2-tailed)

Table 16 Correlation matrix of the significant variables influencing MPI of SCS&UT primary

		GFST	NDW	SWB	SNTS
GFST	Pearson correlation	1	0.071	0.351**	−0.008
	Sig. (2-tailed)		0.341	0.000	0.915
NDW	Pearson correlation	0.071	1	0.120	0.376**
	Sig. (2-tailed)	0.341		0.108	0.000
SWB	Pearson correlation	0.351**	0.120	1	0.182*
	Sig. (2-tailed)	0.000	0.108		0.014
SNTS	Pearson correlation	−0.008	0.376**	0.182*	1
	Sig. (2-tailed)	0.915	0.000	0.014	

*Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)

Table 17 Correlation matrix of the significant variables influencing MPI of SCS&UT upper primary

		SDG	SWB	SNTS
SDG	Pearson correlation	1	−0.095	−0.044
	Sig. (2-tailed)		0.203	0.560
SWB	Pearson correlation	−0.095	1	0.127
	Sig. (2-tailed)	0.203		0.089
SNTS	Pearson correlation	−0.044	0.127	1
	Sig. (2-tailed)	0.560	0.089	

if significant is not very strong. However, in this context, it is worthwhile to mention one important result; the extensive Monte Carlo Simulation reveals that a DEA-based procedure in the first stage followed by regression analysis in the second-stage yield consistent estimators of the impact of contextual variables if the contextual variables, affecting productivity, are independent of the input variables *(although they may be correlated with each other).* Additionally, two-stage DEA-based methods with OLS in the second stage significantly outperform the parametric methods (Banker and Natarajan 2008).

Results of panel regression suggest that infrastructural variables, policy variables, school-specific variables and also the state-level macro-aggregates are important in explaining MPI.

The estimated results for *GCS primary level* suggest that MPI is positively influenced by the policy variables like proportion of schools receiving teaching learning material grant (TLMG), ratio of girls getting free stationary to boys (GFST), the number of para teachers in schools (PTS) and negatively influenced by the poor infrastructural variable like proportion of classroom in bad condition (CBC), proportion of single-teacher school (SNTS), proportion of school without no drinking water (NDW) and the combined effect of proportion of schools without building (SWB), and proportion of single-teacher school (SNTS). The result that the TFPG negatively depends on the combined influence of proportion of school without building and the proportion of single-teacher school basically implies that the effect of a change in the proportion of single-teacher school on TFPG will depend on the magnitude of the proportion of school without building. Similarly, the effect of a change in the proportion of school without building on TFPG will depend on the magnitude of the proportion of single-teacher school.

The results of estimation for *GCS upper primary* level suggest that MPI is positively influenced by the policy variables like proportion of schools receiving school development grant (SDG), the macro-indicator like per capita net state domestic product (PCNSDP), and negatively influenced by the poor infrastructural variables like proportion of classroom in bad condition (CBC) and the proportion of schools without building (SWB).

The estimated results for *SCS&UT primary level* suggest that MPI is positively influenced by the policy variables like proportion of girls getting free stationary to boys (GFST) and negatively influenced by the poor infrastructural variables; like proportion of school without no drinking water (NDW) and the combined effect of proportion of schools without building (SWB) and proportion of single-teacher school (SNTS).

The results of estimation for *SCS&UT upper primary* level suggest that MPI is positively influenced by the policy variables like proportion of schools receiving school development grant (SDG). MPI is negatively influenced by the poor infrastructural variables like proportion of schools without building (SWB) and by the combined influence of proportion of schools without building (SWB) and proportion of single-teacher school (SNTS). The effect of joint interaction variables for SCS&UT primary and upper primary level can similarly be interpreted as in the case of GCS primary.

5 Conclusions

The main thrust of the paper is to find out the extent of total factor productivity growth (TFPG) of elementary education in India during 2005–06 to 2014–15. TFPG is estimated separately for primary and upper primary level of education. The major departure of this paper from the existing literature is its approach in estimating TFPG by constructing two frontiers (i) General Category States (GCS) and (ii) Special Category States (SCS) and Union Territories (UT) as these groups are not homogeneous and operate under different fiscal and social environment. TFPG is measured by MPI and is generated by using non-parametric data envelopment analysis (DEA). The obtained results support that for GCS, MPI is higher for primary level of education as compared to upper primary level. But for SCS&UT, MPI is higher for upper primary level as compared to primary. TFPG is decomposed into technical change, efficiency change and scale efficiency change. The decomposition results suggest that TFPG has mainly been facilitated by technical change and efficiency.

After generating MPI score, a second-stage panel regression is resorted to, for finding out determinants of MPI using random effect panel model as supported by Haussmann's specification test. The results of panel regression suggest that poor infrastructural variables, school-specific variable, policy variables and macro-aggregate significantly affect MPI. The poor infrastructural variables which negatively and significantly affect TFPG are: (i) proportion of classrooms in bad condition (for GCS primary and upper primary level), (ii) proportion of schools without building (for GCS upper primary and SCS&UT upper primary level), (iii) proportion of schools without drinking water (for GCS primary and SCS&UT primary level). Apart from the individual effect of these poor infrastructural variables, there is some interaction effect of the proportion of schools without building and the proportion of single-teacher school (GCS primary, SCS&UT primary and upper primary) implying that sensitivity of TFPG with respect to proportion of single-teacher school depends on the magnitude of the proportion of schools without building. The policy variables which positively and significantly affect TFPG are: (i) proportion of schools receiving teaching learning material grant (for GCS upper primary and SCS&UT upper primary level), (ii) girls getting free stationary to boys (for GCS primary and SCS&UT upper primary level), (iii) proportion of school receiving school development grant (for GCS upper primary and SCS&UT upper primary level). TFPG also depends on the school-specific variables like the proportion of para teachers in total number of teachers (for GCS primary level) and also on the macroeconomic aggregate like per capita net state domestic product (for GCS primary level).

The results of the analysis support the role of the government in promoting TFPG. For example, the government can increase TFPG by providing school development grant, teaching learning material grants to schools. The incentive by the government to the girl students by providing more freer stationary as compared to boys will also promote productivity. For developing country like India, since government operates under many financial constraints, it is not always possible for the government to provide adequate infrastructure to schools. As a result, many schools have to operate with extremely poor infrastructure like: some schools operate without building; some schools have poor condition of classrooms; some schools run only with a single classroom; some schools have just a single teacher and the like. The analyses of the present paper suggest that there is a negative relationship between TFPG and the poor infrastructural variables as mentioned above. Thus, in order to promote TFPG, there is also a need to improve the infrastructural facilities of the school like condition of classrooms, providing school building, providing drinking water facilities. There is also a need to provide more number of permanent teachers to school, as some of the schools operate with the basis of single teacher and there is a negative relationship between single-teacher school and TFPG. Further, there is also negative joint interaction effect of proportion of single-teacher school and the proportion of school without building on TFPG. The TFPG also depends on the school-specific variables like number of para teachers. The need to employ more number of permanent teachers can also be justified by the fact that in the absence of sufficient number of teachers, the schools have to employ para teachers and the para teachers positively and significantly promote TFPG. The result that the TFPG negatively depends on the combined influence of proportion of school without building and the proportion of single-teacher school basically implies that the effect of a change in the proportion of single-teacher school on TFPG will depend on the magnitude of the proportion of school without building. This result in turn implies that increase in either the number of teachers or increase in number of school buildings separately and independently will not be sufficient for increase in TFPG. The effect of an increase in number of reduction in the proportion of single-teacher school on TFPG will properly be felt if the government simultaneously give more attention for the preparation of school building as well.

Since TFPG of GCS primary level depends positively on per capita net state domestic product, any suitable policy by the government which can promote per capita net state domestic product will in turn enhance TFPG of GCS primary level. The estimated results suggest that the factors influencing TFPG varies between GCS and SCS&UT and this result is true both for primary and upper primary level of education which in turn strengthen the argument for creating two separate frontiers for the two groups of states.

There are some limitations of this study. The quality of the delivery of teaching services will also depend on teacher's absenteeism, which has been a major issue nowadays. Teacher's absenteeism basically led to efficiency and productivity difference between private and public school. In order to get the figures on Teacher's absenteeism, one needs to have primary survey data. Further, it is true that performance of the student will also depend on opportunity cost of going to school, which will basically depend on the socio-economic condition of the household. Since this study is based on state-level secondary data, general condition of learning that the student face outside the school is captured by general economics condition of the state, i.e. per capita state net domestic product of the state and inequality in the distribution of income prevailing in the state and the number of population lying below the poverty line. For collection of household's socio-economic data like household income and parent's education, it is necessary to collect primary survey data, which may be one future direction of research.

Acknowledgements The paper is a part of the major project done under UPE-II Programme of Jadavpur University, funded by UGC during 2016–19.

Bibliography

Aigner, D. J., Lovell, C. A. K., & Schmidt, P. J. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21–37.

Anand, S., & Sen, A. K. (1996). *Sustainable human development: Concepts and priorities, office of development studies* (pp. 1–57). Discussion Paper Series, No. 1, United Nations Development Programme, New York. (Originally published as HDRO Occasional Paper 8 for Human Development Report 1994).

Arjomandi, A., Salleh, M. I., & Mohammadzadeh, A. (2015). Measuring productivity change in higher education: An application of Hicks–Moorsteen total factor productivity index to Malaysian public universities. *Journal of the Asia Pacific Economy, 20*(4), 630–643.

Banker, R., Charnes, A., & Cooper. (1984). Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.

Banker, R. D., & Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research, 56*(1), 48–58.

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy, 98*(S5), 103–125.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics, 106*(2), 407–443.

Bessent, A. M., Bessent, E. W., Kennington, E. W., & Regan, B. (1982). An application of mathematical programming to the productivity in the Houstan independent school district. *Management Science, 28*(12), 1355–1367.

Bessent, A. M., Bessent, W., Elam, J., & Long, D. (1984). Educational productivity employs management science method to improve educational quality. *Interfaces, 14,* 1–8.

Bonesronning, H., & Rattsq, J. (1994). Efficiency variation among Norwegian high schools: Consequences of equalization policy. *Economics of Education Review, 3*(4), 289–304.

Bradley, S., Johnes, J., & Little, A. (2010). Measurement and determinants of efficiency and productivity in the further education sector in England. *Bulletin of Economic Research, 62*(1), 0307–3378.

Caves, D. W., Christensen, L. R., & Diewert, D. W. (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica, 50*(6), 1393–1414.

Central Statistical Office, Office of the Registrar General & Census Commissioner, Census Data, 2011, Government of India.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2,* 429–444.

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of DEA to program follow-through. *Management Science, 27*(6), 668–697.

District Information System for Education, National University of Educational Planning and Administration, State Report Cards, different issues.

Dutta, S. (2012). Evaluating the technical efficiency of elementary education in India: An application of DEA. *IUP Journal of Applied Economics, 11*(2), 31–47.

Färe, R., Grosskopf, S., Norris, M., & Zhang, P. (1994). Productivity growth, technical progress and efficiency changes in industrial countries. *American Economic Review, 84*(1), 66–83.

Fare, R., Grosskopf, S., & Weder, W. L. (1989). Measuring school district performance. *Public Finance Quarterly, 17,* 409–428.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society, 120*(3), 253–281.

Ghose, A. (2017). *Efficiency of elementary education in India: Empirical evidence using a non-parametric data envelopment Approach*. Springer.

Government of India. (1986). *New Education Policy*.

Government of India. (2016). *New Education Policy (NEP)*, 30th June 2016. www.my.gov.in.

Government of India. (2019). *National Education Policy*.

Johnes, J. (2008). Efficiency and productivity change in the English higher education sector from 1996–97 to 2004–05. *The Manchester School, 76*(6), 1463–6786, 653–674 (Check).

Johnes, J., & Li, Y. (2008). Measuring the research performance of Chinese higher education institutions using data envelopment analysis. *China Economic Review, 19*(4), 679–696.

Kumbhakar, S. C., & Lovell, C. A. K. (2003). *Stochastic frontier analysis.* Cambridge: Cambridge University Press.

Ludwin, W., & Guthrie, T. (1989). Assessing productivity with data envelopment analysis. *Public Productivity Review, 12,* 361–372.

Lucas, R. E., Jr. (1988). On the mechanism of economic development. *Journal of Monetary Economics, 22*(1), 3–42.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics, 107*, 407–437.

Parteka, A., & Wolszczak-Derlacz, J. (2013). Dynamics of productivity in higher education: Cross-European evidence based on bootstrapped Malmquist indices. *Journal of Productivity Analysis, 40,* 67–82.

Ray, S. C. (2004). *Data envelopment analysis: Theory and technique for economics and operation research*. Cambridge: Cambridge University Press.

Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress and efficiency changes in industrialised countries: Comment. *American Economic Review, 85*(5), 1033–1039.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy, 94*(5), 1002–1038.

Romer, P. M. (1990). Endogenous technological change. *The Journal of Political Economy, 98*(5), Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems, S71–S102.

Sen, A., & Anand, S. (1995). Gender inequality in human development: Theories and measurement, background papers: Human development report 1995. United Nations Development Programme (New York).

Sengupta, A., & Pal, N. P. (2010). Primary education in India: Delivery and outcome—A district level analysis based on DISE Data. *Journal of Educational Planning and Administration, XXIV*(1), 5–21.

Sengupta, A., & Pal, N. P. (2012). Assessing the primary schools multi-dimensional approach: A school level analysis based on Indian data. *International Journal of Educational Development, 32*(2), 264–272.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review, 51*(1), 1–17.

Tyagi, P., Yadav, S. P., Singh, S. P. (2009). *Efficiency analysis of schools using DEA: A case study of Uttar Pradesh state in India*, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.536.9952.

Varian, H. (1984). The non-parametric approach to production analysis. *Economica, 52*(3), 579–599.

Gulf War Syndrome: The Plight of War Veterans

Kajal Lahiri and Hua Lin

Abstract The 2001 National Survey of Veterans, conducted ten years after the end of the Gulf War, provides invaluable information on the longtime chronic health effects of the war. Using the semi-parametric generalized additive model to control for different age distributions in our study subpopulations, we find that it is not the Gulf War deployment or combat per se, but the exposure to toxic chemicals that is the root cause of the myriad of health problems faced by many Gulf War veterans. The simultaneous presence of multiple ailments for these veterans, known as Gulf War syndrome, is also confirmed. Our study provides important additional evidence on the complexity in identifying the genuine health problems faced by the Gulf War veterans that came from widely varying socioeconomic backgrounds. The methodology developed in this paper can be used to identify the presence of, not one but, multiple morbidities due to the simultaneous presence of a number of health problems in other war veterans.

Keywords Gulf War veterans · Generalized additive model · Chemical exposure

JEL Classification I12 · I18 · C49

1 Introduction

The past twenty-five years have witnessed a compelling story of suffering by veterans who served in the 1990–1991 Persian Gulf War.[1] The conditions affecting Gulf War veterans, like many wars around the globe, appear to be complex, involving diverse

[1]For instance, the labor force participation rate for the disabled Gulf War veterans is only around 77.8% in 2009. See Bureau of Labor Statistics (2009).

K. Lahiri (✉)
Department of Economics, University at Albany—SUNY, Albany, NY 12222, USA
e-mail: klahiri@albany.edu

H. Lin
Department of Pediatrics, University of Texas Southwestern Medical Center,
Dallas, TX 75390, USA

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_19

debilitating symptoms associated with multiple organs and physiological systems. The traditional scientific consensus in the years immediately following the war was that psychiatric illness, combat experience, or other deployment-related stressors are the likely reasons for the Gulf War ailments. However, many observers were unconvinced that wartime stress could adequately explain the symptoms of the Gulf War veterans. This is because the war only had four days of ground combat and most of the veterans were never in the combat area. The US Department of Veterans Affairs (VA) has been under increasing pressure from the veteran groups to de-emphasize the view that stress and trauma were chief drivers of Gulf War illness. The US government passed a law in 1998 to conduct an extensive research on Gulf War symptoms. In accordance with this law, Congress and the VA established the Research Advisory Committee on Gulf War Veterans' Illnesses in 2002. One of the prominent findings of this federal panel is a conclusion, based on recent epidemiologic studies, that there is a "probable link" between toxic chemical exposure such as sarin gas and the mysterious ailments that continues to inflict the surviving Gulf War veterans. This conclusion departs from the past official consensus that stress is the possible explanation for Gulf War illnesses. The proponents of the stress theory maintain that it is too early to rule out any specific cause since it is difficult to determine which troops were exposed to what chemicals. Despite recent advances in their understanding of the long-term effect of some of these chemicals, the federal panel concluded that their finding is inconclusive due to the lack of objective government data on Gulf War veterans. They recommended that government should give priority to research studies capable of "making full use of existing federal data resources related to Gulf War veterans' health and military service," and that there is an "urgent need" for additional studies to "reevaluate the association of combinations of neurotoxic exposures with chronic illness in Gulf War veterans." Due to the obvious relevance of the Gulf War experience to recent Iraq and Afghanistan wars, future deployments, and homeland security, the report also recommended that the VA invest at least $60 million over the next few years for additional Gulf War illness research. Because of the limitations in the scope of previous research and the lack of effective follow-up study on long-term chronic health problem for Gulf War veterans, the etiologic basis and clinical significance of Gulf War veterans remain unclear.

The 2001 National Survey of Veterans (*NSV 2001*), the latest comprehensive national survey of veterans, deserves special attention. It provides rich information on Gulf War veterans' health and socioeconomic status, 10 years after the war. Since most studies on Gulf War veterans are based on data collected in the 1990s, this 2001 veteran's data will extend previous research to reveal the longtime potential effects of Gulf War. Quite surprisingly, a preliminary analysis of the *NSV 2001* data revealed that the Gulf War veterans are systematically better in self-reported health status and in terms of the incidence of typical Gulf War-related ailments than the non-Gulf War veterans are. However, the age distribution and the socioeconomic status of the two groups are quite different. Thus, unless the effects of these confounders are appropriately and adequately controlled for, the true health effects of the Gulf War deployment cannot be delineated.

The analysis presented here utilizes a semi-parametric estimation approach to unravel the complexity of Gulf War veterans' health problems using a very comprehensive nationally representative data on all veterans. We study if the veterans who were deployed in the Gulf War have significantly more health problems than other veterans, and whether chemical exposure can be a reason for the typical Gulf War symptoms.

Our study contributes to the existing literature in two ways: We provide epidemiological evidence on the prevalence and nature of Gulf War illnesses. Methodologically, we use generalized additive model (GAM) to control for the age heterogeneity in our national sample of all veterans that would otherwise confound the true effect of chemical exposure on the Gulf War veterans.

2 Study Motivation and the Data

2.1 The National Survey of Veterans

NSV 2001 is the fifth in a series of comprehensive nationwide surveys designed to investigate the demographic profile of US veterans and to help the Department of Veterans Affairs (VA) to plan its future programs and services for veterans. Data collection began on February 12, 2001, and ended on November 12, 2001.[2] A total of 19,961 interviews enter our analysis, consisting of 2267 Gulf War veterans (11.36%) and 17,694 non-Gulf War veterans (88.64%).

2.2 Early Surprise: Preliminary Results from the Survey Data

The main objective of our study is to evaluate the health status of Gulf War veterans and compare them with non-Gulf War veterans. Table 1 provides percent distribution of self-reported health status reported by veterans from our data. Surprisingly, more Gulf War veterans evaluated their health status as "excellent" (15.70% vs. 12.13%), "very good" (28.28 vs. 22.32), and "good" (31.27 vs. 28.72) than non-Gulf War veterans. Much lower percent of Gulf War veterans report poor health compared with other veterans. The p-value for χ^2 test is less than 0.0001, indicating a very significant difference between these two groups. Furthermore, for the typical Gulf War syndromes (viz. gastrointestinal, depression, fatigue, chronic pain, ENT, and concentration problems), Fig. 1 indicates that the percentage of Gulf War veterans who have multisymptoms is almost the same as non-Gulf War veterans (39.35 vs. 38.32). For typical Gulf War symptoms, especially depression, fatigue, and concentration

[2]The next NSV was released on 2010, which did not collect all the information collected in NSV 2001 for the purpose of this study. Also, many of the serious sufferers of the 1990–1991 Gulf War would have died by 2010, creating a serious attrition problem.

Table 1 Distribution of general health status by veterans*

Self-evaluated health status	Gulf War veterans ($n = 2267$)	Non-Gulf War veterans ($n =$ 17,694)
Excellent	15.70	12.13
Very good	28.28	22.32
Good	31.27	28.72
Fair	18.31	20.19
Poor	6.18	12.72

*The hypothesis of independence of deployment status and perceived health status is rejected at 5% level of significance, $\chi^2 =$ 215.0868 ($p < 0.0001$)

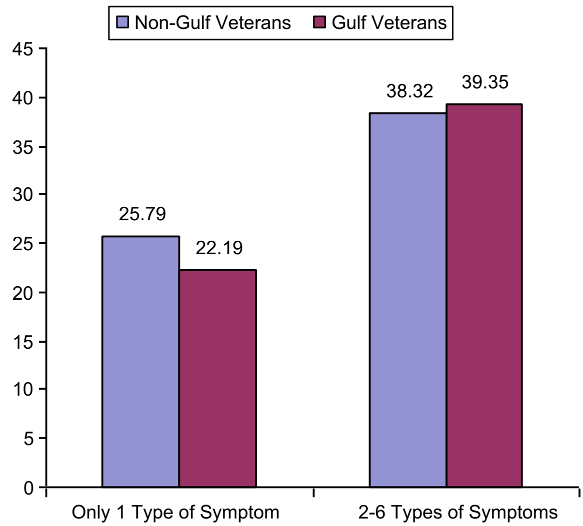


Fig. 1 US Gulf War veterans reporting typical symptoms (%)

problems, Gulf War veterans are significantly less likely to develop these ailments (Table 2). The better health status of Gulf War veterans is further confirmed by the comparison of common medical problems such as cancer, high blood pressure, diabetes, stroke, heart disease, and arthritis. In addition, Table 3 shows that Gulf War veterans have smaller possibility to have all of these common medical problems, including PTSD, one of the main contributing factors for Gulf War illnesses supported by some researchers.

As a first look, these results from the 2001 survey contradict the previous findings on the relative health status of Gulf War veterans and raise an essential question: "Are Gulf War veterans truly more ill compared with other veterans?" The question merits a deeper look into the data.

Table 2 Gulf War veterans reporting typical Gulf War syndromes (%)

	Gulf War veterans	Non-Gulf veterans	*p*-value
Fatigue	39.13	45.24	<0.0001
Depression	23.95	26.40	0.0125
Difficulty concentrating	14.16	16.92	0.0009
Gastrointestinal problem	19.14	17.33	0.0323
ENT problem	20.56	15.51	<0.0001
Chronic pain	25.89	21.62	<0.0001

Table 3 Prevalence rate (%) of US Gulf War veterans reporting common medical problems

	Gulf War veterans	Non-Gulf veterans	*p*-value
Diabetes	3.75	14.98	<0.0001
Kidney problem	8.43	14.41	<0.0001
Stroke	0.79	4.33	<0.0001
Heart disease	7.41	21.36	<0.0001
Cancer	2.78	9.67	<0.0001
High blood pressure	18.00	42.11	<0.0001
Lung problem	7.85	12.63	<0.0001
Arthritis	23.82	34.71	<0.0001
Liver	2.16	2.45	0.2058
PTSD	6.79	7.31	0.3605

2.3 *Characteristics of 2001 NSV Data*

The survey of Gulf War veterans within 10 years of deployment gives us an opportunity to study their health when they are alive and suffering. Unlike previous research on the Gulf War veterans, NSV 2001 is a comprehensive survey and is aimed at providing an overall look at all veterans' health problems in their socioeconomic context. However, it is not specially designed for studying Gulf War illnesses. Therefore, for finding the true effects of Gulf War deployment using this survey, we have to control for potential confounders like the age distributions of the respective subpopulations. Previous studies on Gulf War veterans were based on data concentrating on Gulf War veterans and considered possible confounding effects during data collection design. For example, Kang et al. (2000, 2002), Steele (2000), and Cherry et al. (2001) chose the veterans who served in non-Gulf locations during the same period time of Gulf War as comparison group. The demographic characters of the two comparison groups are, as a result, similar or matched. The *NSV 2001* includes veterans from all periods. One advantage of this comprehensive data set is that we have information on the health status of many non-Gulf War veterans who were exposed to chemicals and also on many Gulf War veterans who were not exposed to chemicals. On the other hand, the non-Gulf War veterans presumably fought longer wars and have had longer time to adjust to them. Thus, their circumstances are different from those of Gulf War veterans. In our comparison, we use a number of observable characteristics to control for these heterogeneities. We also report regressions with only Gulf War veterans.

Table 4 provides sociodemographic characteristics of *NSV 2001* sample of veterans. Compared to non-Gulf veterans, the Gulf War veterans are much younger (40.8 vs. 61.6 years). This is not surprising, given the general design for the survey; Gulf War was the final war period surveyed in *NSV 2001*. Figures 2 and 3 provide direct comparison of the age distributions of Gulf War and non-Gulf War veterans, respectively. We see that less than 10% Gulf War veterans are older than 55, whereas the majority of the non-Gulf War veterans are over age 50. As a result, finding from

Table 4 Distribution of sociodemographic characteristics in 2001 national survey of veterans

Characteristics	Gulf War veterans (n = 2267)	Non-Gulf War veterans (n = 17,694)	p-value
Age (mean age in years in 2001)	40.8	61.6	<0.0001
Gender			<0.0001
Male	83.15	94.94	
Female	16.85	5.06	
Education level			<0.0001
Less than HS	1.76	13.90	
HS diploma	22.45	29.24	
Post-HS training	44.55	32.02	
BA degree or higher	31.14	24.60	
Financial income			<0.0001
$0 to $10,000	3.79	8.53	
$10,001 to $20,000	7.72	18.14	
$20,001 to $30,000	12.53	17.66	
$30,001 to $40,000	15.88	13.33	
$40,001 to $50,000	15.00	10.87	
Over $50,000	45.08	31.46	
Race			<0.0001
White	72.08	83.69	
Black	16.63	7.86	
Other	11.29	8.45	
Chemical exposure	34.10	21.67	<0.0001

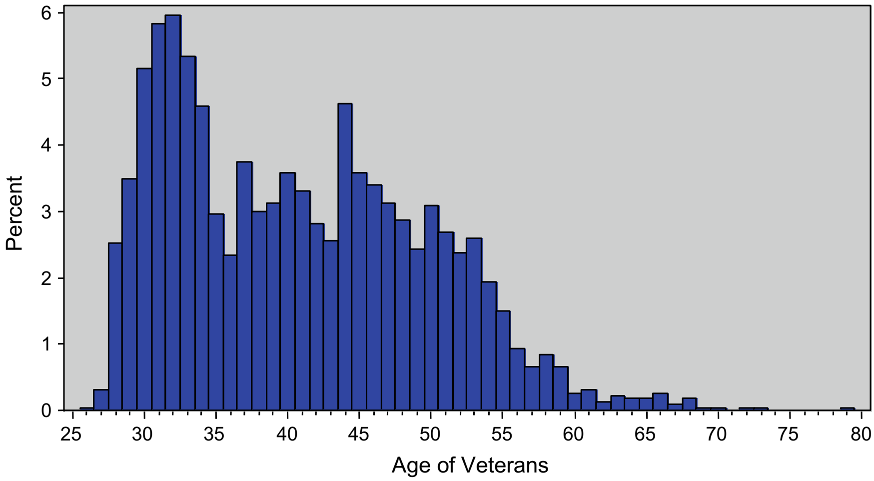


Fig. 2 Age distribution of Gulf War veterans in 2001 NSV

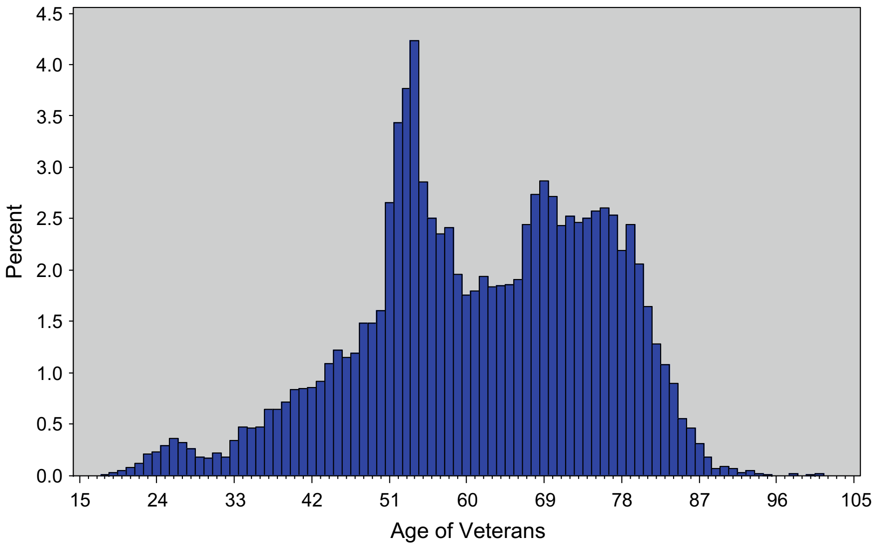


Fig. 3 Age distribution of non-Gulf War veterans in 2001 NSV

NSV 2001 might be affected by the "age effect" that arises from the fact that younger veterans tend to be healthier and less likely to develop illnesses than older veterans do.

The education level and the economic status of the Gulf War veterans are also much higher than that of the non-Gulf veterans. The percentage of whites compared to other races is also significantly smaller for Gulf veterans (72.08% vs. 83.69%). We also find that relatively more number of females is Gulf veterans compared to other veterans (16.85% vs. 5.06%). It is clear that unless we control for these confounders, the effect of chemical exposure on the Gulf War veterans cannot be properly ascertained.

3 The Logistic Model

With a binary dependent variable (health problem = 1, otherwise 0), the most widely used method to control for potential confounders is the logistic model with multiple covariates:

$$\log[(p(X)/(1 - p(X)] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p,$$

where $p(X) = \text{Prob}(Y = 1|X)$

The assumption behind this model is that logit $p(X)$ is a linear function of a set of covariates $X_1, X_2, \ldots X_p$. The parameters of the linear function are then estimated by maximum likelihood method. In our analysis, the logit of probability of developing specific illness is modeled as a linear function of Gulf War dummy and other explanatory variables like age, education, race, and others. Two likely factors proposed by various researchers—chemical exposure and PTSD—are also included in the explanatory variables.

In order to better capture the age effect in our data, we added an interaction variable—age times Gulf War veterans (age*Gulf)—in our regression. The other cross-variable used on the right-hand side of logistic regression is Gulf War dummy times chemical exposure (Gulf*chem) to capture the effect of the specific types of chemicals that the Gulf veterans were exposed as compared with other veterans. Relatively more number of Gulf veterans report having exposed to chemicals than other veterans (34.10 vs. 21.67%). The logistic regression results are shown in Tables 5a, 6a, 7a, 8a, 9a, and 10a.

We evaluate six typical Gulf War illnesses: chronic pain, gastrointestinal, depression, fatigue, concentration, and ENT problems. Despite specifying age and other potential confounders on the explanatory side of the model, logistic model estimates provide almost the same conclusion as the preliminary analysis of the data. Gulf War veterans show lower possibility to develop four of the six typical symptoms: chronic pain (OR = 0.283), gastrointestinal problem (OR = 0.291), depression problem (OR = 0.633), and concentration problem (OR = 0.474). For fatigue and ENT, the odds ratios were not statistically different from one. One interesting finding is that the

Table 5 **a** Logistic model for chronic pain problem. **b** Logistic model and GAM for chronic pain problem (age: 28–60)

a

	Odds ratio	*p*-value
Male	0.594	<0.0001
Gulf	0.283	<0.0001
Chemi_exposure	1.952	<0.0001
White	0.785	<0.0001
Edu $\geq$ HS	0.817	0.0002
Income $\geq$ 10 K	0.614	<0.0001
PTSD	4.905	<0.0001
Gulf*chemi	1.055	0.6327
Age*Gulf	1.035	<0.0001
Age	1.003	0.0231

b

	Logistic odds ratio	Logistic *p*-value	GAM odds ratio	GAM *p*-value
Male	0.576	<0.0001	0.589	<0.0001
Gulf	1.276	0.0036	1.278	0.0033
Chemi_exposure	1.778	<0.0001	1.843	<0.0001
White	0.799	<0.0001	0.805	0.0002
Edu $\geq$ HS	0.747	0.0040	0.748	0.0042
Income $\geq$ 10 K	0.483	<0.0001	0.486	<0.0001
PTSD	4.610	<0.0001	4.689	<0.0001
Gulf*chemi	1.211	0.1080	1.173	0.1815
Age	1.223	<0.0001	1.013	0.0004
Age square	0.998	<0.0001	–	–

$^*\chi^2$ test for nonparametric age effect in GAM is 61.1960 ($p < 0.0001$)

interaction variable age*Gulf has odds ratios significantly over one for chronic pain, and gastrointestinal and concentration problems suggesting that with comparable age as other veterans, the Gulf veterans have a higher probability of getting these health problems. Veterans who have relatively better socioeconomic status (income and education) and are white have less likelihood of suffering from all these symptoms, possibly because they can receive better treatment and management of their health problems. The variable Gulf*chemi is significant for depression and fatigue, but not for other ailments. The variable chemi_exposure is very significant for all symptoms, and the odds ratios are always in excess of 1.5. However, since the Gulf dummy comes out to be significantly negative for all symptoms except ENT (for which it was statistically insignificant), the question still remains if we are adequately controlling for the differential age distributions between these two groups.

Table 6 **a** Logistic model for gastrointestinal problem. **b** Logistic model and GAM for gastrointestinal problem (age: 28–60)

a

	Odds ratio	*p*-value
Male	0.723	<0.0001
Gulf	0.291	<0.0001
Chemi_exposure	1.638	<0.0001
White	0.983	0.7411
Edu $\geq$ HS	0.761	<0.0001
Income $\geq$ 10 K	0.803	0.0009
PTSD	3.369	<0.0001
Gulf*chemi	1.110	0.3901
Age*Gulf	1.036	<0.0001
Age	1.011	<0.0001

b

	Logistic odds ratio	Logistic *p*-value	GAM odds ratio	GAM *p*-value
Male	0.730	0.0007	0.742	0.0014
Gulf	1.392	0.0004	1.393	0.0003
Chemi_exposure	1.539	<0.0001	1.574	<0.0001
White	0.922	0.1992	0.926	0.2232
Edu $\geq$ HS	0.882	0.2576	0.881	0.2547
Income $\geq$ 10 K	0.656	<0.0001	0.659	<0.0001
PTSD	3.134	<0.0001	3.169	<0.0001
Gulf*chemi	1.217	0.1290	1.191	0.1771
Age	1.289	<0.0001	1.021	<0.0001
Age square	0.997	<0.0001	–	–

$^*\chi^2$ test for nonparametric age effect in GAM is 45.1556 ($p < 0.0001$)

4 The Generalized Additive Model (GAM)

4.1 *Limitations of Logistic Model for Controlling Age Effect*

The logistic model is limited in controlling age effects since it assumes a rather simple relationship between the transformations of the response variable and independent variables. If the independent variable is related to the response variable in a complicated nonlinear fashion, then the linearity assumption is not reasonable and the logistic model can yield misleading coefficients for all explanatory variables. Based on the generalized additive model (that we will explain shortly), Figs. 4, 5, 6, 7, 8, and 9 show plots of the partial effects of age on the possibility of developing the six typical Gulf War symptoms with a 95% confidence interval. Except for the ENT

Table 7 **a** Logistic model for depression problem. **b** Logistic model and GAM for depression problem (age: 28–60)

a

	Odds ratio	*p*-value
Male	0.867	0.0432
Gulf	0.633	0.0833
Chemi_exposure	1.466	<0.0001
White	0.662	<0.0001
Edu $\geq$ HS	0.693	<0.0001
Income $\geq$ 10 K	0.459	<0.0001
PTSD	8.127	<0.0001
Gulf*chemi	1.393	0.0043
Age*Gulf	1.002	0.6919
Age	0.997	0.0195

b

	Logistic odds ratio	Logistic *p*-value	GAM odds ratio	GAM *p*-value
Male	0.789	0.0048	0.786	0.0041
Gulf	0.713	<0.0001	0.708	<0.0001
Chemi_exposure	1.575	<0.0001	1.568	<0.0001
White	0.656	<0.0001	0.656	<0.0001
Edu $\geq$ HS	0.723	0.0011	0.723	0.0011
Income $\geq$ 10 K	0.344	<0.0001	0.344	<0.0001
PTSD	8.510	<0.0001	8.480	<0.0001
Gulf*chemi	1.262	0.0571	1.273	0.0476
Age	1.119	0.0005	1.000	0.9389
Age square	0.999	0.0005	–	–

$*\chi^2$ test for nonparametric age effect in GAM is 28.1370 ($p < 0.0001$)

case, none of the effects looks linear. This highly nonlinear relationship between age and response variable is the key to understand why our logistic model provides little control for the age heterogeneity in the two samples (see Fig. 10 for the presence of having more than three symptoms).

4.2 Methodology for Generalized Additive Model

One approach to model an unknown nonlinear relation between dependent and independent variables is nonparametric regression. Nonparametric regression relaxes the usual assumption of linearity and enables one to explore the data more flexibly, uncovering structure in the data that might otherwise be missed. However, many forms of

Table 8 **a** Logistic model for fatigue problem. **b** Logistic model and GAM for fatigue problem (age: 28–60)

a

	Odds ratio	*p*-value
Male	0.689	<0.0001
Gulf	0.826	0.3966
Chemi_exposure	1.644	<0.0001
White	0.830	<0.0001
Edu ≥ HS	0.735	<0.0001
Income ≥ 10 K	0.464	<0.0001
PTSD	4.645	<0.0001
Gulf*chemi	1.267	0.0206
Age*Gulf	1.003	0.6247
Age	1.015	<0.0001

b

	Logistic odds ratio	Logistic *p*-value	GAM odds ratio	GAM *p*-value
Male	0.702	<0.0001	0.702	<0.0001
Gulf	0.940	0.3828	0.937	0.3576
Chemi_exposure	1.745	<0.0001	1.748	<0.0001
White	0.796	<0.0001	0.796	<0.0001
Edu ≥ HS	0.656	<0.0001	0.657	<0.0001
Income ≥ 10 K	0.395	<0.0001	0.395	<0.0001
PTSD	5.723	<0.0001	5.721	<0.0001
Gulf*chemi	1.175	0.1389	1.174	0.1393
Age	1.151	<0.0001	1.013	<0.0001
Age square	0.999	<0.0001	–	–

$^*\chi^2$ test for nonparametric age effect in GAM is 28.1370 ($p < 0.0001$)

nonparametric regression do not perform well when the number of independent variables in the model is large. The sparseness of data in this setting causes the variances of the estimates to be unacceptably large. The problem of rapidly increasing variance for increasing dimensionality is referred to as the "curse of dimensionality." Interpretability is another problem with nonparametric regression based on kernel and smoothing spline estimates. The information these estimates contain about the relationship between the dependent and the independent variables is often difficult to comprehend.

To overcome these difficulties, Stone (1985) proposed additive models. These models estimate an additive approximation to the multivariate regression function. The benefits of an additive approximation are at least twofold. First, since each of the individual additive terms is estimated using a univariate smoother, the curse of dimensionality is avoided, at the cost of not being able to approximate universally.

Table 9 **a** Logistic model for concentration problem. **b** Logistic model and GAM for concentration problem (age: 28–60)

a

	Odds ratio	p-value
Male	0.859	0.0803
Gulf	0.474	0.0250
Chemi_exposure	1.789	<0.0001
White	0.625	<0.0001
Edu $\geq$ HS	0.565	<0.0001
Income $\geq$ 10 K	0.435	<0.0001
PTSD	10.244	<0.0001
Gulf*chemi	1.034	0.8152
Age*Gulf	1.013	0.0935
Age	1.002	0.1916

b

	Logistic odds ratio	Logistic p-value	GAM odds ratio	GAM p-value
Male	0.784	0.0176	0.783	0.0170
Gulf	0.870	0.1993	0.861	0.1654
Chemi_exposure	1.796	<0.0001	1.807	<0.0001
White	0.637	<0.0001	0.638	<0.0001
Edu $\geq$ HS	0.619	<0.0001	0.620	<0.0001
Income $\geq$ 10 K	0.354	<0.0001	0.355	<0.0001
PTSD	11.086	<0.0001	11.092	<0.0001
Gulf*chemi	0.985	0.9210	0.989	0.9398
Age	1.225	<0.0001	1.013	0.0044
Age square	0.998	<0.0001	–	–

$^*\chi^2$ test for nonparametric age effect in GAM is 34.9559 ($p < 0.0001$)

Second, estimates of the individual terms explain how the dependent variable changes with the corresponding independent variables.

To extend the additive model to a wide range of distribution families, Hastie and Tibshirani (1990) proposed generalized additive models. These models enable the mean of the dependent variable to depend on an additive predictor through a nonlinear link function. The models permit the response probability distribution to be any member of the exponential family of distributions. Many widely used statistical models belong to this general class; they include additive models for Gaussian data, nonparametric logistic models for binary data, and nonparametric log-linear models for Poisson data.

For response random variable Y and a set of predictors $X_1, X_2, \ldots, X_p$, instead of defining

Table 10 **a** Logistic model for ENT problem. **b** Logistic model and GAM for ENT problem (age: 28–60)

a

	Odds ratio	p-value
Male	0.557	<0.0001
Gulf	1.290	0.3519
Chemi_exposure	1.744	<0.0001
White	1.074	0.1719
Edu $\geq$ HS	0.941	0.3213
Income $\geq$ 10 K	0.922	0.2495
PTSD	2.286	<0.0001
Gulf*chemi	1.036	0.7645
Age*Gulf	1.003	0.6126
Age	1.009	<0.0001

b

	Logistic odds ratio	Logistic p-value	GAM odds ratio	GAM p-value
Male	0.519	<0.0001	0.524	<0.0001
Gulf	1.400	0.0002	1.402	0.0002
Chemi_exposure	1.648	<0.0001	1.667	<0.0001
White	1.090	0.1884	1.092	0.1769
Edu $\geq$ HS	0.976	0.8342	0.975	0.8319
Income $\geq$ 10 K	0.994	0.9551	0.996	0.9688
PTSD	2.152	<0.0001	2.168	<0.0001
Gulf*chemi	1.123	0.3574	1.109	0.4092
Age	1.009	0.0178	1.009	0.0213
Age square				

*χ^2 test for nonparametric age effect in GAM is 2.9308 ($p = 0.4024$)

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \text{Error},$$

the additive model generalizes the linear model by modeling the conditional expectation as

$$Y = S_1(X_1) + S_2(X_2) \cdots + S_p(X_p) + \text{Error}$$

where $s_i(X)$, $i = 1, 2, \ldots, p$ are smooth functions. In order to be estimable, the smooth functions s_i have to satisfy the standardized conditions, viz. $Es_j\ (X_j) = 0$.

Since our analysis includes potentially a large number of explanatory variables, we specified a semi-parametric generalized additive model with logit link:

$$g(p) = \log[p(X)/(1 - p(X)] = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + S(\text{age})$$

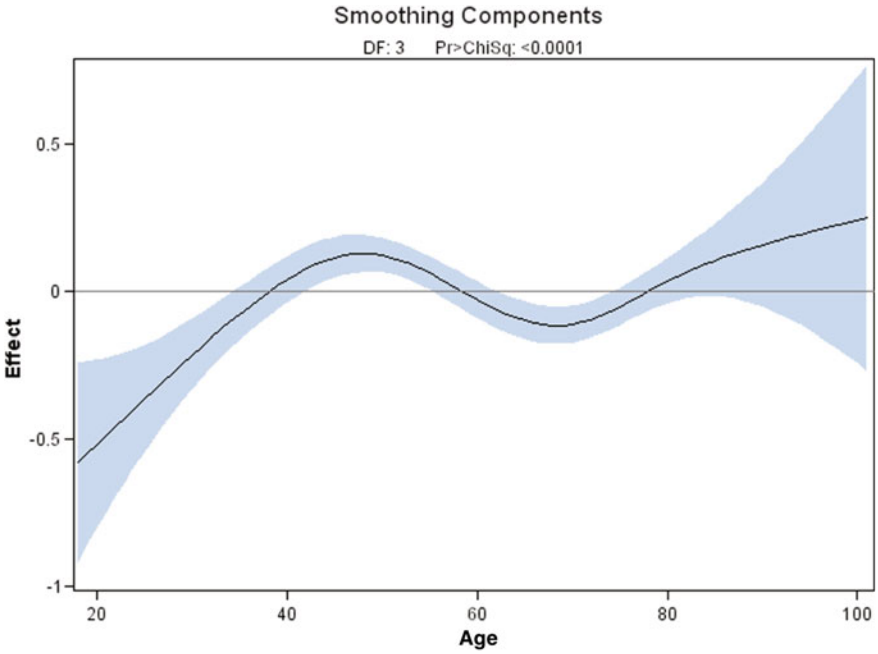


Fig. 4 Partial effect of age on depression problem

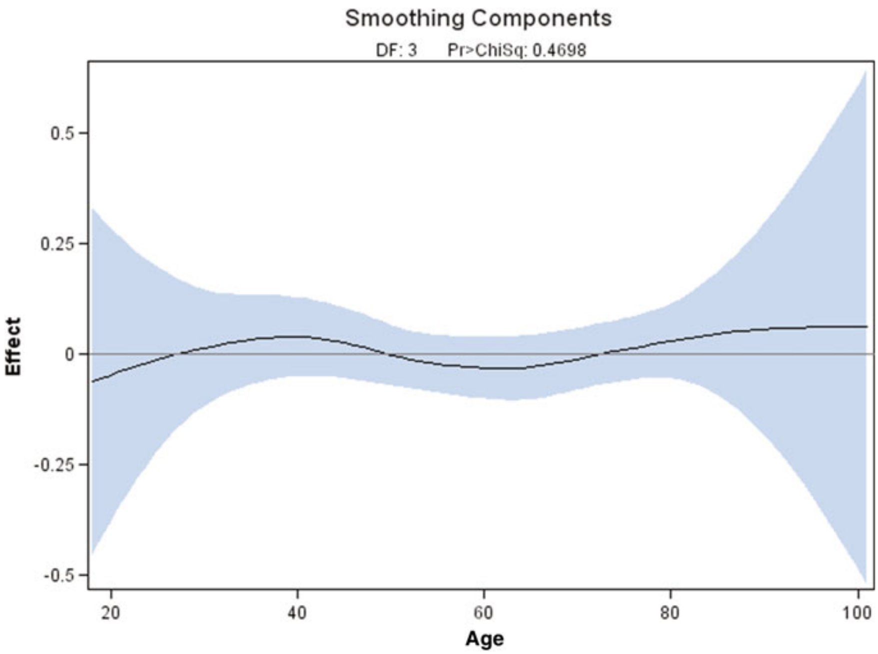


Fig. 5 Partial effect of age on ENT problem

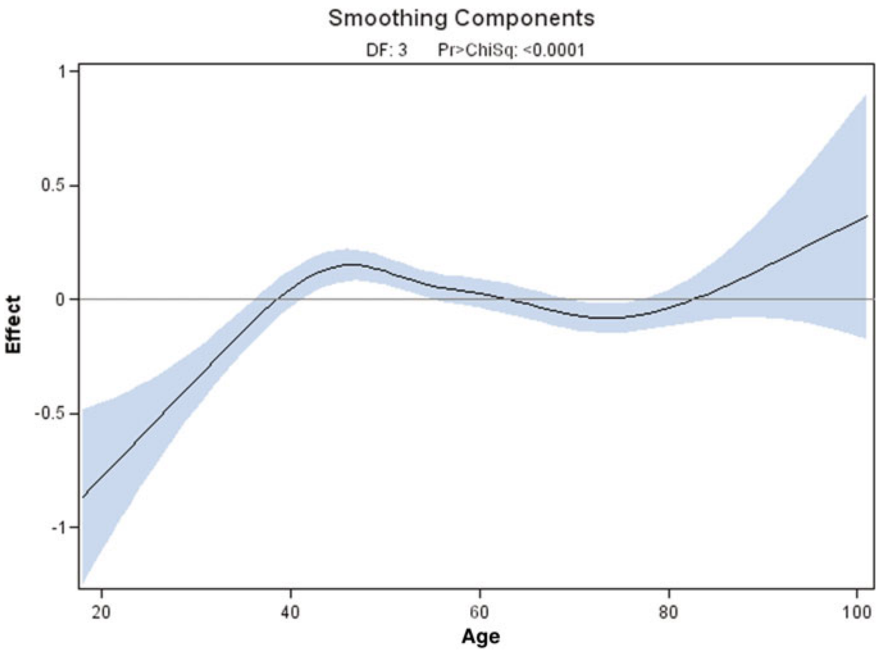


Fig. 6 Partial effect of age on chronic pain problem

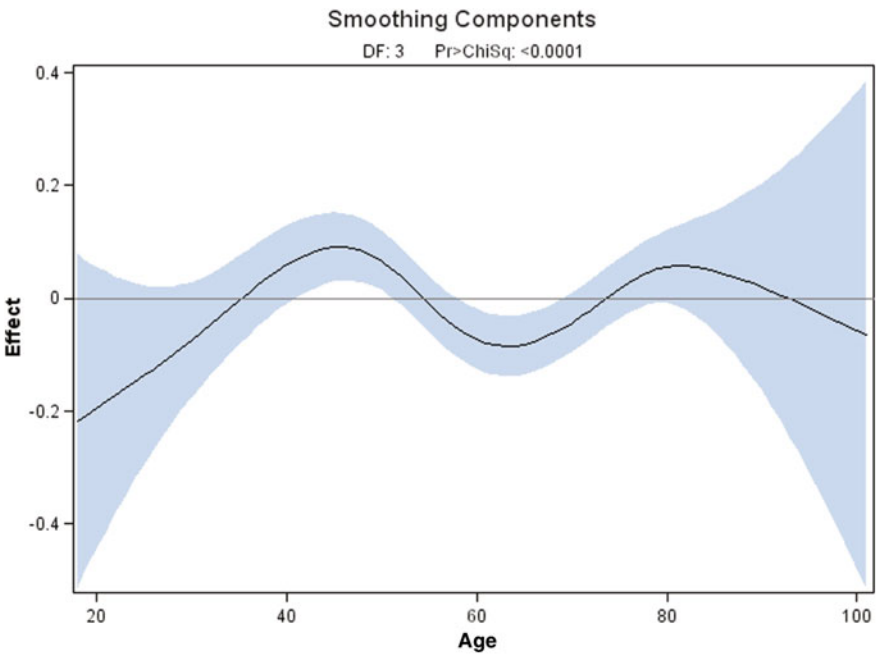


Fig. 7 Partial effect of age on fatigue problem

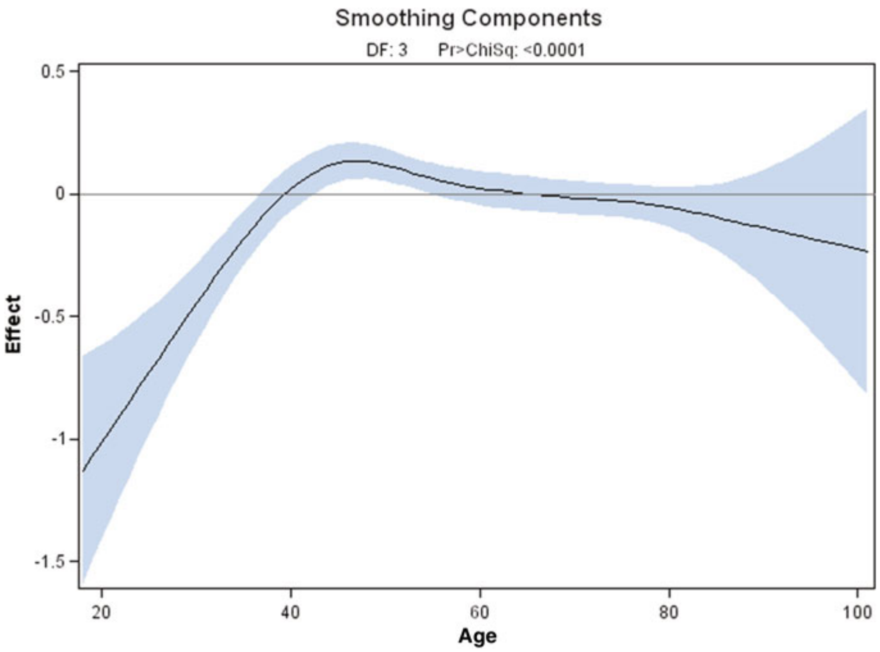


Fig. 8 Partial effect of age on gastrointestinal problem

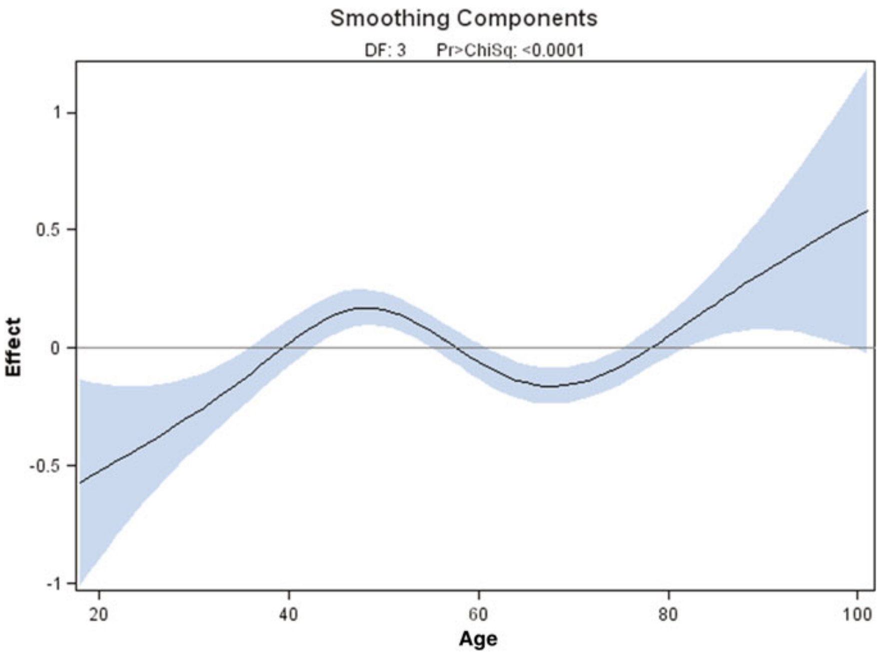


Fig. 9 Partial effect of age on concentration problem

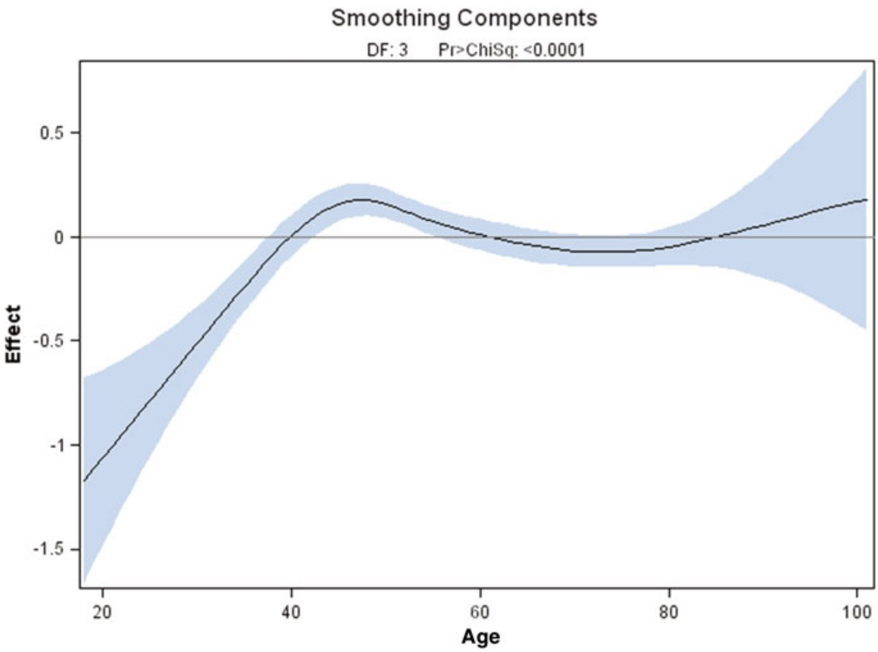


Fig. 10 Partial effect of age for having more than three symptoms

where X_1 to X_{p-1} stand for other independent variables such as gender, education and chemical exposure. Age is the main variable of interest we try to control with semi-parametric specification S. There are many ways to estimate generalized additive models. We estimate the parameters with the local scoring algorithm using a cubic spline smoother, as suggested by Hastie and Tibshirani (1986).

Since age of a veteran is a very important characteristic that determines health with a complicated nonlinear link, we restricted our sample to all veterans who are between ages 28 and 60. Figures 2 and 3 indicate that between ages 28 and 60 the most of the Gulf War veterans match with non-Gulf veterans. Also, between ages 28 and 60, the nonlinear relation between age and the incidence of the health problems can reasonably be approximated by a quadratic function; see Figs. 4, 5, 6, 7, 8, and 9. The confidence interval for the fitted line beyond age 60 is also very wide, implying less precise approximation. Therefore, we have good reasons to restrict our sample to all veterans who are between 28 and 60 years old. Another reason for restricting veterans in our sample to be older than 28 is that a Gulf War veteran cannot possibly be less than 28 years old when the survey was taken.

4.3 Results from GAM Estimation

The last two columns of results from Tables 5b, 6b, 7b, 8b, 9b, and 10b provide regression results from the GAM, with the sample restricted to ages 28–60. The first two columns show the results from logistic regression with age, age^2, and other covariates. Not surprisingly, logistic regression and GAM provide almost the same results due to the quadratic age specification in the former, as diagnosed by GAM. This way, our GAM estimates can be given a very straightforward interpretation; see Hastie and Tibshirani (1990). The nonparametric age specification is significant in five of the six typical Gulf War symptoms, with p-value less than 0.05. As expected from the corresponding graph, ENT is the only symptom with insignificant nonlinear age effect. For this problem, logistic model without age square specification provides very similar results with that from GAM. This regression results further confirm the value of GAM in controlling for the age confounder.

Most prominently and consistently, chemical exposure remains the chief driver for all these medical problems, even after we control the effect of stress factor PTSD. Especially for the Gulf veterans who have been exposed to the toxic exposure (Gulf*chemi_exposure), the odds ratios are always significantly above 1 for five of the six symptoms. Although PTSD is also a significant contributor to all the problems we evaluated, it cannot explain the incidence of the individual medical problems we are analyzing. Studies of veterans from different wars have consistently found that patients with PTSD generally have more symptomatic complaints than those without PTSD. Also, only a small fraction of veterans (6.79%) report PTSD in our sample. The reason is that the war was short, requiring only four days of ground combat to achieve a decisive victory. Casualty rates were very low, and the vast majority of veterans was never in combat areas and did not witness any deaths during deployment. VA have also reported that less than four percent of veterans examined in its Gulf War registry have either a primary or a secondary diagnosis of PTSD. Similarly, a RAND report commissioned by the Department of Defense to review the scientific evidence concerning stress and Gulf War illnesses concluded that overall rates of PTSD are low in Gulf War veterans. As noted, the chemical exposure remains significant even after we control for the presence of PTSD problem in both logistic and GAM specifications. This consensus between the two model specifications demonstrates that at least, deployment stress is not enough to explain the relative health status of Gulf War veterans.

Remarkably, for chronic pain and gastrointestinal problems, the OR for the Gulf dummy changes from significantly less than one to significantly greater than one in GAM estimation implying that once age is properly accounted for, Gulf veterans indeed have more incidence of these problems. For depression and fatigue problems, OR for Gulf dummy increased significantly, but did not exceed the value of one. However, for these ailments, the Gulf*chemi dummy is significantly more than one. This variable together with chemi_exposure implies that chemical exposure in the Gulf War increased the odds of being inflicted with these ailments. For concentration problem, the Gulf dummy in GAM estimation again increased in OR, but did not exceed one. The Gulf*chemi variable is also not statistically significant. For this medical problem, the only evidence of the Gulf War comes via the chemi_exposure variable. For ENT, GAM and the logistic model with linear and quadratic age term estimations, as expected, are very similar because our nonparametric estimation identified the age effect as close to being linear. For this medical problem (i.e., ENT), Gulf dummy is always positive and highly significant. Thus, the preponderance of the evidence is that once the effects of the cofounders, most notably that of age, are appropriately taken care off, the Gulf War veterans, due to their exposure to chemicals, have significantly higher odds in developing these chronic medical problems that did not fade away even 10 years after the exposure.

However, it is the occurrence of multiple symptoms that distinguishes Gulf War veterans from other veterans. Table 11 indicates that Gulf War veterans are significantly more likely to have more than three of the six typical symptoms all simultaneously; this is especially true for Gulf War veterans who had toxic exposure. This finding supports the conclusion of the recent report released by the federal committee: "A similar proportion of Gulf and non-Gulf veterans reported a relatively low level of symptoms—that is, symptoms in just one or two of the six defined symptom groups." However, more Gulf War veterans "reported having moderate-to-severe symptoms in three or more of the symptom domains."

Generally speaking, our results indicate that the pattern of symptoms in Gulf War veterans is unique in terms of the occurrence of multiple symptoms. To be a Gulf War veteran may not result in worse health with respect to only one medical condition. The Gulf War veterans who were exposed to chemical exposure during the war are most likely to be troubled by a multitude of symptoms. Therefore, after reevaluating the data, we reach the conclusion that many Gulf War veterans are ill in a special way in that they are inflicted by multiple problems, and the likely causal link to this syndrome is the toxic chemical exposure.

Table 11 Logistic and GAM analysis for having more than three typical symptoms

	Logistic odds ratio	Logistic p-value	GAM odds ratio (28–60)	GAM p-value (28–60)
Male	0.624	<0.0001	0.614	<0.0001
Gulf	0.235	<0.0001	1.202	0.0819
Chemi_exposure	2.131	<0.0001	2.097	<0.0001
White	0.750	<0.0001	0.755	<0.0001
Edu $\geq$ HS	0.691	<0.0001	0.615	<0.0001
Income $\geq$ 10 K	0.566	<0.0001	0.452	<0.0001
PTSD	7.954	<0.0001	8.232	<0.0001
Gulf*chemi	1.210	0.1505	1.269	0.0924
Age*Gulf	1.037	<0.0001	–	–
Age	1.008	<0.0001	1.020	<0.0001

*χ^2 test for nonparametric age effect in GAM is 69.5561 ($p < 0.0001$)

This conclusion is further confirmed when we restrict our analysis to a sample of only Gulf War veterans. We compare Gulf War veterans who had experienced toxic exposure during the war with those who did not report chemical exposure. Table 12 indicates that veterans with chemical exposure suffer from worse self-reported health status, more service-connected disability, higher outpatient and inpatient hospitalizations, and abnormally higher rate of the typical Gulf War symptoms. Moreover, Table 13 and Fig. 11 show that a similar proportion of chemically exposed and non-exposed Gulf War veterans reported a relatively low number of symptoms—for instance, only one medical problem out of the six defined symptom groups. However, over 50% of exposure veterans reported symptoms in two or more of the symptom domains, compared to less than 30% of the comparison group. Thus, the Gulf War veterans experiencing chemical exposure have worse health status, in terms of frequency, severity, duration, and the occurrence of multiple-symptom health problems. In Table 14, we report an ordered probit regression of self-reported health status on

Table 12 General health status of Gulf War veterans

	Chemical exposure	No chemical exposure	*p*-value
Self-evaluated health status			
Excellent	9.44	18.94	<0.0001
Very good	20.05	32.53	
Good	32.99	30.39	
Fair	26.52	14.06	
Poor	10.61	3.88	
Service-connected disability	60.03	39.89	<0.0001
Number of outpatient visits	8.49	4.92	<0.0001
Usage of emergency service	32.99	26.57	0.0069
Usage of inpatient service	8.15	6.36	0.2885
ENT problem	28.07	16.67	<0.0001
Chronic pain	37.77	19.75	<0.0001
Fatigue	52.78	32.06	<0.0001
Depression	35.06	18.21	<0.0001
Concentration problem	21.99	10.11	<0.0001
Gastrointestinal problem	27.81	14.66	<0.0001

Table 13 Proportion of US Gulf War veterans reporting typical symptoms

Number of symptoms	Chemical exposure (%)	No chemical exposure (%)	p-value
0	25.10	45.38	<0.0001
1	20.44	23.09	
2	18.37	15.86	
3	13.84	8.63	
4	9.96	4.75	
5	8.80	1.94	
6	3.49	0.33	

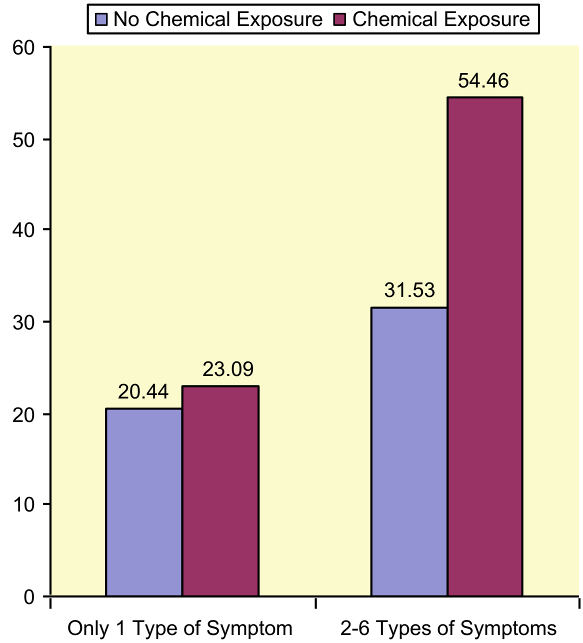


Fig. 11 Proportion of US Gulf War veterans reporting typical symptoms

Table 14 Ordered probit model for self-reported health status (Gulf veterans only)

	Marginal effect for health = poor	Marginal effect for health = fair	Marginal effect for health = good	Marginal effect for health = very good	p-value
Chemi_exposure	0.0535	0.1053	0.0467	−0.0885	0.0000
White	−0.0344	−0.0676	−0.0300	0.0568	0.0000
Male	−0.0124	−0.0243	−0.0108	0.0205	0.0458
Age	0.0020	0.0039	0.0017	−0.0033	0.0000
Edu ≥ HS	−0.0264	−0.0520	−0.0230	0.0437	0.1400
Income ≥ 10 K	−0.0627	−0.1234	−0.0547	0.1038	0.0000

chemical exposure with controls for race, gender, education, age, and income. Based on this regression, Fig. 12 depicts a significant left shift of the distribution of the predicted probabilities of falling into different self-reported health groups for these two Gulf veteran groups. This graph shows that for chemically exposed Gulf veterans, the probabilities of reporting poor and fair health are higher, and those for very good and excellent health, they are lower compared to Gulf veterans who were not exposed to the chemicals. In Table 15, we report a logistic regression of having three or more Gulf illness syndromes as a function of chemical exposure, gender, race, education, income, and age. Again, the effect of chemi_exposure is found to be very significant with an odds ratio in excess of 3. Based on this regression, Fig. 13 reveals that a gulf veteran's probability for having three or more health symptoms increases substantially if they were exposed to chemicals. The vertical difference between the

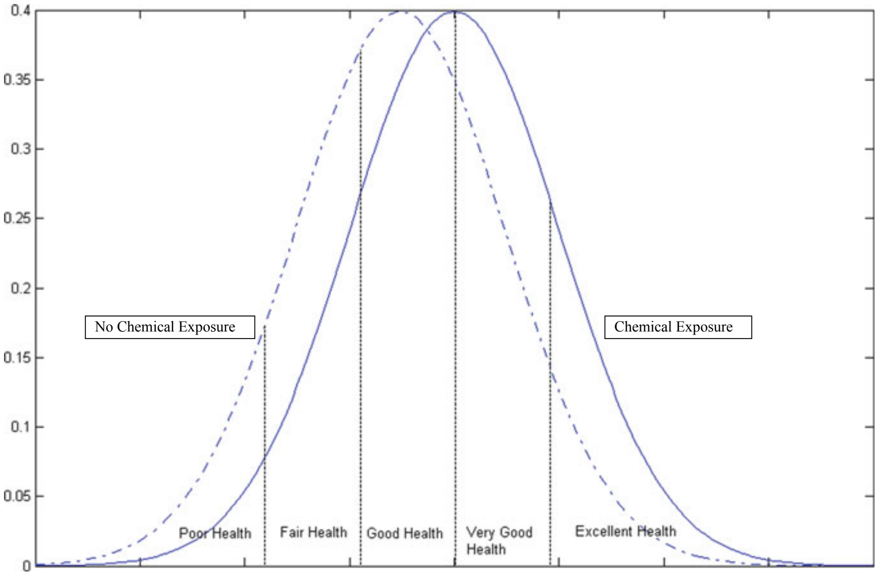


Fig. 12 Effects of chemical exposure on predicted probabilities of self-reported health for Gulf veterans

Table 15 Logistic regression for having more than three problems (Gulf veterans only)

	Odds ratio	*p*-value
Male	0.543	<0.0001
Chemi_exposure	3.066	<0.0001
White	0.705	0.0022
Edu $\geq$ HS	0.499	0.0518
Income $\geq$ 10 K	0.354	<0.0001
Age	1.034	<0.0001

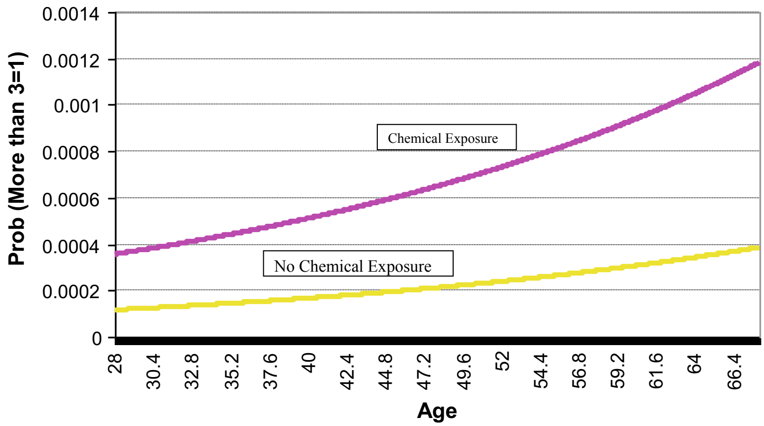


Fig. 13 Effect of chemical exposure on the probability of having more than three typical Gulf War syndromes

two lines depicts the marginal effect of chemical exposure as a function of age; we find that this difference also increases with age.

These regressions clearly show that our analysis of the NSV 2001 data overwhelmingly supports the conclusion of the federal panel that chemical exposure during the Gulf War is the principal driver of the much-documented Gulf War illnesses.

5 Conclusions

About one-fourth of the 697,000 US servicemen and women in the first Gulf War have shown symptoms related to Gulf War syndrome. Despite recent research breakthroughs on the effects of clinical exposure on Gulf War illnesses, the etiologic basis and clinical significance of Gulf War veterans still remain unclear. Due to the lack of objective long-term follow-up data, there has been no conclusive answer to the fundamental question related to the pattern and nature of these mysterious and often debilitating ailments. In their 2004 report, the Research Advisory Committee on Gulf War Veterans' Illnesses, appointed by Congress and the VA, recommended that important social priority should be given to expansive studies capable of reevaluating the possible reasons for the Gulf War illnesses. The 2001 National Survey of Veterans, conducted 10 years after the Gulf War, provides invaluable information for longtime chronic health effect of the war and deserves special investigation. Our preliminary data analysis, contrary to previous evidence on the relative health for Gulf War veterans, indicates better health for Gulf War veterans compared to other veterans. Using this data and appropriate statistical methods, we assessed the health of Gulf War veterans, compared their health to veterans who did not serve in Persian

Gulf, and reevaluated the probable link between chemical exposure and Gulf War illnesses.

Since NSV 2001 is a comprehensive data set where Gulf veterans represent only a small part of it, we showed that age distribution of veterans in different subpopulations is an extremely important confounding factor. The widely used logistic model contributes little in controlling age effect due to its simplifying assumption. We show that the marginal effect of age in our sample is highly nonlinear requiring a more flexible specification of the effect of age on health. We used the semi-parametric generalized additive model and estimated it using local scoring algorithms. The results are striking. After successfully controlling the effect of age with GAM specification, we could unravel the true underlying causes of the special health problems of Gulf War veterans. Gulf War deployment, by itself, is not the root cause of the myriad of health problems faced by some of the Gulf War veterans. Only the Gulf War veterans who were exposed to toxic chemicals are found to be strongly associated with an abnormally high risk of the typical Gulf War illnesses. Moreover, an important feature of the health problems of Gulf War veterans is the presence of symptoms for a number of ailments.

We also reevaluated the two possible contributors to the ill health of Gulf War veterans: deployment stress and chemical exposure. Since the war was short and the overall rate of PTSD in Gulf War veterans is low, PTSD is less likely to be the main factor causing health problems in Gulf War veterans. Instead, both logistic regression and GAM specifications indicate a significant contribution of chemical exposure even after we control the deployment stress factor. This finding suggests a strong association between the higher prevalence of Gulf syndromes and the toxic exposure specific to the Gulf War.

Our study provides important evidence on the complexity in identifying and recognizing health problems of veterans after deployment from a specific war. Methodologically, we addressed the potential confounding effect of different age distributions in study subsamples with a semi-parametric statistical approach that should help future empirical research.

Acknowledgements This research was supported by the National Center on Minority Health and Health Disparities and National Institutes of Health (grant number P20MD003373). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center on Minority Health and Health Disparities or the National Institutes of Health. We thank John Jones, Terrence Kinal, Zulkarnain Pulungan, and an anonymous referee for many comments and suggestions.

References

Australian Commonwealth Department of Veterans' Affairs. Australian Gulf War veterans' health study report. *ACDVA*, 2003.

Bureau of Labor Statistics (BLS). U.S. Department of Labor. Employment situation of veterans—2009. News Release, March 12, 1010. USDL-10-0285.

Chemicals sickened Gulf War veterans, latest study finds, *New York Times* (2004, October 15).
Cherry, N., Creed, F., & Silman, A., et al. (2001). Health and exposures of United Kingdom Gulf War veterans. Part II: The relation of health to exposure. *Occupational and Environment Medicine, 58*, 299–306.
Gray, G. C., Smith, T. C., Kang, H. K., et al. (2000). Are Gulf War veterans suffering war-related illnesses? Federal and civilian hospitalizations examined, June 1991 to December 1994. *American Journal of Epidemiology, 151,* 63–71.
Hastie, T. J., & Tibshirani, R. J. (1986). Generalized additive models (with discussion). *Statistical Science, 1,* 297–318.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.
Hotopf, M., David, A. S., Hull, L., Ismail, K., Unwin, C., & Wessely, S. (2003a). The health effects of peacekeeping (Bosnia, 1992–1996): A cross-sectional study—Comparison with nondeployed military personnel. *Military Medicine, 168,* 408–413.
Hotopf, M., David, A. S., Hull, L., Nikalaou, V., Unwin, C., & Wessely, S. (2003b). Gulf war illness—Better, worse, or just the same? A cohort study. *British Medical Journal, 327,* 1370–1374.
Hyams, K. C., Signall, F. S., & Roswell, R. (1996). War syndromes and their evaluation: from the U.S. Civil War to the Persian Gulf War. *Annals of Internal Medicine, 125*, 398–405.
Iowa Persian Gulf Study Group. (1997). Self-reported illness and health status among Gulf War veterans—A population-based study. *Journal of the American Medical Association, 277,* 238–245.
Ishoy, T., Suadicani, P., Guldager, B., Appleyard, M., Hein, H. O., & Gyntelberg, F. (1999). State of health after deployment in the Persian Gulf: The Danish Gulf War study. *Danish Medical Bulletin, 46,* 416–419.
Kang, H. K., Mahan, C. M., & Lee, K. Y. (2000). Illnesses among United States veterans of the Gulf War: A population-based survey of 30,000 veterans. *Journal of Occupational and Environmental Medicine, 42,* 491–501.
Kang, H. K., Mahan, C. M., Lee, K. Y., Murphy, F. M., Simmens, S. J., Young, H. A., & Levine P. H. (2002). Evidence for a deployment-related Gulf War syndrome by factor analysis. *Achieves of Environmental Health, 57*(1), 61–68.
Landrigan, P. J. (1997). Illness in Gulf War veterans: Causes and consequences. *Journal of the American Medical Association, 277,* 259–261.
Lange, G., Tiersky, L., DeLuca, J., et al. (1999). Psychiatric diagnoses in Gulf War veterans with fatiguing illness. *Psychiatry Research, 89*, 399–448.
Murphy, F. M. (1999). Gulf War syndrome: There may be no specific syndrome, but troops suffer after most wars. (Editorial). *British Medical Journal, 318*, 274–275.
Presidential Advisory Committee on Gulf War Veterans' Illnesses: Final report. Washington, DC: US Government Printing Office, 1996.
SAS Institute Inc. (2000). *SAS/STAT user's guide, version 9*. Cary, NC: SAS Institute Inc.
Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics, 13,* 689–705.
Scientific Progress in Understanding Gulf War veterans' Illnesses: Report and Recommendations, Research Advisory Committee on Gulf War Veterans' Illness, September, 2004.
Steele, L. (2000). Prevalence and patterns of Gulf War illness in Kansas veterans: Association of symptoms with characteristics of person, place, and time of military service. *American Journal of Epidemiology, 152,* 991–1001.
VA advisers link Gulf War illnesses to neurotoxins. *Science, 306*, 26–27 (2004).
Verbrugge, L. N., & Ascione, F. J. (1987). Exploring the iceberg: Common symptoms and how people care for them. *Medical Care, 25,* 539–563.
Wolfe, J., Proctor, S. P., Erickson, D. J., et al. (1999). Relationship of psychiatric status to Gulf War veterans' health problems. *Psychosomatic Medicine, 61,* 532–540.

Foetal Starvation, Economic Adversity and Health a Difference-in-Difference Approach

Zakir Husain, Diganta Mukherjee, Mousumi Dutta and Susmita Mukhopadhyay

Abstract The foetal origin hypothesis argues that starvation during the foetal stage increases the probability of the onset of non-communicable diseases in midlife. The theory, however, fails to identify the mechanisms underlying the outcome. Nor does it succeed in distinguishing between study and control groups. The predictive adaptive response theory addresses the former deficiency by hypothesising that nutrition supply at the foetal stage signals the future nutrition supply and leads to adaptation of the foetus to the future expected environment. Mismatch between expected and actual environment will increase the likelihood of non-communicable diseases. The study examines the long-term impact of foetal starvation on anthropometric indicators among residents in the Sundarban region in India. We hypothesise that nutrition deficiency in the foetal stage signals the future expected environment to the foetus. This leads to the growth of a thrifty phenotype ensuring optimal performance of the offspring in a nutrition-deficit environment. A primary survey, undertaken between May 2014 and April 2015, was used to collect the data. In the first stage of the survey, Muslim women who had offspring in the period 1993–1997 were listed. In the second phase, anthropometric measurements of their offspring were taken. Respondents are placed within the study group if their mothers had kept the Ramadan fast (provided it coincided with conception); remaining respondents were defined as controls. Differences in mean anthropometric outcomes are tested using Monte Carlo simulations. A Difference-in-difference method is also applied. Respondents exposed to foetal starvation had better outcomes than those in the control group if they remained in poverty or their economic status deteriorated. Results were reversed

Z. Husain (✉) · M. Dutta
Economics Department, Presidency University, Kolkata, India
e-mail: dzhusain@gmail.com

M. Dutta
e-mail: dmousumi1970@gmail.com

D. Mukherjee
Sampling & Official Statistics Unit, Indian Statistical Institute, Kolkata, India
e-mail: digantam@hotmail.com

S. Mukhopadhyay
Anthropology Unit, Indian Statistical Institute, Kolkata, India
e-mail: susmi.mukho@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_20

for children with a sustained high standard of living, or those whose economic conditions improve. Findings are interpreted to provide support for the predictive adaptive response theory. However, tests using larger samples are required before arriving at a firm conclusion.

Keywords Foetal origin hypothesis · Predictive adaptive response · Anthropometric outcomes · Monte Carlo simulations · Difference-in-difference · India

1 Background

In the late 1950s, epidemiologists believed that the placenta was a "perfect filter, protecting the foetus from harmful substances in the mother's body and letting through helpful ones" (Landro 2010). This belief was shaken by the Thalidomide episode in the 1950s and 1960s[1] and by the research of David Barker and his associates. The latter proposed that the nine months in utero was one of the most critical periods in a person's life, shaping future health trajectories, endowments and capabilities. Based on the finding that several chronic diseases associated with ageing may be programmed before birth, Barker proposed that the old model of adult degenerative disease, based on an interaction of genes and an adverse environment in adult life, should be replaced by a new approach based on "programming by the environment in foetal and infant life" (Barker 1990, p. 1111). This is called the foetal origin hypothesis (FOH).

Barker viewed the effects of foetal under-nutrition on adult diseases as side effects of foetal adjustments instigated to boost survival of the foetus. The permanent changes in the structure and the functioning of the body resulting from foetal starvation becomes disadvantageous later in adult life. Hales and Barker later published the "thrifty phenotype" hypothesis, proposing that adjustment of foetal growth rate might set nutritional expectations and thereby condition the organism's response to nutrition later in life (Hales and Barker 2001). The idea was subsequently expanded to argue that the mother provides signals about the future environment to the foetus who adapts accordingly (Bateson 2001; Bateson et al. 2004; Gluckman et al. 2005a, b, 2008; Gluckman and Hanson 2007). Bateson, for instance, argues that "… the pregnant woman in poor nutritional condition may unwittingly signal to her unborn baby that it is about to enter a harsh world." (Bateson et al. 2004: 420). This leads to the baby being born with adaptations that help it to cope in food-deficit environments. This is referred to as predictive adaptive response (PAR).

One important adaptation to the nutritional shortage is that the foetus will use most of the scarce energy for the most vital organs (particularly the brain) and their metabolism so that they will be protected against lack of fuel (Godfrey and Barker

[1] Thalidomide was licensed in 1957 (except in the USA) and widely prescribed to pregnant women for morning sickness, until 1961 when it was identified as the cause of an epidemic of severe birth defects such as missing arms and legs (Haymann 1962; Smithells 1962).

2001). This "brain sparing" may slow down cell division in organs undergoing their critical growth period. Other foetal adaptations consist of increased maternal concentration levels of the stress hormone CRH (corticotropin-releasing hormones). This prepares the foetus for an expected preterm delivery—foetal growth is reduced and tissue maturation accelerated (Hobel and Culhane 2003).

However, the signal received at the foetal stage may also fail to correctly anticipate the future environment. If individuals with a thrifty phenotype find themselves growing up in conditions of relative affluence, there will be an environmental mismatch. Changes that are adaptive under poor conditions, such as storing food as fat and maintaining high blood glucose levels, now become maladaptive under good conditions (Bateson et al. 2004; Gluckman et al. 2005a). In such cases, the costs of such adaptations emerge in the form of increased risks of coronary heart disease and its biological risk factors, such as hypertension and type II diabetes mellitus (Seckl et al. 1997; Godfrey and Barker 2001; Seckl and Holmes 2007; Van Abeelen et al. 2012).

While there is a substantial body of the empirical literature on the FOH, most of the works suffer from an identification problem. The lengthy time gap between the nutrition shock/signal and the manifestation of the outcome of interest creates an identification problem. Short of time travelling, it is not easy to obtain reliable data for *both* these time periods. This creates an identification problem (Paneth and Susser 1995).[2]

To solve this problem, early works either studied animals which mature rapidly or relied on proxies of foetal starvation. Both methods are flawed. Wells points out that human beings have longer periods of growth before attaining maturity the random component in the "future" environment increases. It is doubtful, therefore, whether signals received at either the foetal or other stages produce "a developmental trajectory (that) … is genuinely predictive of the environment experienced as an adult" (Wells 2006: 424). The use of birth weights, a commonly used proxy for foetal starvation, has also been criticised. In developing countries, mothers are short in general and produce stunted or underweight babies. Focus on simply the weight or height at birth, however, ignores the *thin-fat* syndrome, as such babies may have low skeletal muscle mass but high central adiposity (Yajnik 2004).

In response, economists have attempted to model starvation in terms of a national or regional shock (generally famines) occurring over a single period and affecting a single cohort. All persons in the cohort are assumed to be exposed to the in utero shock, and their health outcomes compared to preceding or subsequent birth cohorts (who are *all* assumed to be sheltered from any nutritional deprivation). Using the nineteenth-century blight to French vineyards caused by the phylloxera insect as a shock, a study found that prenatal exposure reduced the height of army recruits aged 20 years by 0.5–0.9 cm (Banerjee et al. 2010). Studies on pandemic influenza in Brazil (Nelson 2010) and Britain (Kelly 2009) also reported long-term health

[2]Douglas Miller refers to this as the science fiction problem in analysing the foetal origins hypothesis—sans time travelling, we have to wait for a generation for the effects of a prenatal intervention of interest to be observed in adulthood (cited in Almond and Currie 2011: 157).

impacts on children born to women during the pandemic; an adverse labour outcome was reported for American (Almond 2006) and Danish (Schwandt 2017) children. Increased risk of diabetes and cardiovascular disease was observed for children born during the Dutch famine at the time of German occupation (Ravelli et al. 1998) and in the course of the Leningrad siege (Stanner et al. 1997), respectively.

Another method that has been used in recent studies is to use fasting during the lunar month of Ramadan as a proxy for foetal shock (Almond and Mazumder 2011; Van Ewijk 2011). A study found that Muslim children exposed to Ramadan in early pregnancy have significant health effects, with the likelihood of adult disability increasing by over 20% among Iraqis and Ugandan Muslims; substantially larger effects are observed for mental and learning disabilities (Almond and Mazumder 2011). The second study is based on the Indonesian Family Life Survey data. Using a similar intent to treat approach, the study found that Ramadan fast during pregnancy may increase risk of anaemia, coronary heart disease and type II diabetes in midlife, irrespective of the stage of pregnancy in which Ramadan took place (Van Ewijk 2011). Another study using the Indonesian Family Life Survey data reported a difference in height (Kunto and Mandemakers 2019). A recent meta-analysis of 22 studies using the Ramadan approach reported the need for further studies using a more carefully design approach before coming to any conclusion (Glazier et al. 2018).

This approach, however, suffers from the identification problem because all pregnant Muslim women need not fast. The second problem with using Ramadan as a proxy for foetal starvation is that the approach focuses on the signals received in one month, ignoring the signals received by the foetus during the remaining eight months of pregnancy. It is quite possible that the nutrition intake of the mother is high before and after Ramadan. In that case, the question is whether the signal received by the foetus during the month-long Ramadan is sufficiently strong enough to override the signal received during the remaining 8-month period becomes important.

Finally, such studies tend to view the onset of adulthood disease as a cost of the adaptation triggered by foetal starvation. Literature on FOH is marked by a failure to consider the causal pathways through which the foetal origin hypothesis operates—the mismatch of the thrifty phenotype to the adult environment. A close examination of FOH reveals that four key predictions relating to environmental mismatch emerge from the PAR theory:

(i) "For those born in good conditions, fitness is highest under good adult conditions and progressively decreases as the adult environment deteriorates.
(ii) In good adult conditions, those individuals that developed under good conditions have higher fitness than those that developed under poor conditions.
(iii) For those born in poor conditions, fitness is highest when the environment matches that of development, that is poor, and progressively decreases as the adult environment departs from this, i.e. as the conditions in the adult environment improve.
(iv) In poor adult conditions, those individuals that developed under poor conditions have higher fitness than those that developed under good conditions" (27: 1638).

Rather than examining these hypotheses linked to the causal pathways linked to the impact of foetal starvation, studies have generally assumed that early signals of nutrition deficient environment induce adaptations, irrespective of whether the signals correctly predict the future environment, costly in the long term. Attempts to test PAR have been rare, limited to animals and produced mixed evidence among rodents (McCay and Crowell 1934; Mccay et al. 1935; Taborsky 2006).

In this study, we examine whether signals about the future environment received in the form of nutritional supply at the foetus stage enable the respondent to utilise its development plasticity and adapt accordingly. The distinctive features of the study are as follows:

(1) The study uses data from a primary survey of Muslim youths aged 18–22 years, along with their mothers. We have collected information on the current and past standard of living, as well as on Ramadan fasting during conception. This provides better proxies of the signals received at the foetal stage.
(2) Unlike studies that focus only on verification of the FOH, this study also takes into consideration present conditions. This enables us to verify whether the thrifty phenotype is optimal designed only if the signals accurately predict the future environment. In brief, we are able to link the FOH with PAR and test the latter.
(3) We are able to assess the impact of a short-term nutritional shock (the Ramadan effect), overlapping the average nutritional supply during conception (proxied by the past standard of living).
(4) Finally, most of the studies have been undertaken in developed societies, where food shortage is generally not a problem.[3] The present study examines the FOH in a nutrition-deficit context, a situation more relevant in Afro-Asian developing countries.

2 Hypotheses and Method of Data Collection

2.1 Research Hypotheses

The basic premise of this study is that an organism's phenotype is influenced by its own environment. But past environmental effects on its parent(s) are also transmitted to their developing offspring. The nutritional intake received by the mother during her pregnancy is an example; it indicates the level of nutrition in the future expected environment. Accordingly, such signals interact with the genetic make-up to prepare the offspring to perform optimally in the expected environment. If there is a mismatch between actual and expected environment, however, performance may be adversely affected.

[3] There are only a few studies of FOH in India (Stein et al. 1996, 1997; Fall et al. 1998; Kumaran et al. 2000; Yajnik 2004). Instances of studies using the Ramadan approach in other developing societies are (Kunto and Mandemakers 2018; Savitri et al. 2018).

Three aspects of this hypothesis are tested in our study:

(1) The level of nutritional intake at the foetal stage signals a particular type of future environment. Depending upon the signal, the offspring is prepared for optimal survival in a nutrition-deficit or nutrition-rich environment. If the actual environment matches the expected environment, the outcome will be better than if there is a mismatch. This consists of testing the four propositions derived by Monaghan (2008), stated earlier.
(2) Ramadan fasting provides a short-term intensive nutritional shock lasting for a month. Given its short duration, can it override the signal provided by nutritional intake during the eight months comprising the rest of the foetal stage?
(3) Finally, we examine whether foetal starvation genetically programs the offspring to face latter life economic adversity more efficiently than an offspring anticipating a nutrition-plenty future.

2.2 Survey Site

The survey was undertaken in the chronically poor and underdeveloped Basanti block in the Sundarban areas. Basanti is an administrative division in Canning subdivision of South 24 Parganas district in the Indian state of West Bengal. The headquarters of this block is at Sonakhali. A brief profile of the district is given in Table 1.

About 65% of the population lies below the poverty line, while 37% do not have two square meals per day. Basanti ranks as the worst block in South 24 Parganas according to the Standard of Living Index calculated by the 2005 Rural Household Survey (Government of West Bengal 2008). The main occupations in Basanti are farming and pisciculture. About 42% of the workers are daily labourers, while 36% are cultivators.

Table 1 Profile of Basanti block

Indicators	2001	2011
Households	50,751	70,818
Population	278,543	336,717
Urban population	–	6625 (1.97%)
Percentage of literate population	56.98	68.32
Percentage of SC/ST population	47.84	41.49
Percentage of Muslim population[a]	41.18	44.87

[a]In 2011, the share of Muslim male and female population was 44.38 and 45.38%, respectively. Corresponding percentages for 2001 were 41.18 and 40.63%, respectively
Source http://www.censusindia.gov.in/2011census/population_enumeration.html

2.3 Sampling Strategy

The study was based on a primary survey of Muslim youth aged between 18 and 22 years (born between 1993 and 1997). This age group was selected keeping in mind two factors:

(a) Lowering the age below 18 years might have included respondents whose anthropometric development is not fully complete; and
(b) Increasing the age above 22 years would mean that time gap between birth of respondent and time of survey would also increase, increasing risk of recall error.

Based on the birth date of the child, we calculated whether the foetal stage had coincided with Ramadan.[4] The mother was asked if she had fasted during the pregnancy. The mother's memory was tested by asking her questions on incidents that had occurred in the 1990s. Only those who passed the recall test were included in the survey. Thus, we are able to clearly distinguish between children exposed to foetal starvation and those who were not. In addition, information on the asset holding and economic status of the family and occupation of parents (at the time of pregnancy and current), along with education of parents, was collected.

The field survey was undertaken by a local NGO, Sundarban Unnayan Niketan (SUN), under the supervision of the Investigators. The NGO identified Muslim women with children born between our reference period (1993–1997) and administered the first round of survey, in which birth dates of the children born in this period were noted, recall tests performed to assess reliability of the respondent's memory and information collected on the educational level of the respondent and her husband, past and current occupation of the couple and asset ownership. This part of the survey was undertaken from July to October 2013.

In the second stage of the survey, undertaken between December 2013 and April 2014, camps were held in different localities within the survey region, and the following measurements of children listed in the first stage were taken:

(1) *Body mass index* (BMI), kilograms/metre^2: BMI is a measure of one's height in relation to one's weight and used to categorise a person as underweight, normal weight, overweight or obese.
(2) *Mid-upper arm circumference* (MUAC), centimetre: the circumference of the left upper arm, measured at the mid-point between the tip of the shoulder and the tip of the elbow (olecranon process and the acromion). MUAC is used for the assessment of nutritional status. It is a good predictor of mortality, and in many studies, MUAC predicted death in children better than any other anthropometric indicator.

[4]We used the Muslim calendar available in the webpage http://www.al-islam.com, cross-checking the dates with http://www.islamicfinder.org/dateConversion.php. Birth dates were cross-checked with official certificates and documents.

(3) *Skinfold thickness (in the bicep, tricep, subscapular and suprailiac regions)*, millimetre: measures of subcutaneous fat at different regions of the human body.
(4) *Body density*, grams/millilitre: measure of body fat, based on measurements of subcutaneous fat in the bicep, tricep, subscapular and suprailiac regions.[5]
(5) *Body fat*, percentage: amount of body fat mass with regard to total body weight, expressed as a percentage. This was measured using the bio-electrical impedance method.

The instruments used for measuring were Galaxy anthropometer, Harpenden skinfold calliper and Omron Karada HBF 375 Body Fat Monitor. Only those respondents who did not have a meal in the two hours preceding the survey were measured as the metabolic activity may interfere with readings of body fat monitor.

2.4 Sample Profile

A total of 684 women and 447 youths who were offspring of the women respondents (of which 57.40% were males) were surveyed using questionnaires approved by the Ethical Committee of Indian Statistical Institute, Kolkata. At the time of the survey, informed consent of the respondent (both mother and child) was obtained. After entering the data, inconsistent or incomplete entries were discarded, leaving a sample of 424 respondents. Out of them, 262 were males and 162 females. The percentage exposed to foetal starvation was 64.89 and 77.16 for males and females, respectively.[6] The sample profile is given in Table 2.

2.5 Statistical Methodology

Our study assumes that consumption is positively related to the standard of living (SLI).[7] Therefore, information on past SLI provides cues about the consumption standard of the family and, more importantly, the mother. To test the first objective, we divided the sample into four groups based on past (during pregnancy) and current

[5]Estimated using the formula given in Durnin and Womersley (Durnin and Womersley 1974).

[6]It implies that 120 and 75 male and females formed the treatment group; and 92 and 37 male and female respondents formed the control group.

[7]Standard of living indices were constructed based on information elicited on asset ownership (radio and cycle for past SLI; TV, motorbike and electricity for current SLI) quality of housing (whether *pucca* and roof type) and ownership of land. Principal component analysis was used to construct the index. The values of KMO and χ^2 (for Bartlett's test for Sphericity) were 0.63 and 191.84, respectively, for the current SLI index. For the past SLI index, these values were 0.52 and 274.72, respectively.

Respondents were divided into three groups based on values of SLI index—low, medium and high. To sharpen the contrast, only individuals with high and low SLI were taken in the study.

Table 2 Sample profile—by control/study group (percentages)

	No exposure to foetal starvation	Exposure to foetal starvation
Education of mother		
Illiterate	39.29	60.71
Can read and write	20.83	79.17
Class 1–4	22.95	77.05
Class 5–8	30.65	69.35
Class 9–10	28.57	71.43
Others	0	100
Education of father		
Illiterate	40.86	59.14
Can read and write	30.38	69.62
Class 1–4	38.98	61.02
Class 5–8	25	75
Class 9–10	19.4	80.6
Class 11–12	25	75
Graduate and above	33.33	66.67
Others	0	100
Past occupation of mother		
Housewife/unemployed	31.31	68.69
Labourer	60	40
Artisans	4.76	95.24
Primary sector	50	50
Current occupation		
Housewife/unemployed	31.22	68.78
Labourer	50	50
Artisans	19.23	80.77
Service (salaried)	66.67	33.33
Others	66.67	33.33
Respondent's gender		
Male	35.11	64.89
Female	22.84	77.16
Respondent's occupation		
Housewife/unemployed	30.15	69.85
Labourer	41.57	58.43
Agriculture (-related)	33.33	66.67
Pisciculture	75	25
Transport	12.5	87.5

(continued)

Table 2 (continued)

	No exposure to foetal starvation	Exposure to foetal starvation
Handicrafts	20.22	79.78
Other trade	40.74	59.26
Salaried	66.67	33.33
Student	24	76
Others	11.11	88.89
Respondent's education		
Illiterate	31.25	68.75
Can read/write	33.33	66.67
Class 1–4	35.82	64.18
Class 5–8	27.66	72.34
Class 9–10	27.94	72.06
Class 11–12	31.37	68.63
Graduate	27.78	72.22
Other	100	0
Previous financial status		
Poorest	42.40	52.22
Middle	13.60	13.31
Richest	44.00	34.47
Current financial status (5 quintiles)		
Poorest	47.29	4.00
Middle	25.58	28.47
Richest	27.13	44.00

SLI—HH (SLI was high in both periods), HL (past SLI was high, while current SLI is low), LL (SLI was low in both periods) and LH (individuals had low SLI in the past, but this improved subsequently). High past SLI implies a high nutritional intake during pregnancy. Based on this signal, the foetus anticipates a nutrition-plenty environment. The reverse holds for respondents with low past SLI. Predictions are correct for the two groups, HH and LL; LH and HL, on the other hand, represent mismatches.

Monaghan's four propositions, stated in the introduction, can be restated as:

(1) **H11**: Outcome of respondents with high past SLI worsens if SLI decreases ($x^{HL} < x^{HH}$);
(2) **H12**: Outcome of respondents with low past SLI worsens if SLI increases ($x^{LH} < x^{LL}$);
(3) **H13**: Outcome of respondents with high current SLI will be better if their past SLI was also high ($x^{LH} < x^{HH}$); and,

(4) **H14**: Outcome of respondents with low current SLI will be better if their past SLI was also low ($x^{HL} < x^{LL}$).

when x is the outcome of interest described earlier.

To test whether such differences are statistically significant is not an easy task given that the samples are small (and will decrease as we progress to further tests) and unequal. The reliability of the Student's t-statistic has been questioned as it assumes that sample sizes are equal, have equal variances and are normally distributed (Ruxton 2006). Although the use of the Mann–Whitney U-test has sometimes been suggested, its inability to effectively substitute for the Student's t has been demonstrated (Kasuya 2001; Neuhäuser 2002). In such a situation, a modification of the degrees of freedom (Welch 1938, 1947) to increase the power of the t-test for unequal samples with unequal variances has been found to be useful (Ruxton 2006), particularly as it ensures better control of Type 1 error rates (Delacre et al. 2017). Resampling is another technique that has been suggested (Edgington 1964, 1969). This consists of taking repeated permutations of the original sample and estimating in how many cases the hypothesis is proved. In this study, we have taken 1000 simulations of the sample based on the Monte Carlo method. Welch's t-test was performed for all the permutations, and the proportion of times, our hypothesis was accepted, is reported.

The second set of hypotheses comprises of checking whether the short-term intensive nutritional shock of Ramadan fasting overrides the signal provided by nutritional intake during the eight months comprising the rest of the foetal stage.

To test this, we divided the sample into four groups based on their past and current SLI. We examined the impact of a short-term nutritional shock, in the form of Ramadan fasting, for respondents with high past SLI scores (HH and HL). Respondents belonging to the LH and HL groups are not considered as Ramadan fasting merely accentuates the low levels of nutritional intake of the mother. Our hypotheses are:

(1) **H21**: For families whose economic status has deteriorated (HL), children exposed to foetal starvation have better outcomes than those children not exposed to any such nutritional shock $\left(x_{FS}^{HL} > x_{NE}^{HL}\right)$, and
(2) **H22**: For families whose economic status has remained about the same (HH), children exposed to foetal starvation have worse outcomes than those children not exposed to any such nutritional shock $\left(x_{FS}^{HH} < x_{NE}^{HH}\right)$

when

FS indicates that the respondent has been exposed to foetal starvation, and
NE indicates that respondent has not been exposed to foetal starvation.

In the third step of our analysis, we test whether foetal starvation genetically programs the offspring to face latter life economic adversity more effectively than an offspring anticipating a nutrition-rich future. It is tested using the double-difference method. The first differences are:

$$D_{FE} = (x_{HH} - x_{HL})_{FE} \text{ for the study group}$$

$$D_{\mathrm{NE}} = (x_{\mathrm{HH}} - x_{\mathrm{HL}})_{\mathrm{NE}} \text{ for the control group}$$

It represents the decline in outcome due to deterioration in SLI. We then take the second difference:

$$\mathrm{DIDM} = D^{i}_{\mathrm{NE}} - D^{i}_{\mathrm{FE}}$$

Given that foetal starvation prepares respondents to face a nutrition deficient environment, it should be the respondents exposed to foetal starvation who are better able to face the adversity relative to respondents not exposed to foetal starvation. Although the anthropometric outcome of both groups will fall, the decline will be less for the control group vis-à-vis for the study group. The hypothesis is:

(1) **H31**: For families whose economic status has deteriorated (HL), the mean decline in outcome for children exposed to foetal starvation is less than the mean decline in outcome for those children not exposed to any such nutritional shock, i.e. DIDM > 0.

Since this is a difference-in-difference, a regression of the outcomes studied should be regressed on an EXPOSURE dummy (=1 if exposed to Ramadan fasting in the foetal stage, =0 otherwise), an indicator labelled TREATMENT (indicating whether respondent belongs to HH or HL group) and the interaction of these two dummies. The coefficient of the interaction is an estimate of the average treatment effect (ATE). As control variables, we have used birth order, and dummies indicating whether respondents were delivered prematurely, were breastfed and are currently engaged in any manual work. Past and current economic status is taken care of in the TREATMENT dummy. For households belonging to the HH group (high SLI in both periods), TREATMENT takes a value of unity; in case of households whose economic status has fallen (i.e. HL), TREATMENT equals 0.[8]

3 Findings and Discussion

3.1 Mismatch Between Actual and Expected Environment

Results are reported in Table 3. For male respondents, the percentage of indicators confirming the hypotheses are 100 (HH > HL), 88 (HH > LH), 63 (LL > LH) and 75 (LL > HL). Welch's test supported our hypothesis in only 23% of the simulations for suprailiac skinfold thickness (males, HH vs. HL) and in 32 and 38% cases for BMI and skinfold thickness at bicep region (male: LL vs. HL). On the other hand, the proportion of simulations supporting our hypothesis was high for all skinfold thicknesses (except bicep), body density and body fat per cent (Male: LL vs. LH).

[8]LH and LL households are not considered in these hypotheses.

Table 3 Mean outcomes for respondents belonging to HH, HL, LH and LL households—male and female respondents

Anthropometric indicators: male	HH (33)	HL (35)	Welch's t	Prob.	c/n	HH (33)	LH (30)	Welch's t	Prob.	c/n
BMI	21.51	19.42	1.83	0.04	0.9980	21.51	20.08	1.24	0.11	1.0000
Mid-upper arm circumference	24.28	22.48	2.01	0.02	0.9800	24.28	23.47	1.16	0.13	0.9750
Biceps: skinfold thickness	5.2	4.36	1.15	0.13	0.8650	5.2	4.71	0.76	0.23	0.9380
Triceps: skinfold thickness	7.85	5.94	1.57	0.06	0.9630	7.85	5.54	2.02	0.02	0.7950
Subscapular skinfold thickness	12.47	8.59	2.89	0.00	0.9980	12.47	10.45	1.40	0.08	0.8850
Suprailiac skinfold thickness	6.64	5.51	1.11	0.14	0.8180	6.64	6.04	0.54	0.30	0.2350
Body density	8.94	8.58	2.49	0.01	0.9870	8.94	7.56	1.40	0.08	0.8380
Percentage of body fat	15.57	13.13	1.84	0.04	0.8980	15.57	13.34	1.61	0.06	0.5590
Anthropometric indicators: male	LL (59)	LH (30)	Welch's t	Prob.	c/n	LL (59)	HL (35)	Welch's t	Prob.	c/n
BMI	19.09	20.08	−1.93	0.97	0.0800	19.09	19.42	−0.69	0.76	0.3170
Mid-upper arm circumference	22.66	23.47	−1.30	0.90	0.1460	22.66	22.48	0.22	0.41	0.5650
Biceps: skinfold thickness	4.2	4.71	−1.46	0.93	0.1990	4.2	4.36	−0.34	0.63	0.3820
Triceps: skinfold thickness	7.02	5.54	2.01	0.02	0.9040	7.02	5.94	1.28	0.10	0.8550
Subscapular skinfold thickness	11.07	10.45	0.54	0.29	0.7070	11.07	8.59	2.46	0.01	0.9820
Suprailiac skinfold thickness	7.48	6.04	1.35	0.09	0.8820	7.48	5.51	2.03	0.02	0.9580
Body density	7.91	7.56	0.60	0.27	0.7150	7.91	8.58	1.85	0.03	0.9530
Percentage of body fat	15.27	13.34	1.65	0.05	0.8550	15.27	13.13	1.95	0.03	0.8970

(continued)

Table 3 (continued)

Anthropometric indicators: female	HH (26)	HL (19)	Welch's t	Prob.	c/n	HH (26)	LH (16)	Welch's t	Prob.	c/n
BMI	19.25	17.72	2.07	0.02	0.9470	19.25	19.99	−0.91	0.82	0.7160
Mid-upper arm circumference	21.3	20.56	0.97	0.17	0.7900	21.3	22.16	−1.21	0.88	0.3940
Biceps: skinfold thickness	6.57	6.84	−0.26	0.60	0.3790	6.57	6.37	0.22	0.41	0.0380
Triceps: skinfold thickness	11.07	11.03	0.02	0.49	0.5190	11.07	10.15	0.54	0.30	0.8520
Subscapular skinfold thickness	12.77	10.63	1.14	0.13	0.9040	12.77	11.64	0.77	0.22	0.8810
Suprailiac skinfold thickness	3.63	3.38	0.19	0.43	0.5930	3.63	3.48	0.85	0.20	0.7640
Body density	3.86	3.78	−0.55	0.29	0.7450	3.86	3.79	0.50	0.31	0.6510
Percentage of body fat	25.08	22.75	1.51	0.07	0.8280	25.08	25.26	−0.13	0.55	0.1580
Anthropometric indicators: female	LL (41)	LH (16)	Welch's t	Prob.	c/n	LL (41)	HL (19)	Welch's t	Prob.	c/n
BMI	18.83	19.99	−1.49	0.93	0.0840	18.83	17.72	1.60	0.06	0.9210
Mid-upper arm circumference	21.58	22.16	−0.91	0.82	0.2750	21.58	20.56	1.44	0.08	0.8800
Biceps: skinfold thickness	7.81	6.37	1.71	0.05	0.9580	7.81	6.84	0.96	0.17	0.8980
Triceps: skinfold thickness	9.71	10.15	−0.32	0.62	0.3590	9.71	11.03	−0.81	0.79	0.1630
Subscapular skinfold thickness	11.21	11.64	−0.37	0.64	0.3720	11.21	10.63	0.35	0.36	0.6680
Suprailiac skinfold thickness	3.54	3.48	0.24	0.41	0.0573	3.54	3.38	−0.40	0.65	0.2910
Body density	3.82	3.79	0.22	0.41	0.5710	3.82	3.78	0.31	0.38	0.6210
Percentage of body fat	27.12	25.26	1.43	0.08	0.7900	27.12	22.75	2.87	0.00	0.9790

Note Figures in parentheses below HH, HL, LH and LL are sample size

Among females, three of our hypotheses (HH > HL, HH > LH and LL > HL) receive support from results of more than half the indicators—for seven, five and six out of the eight indicators, respectively. In these three cases, a low proportion of simulations support our hypothesis for biceps (HH vs. HL), mid-upper arm circumference, skinfold thickness in bicep and body fat per cent (HH vs. LH) and skinfold thickness in bicep and suprailiac regions (LL vs. HL). Therefore, there seems considerable evidence to support the starting group of hypotheses (H11–H14), stating that outcomes will be better if the signal received at the foetal stage matches with the actual future environment, vis-à-vis when there is a mismatch.

Overall, when signals of nutrition deficiency at the foetal stage match with the subsequent environment, male respondents were found to have higher MUAC, skinfold thickness at the triceps and subscapular region, body density and body fat; among female respondents, a similar finding was observed for skinfold thickness in the subscapular region, body density and body fat. In addition, BMI was also higher in general. Further, mean values of such respondents were in the normal range. The control group, in contrast, had lower values for these outcomes, bordering on the low-normal boundary. It indicates that mismatch in signals of expected environment and the ensuing environment may reduce body fat levels. This may have long-term health consequences (Blackburn and Phillips 2001)—affecting the ability to store energy (leading to fatigue), affect the functioning of the brain and cardiovascular system, reduce thyroid levels, affect the absorption of calcium and weaken muscles.

3.2 Impact of Ramadan Fasting

The second question that we turn to is the impact of Ramadan fasting. Past SLI provides information about consumption levels over the entire gestation period. But are short-term nutritional shocks also important as indications of the future environment? In particular, do short-run nutritional shocks, like Ramadan fasting, override the longer run signals provided by past SLI (and consumption) to the foetus?

Results are given in Table 4. Here, the evidence seems to be against the hypothesis. This confirms studies of the impact of Ramadan fasting among Muslims in diverse countries (Kavehmanesh and Abolghasemi 2004; Ziaee et al. 2010; Alwasel et al. 2013a, b; Savitri et al. 2014, 2018; Glazier et al. 2018). Results—whether percentage difference, Welch's *t*-test or *t*-tests after Monte Carlo simulations—supports our hypothesis to some extent only for male respondents in the HH group. For this group, *t*-tests confirm our hypothesis for skinfold thickness in tricep, subscapular and suprailiac regions and percentage of body fat. After Monte Carlo simulations, t-tests support our hypothesis for six out of eight indicators. The exceptions are mid-upper arm circumference and bicep skinfold thickness.

The failure to find evidence in support of the impact of Ramadan fasting is in contrast to the favourable results obtained for the impact of a sustained nutritional deficit. The present study is not directly comparable with existing studies of Ramadan fasting that uses large-scale survey data and relate coincidence of Ramadan with ges-

Table 4 Mean outcomes for male respondents belonging to HH and HL households—by exposure to foetal starvation

Anthropometric indicators: male	SLI has deteriorated: HL (17, 18)					SLI has always remained high: HH (12, 21)				
	No exposure	Exposure to foetal starvation	Welch's *t*	Probability	*c*/*n*	No exposure	Exposure to foetal starvation	Welch's *t*	Probability	*c*/*n*
BMI	19.64	18.66	0.54	0.70	0.0270	23.01	20.34	0.82	0.21	0.6450
Mid-upper arm circumference	21.46	23.08	−1.38	0.09	0.3800	24.48	23.78	0.25	0.40	0.0570
Biceps: skinfold thickness	5.52	6.47	−0.51	0.31	0.2720	8.4	6.36	0.43	0.34	0.3430
Triceps skinfold thickness	7.56	12.28	0.94	0.82	0.0150	14.65	11.1	1.37	0.09	0.7400
Subscapular skinfold thickness	9.65	6.98	1.59	0.93	0.0030	16.38	10.49	2.36	0.02	0.8520
Suprailiac skinfold thickness	6.18	4.95	1.14	0.87	0.0640	8.6	5.46	1.43	0.09	0.7650
Body density	3.45	3.32	1.17	0.87	0.0160	3.85	3.52	1.98	0.03	0.8190
Percentage of body fat	9.7	8.32	1.10	0.86	0.2690	12.18	10.3	0.78	0.22	0.7500

(continued)

Table 4 (continued)

Anthropometric indicators: female	SLI has deteriorated: HL (5, 10)					SLI has always remained high: HH (4, 18)				
	No exposure	Exposure to foetal starvation	Welch's *t*	Probability	*c*/*n*	No exposure	Exposure to foetal starvation	Welch's *t*	Probability	*c*/*n*
BMI	17.12	17.69	−0.97	0.83	0.6540	18.78	19.58	−0.50	0.70	0.2980
Mid-upper arm circumference	20.84	20.41	0.26	0.60	0.6990	20.53	21.66	−0.87	0.70	0.6050
Biceps: skinfold thickness	4.14	5.82	−1.10	0.14	0.0940	5.61	5.53	0.70	0.26	0.0800
Triceps skinfold thickness	6.47	5.13	−1.90	0.04	0.0620	10.3	6.6	0.09	0.20	0.0110
Subscapular skinfold thickness	8.16	10.72	−1.62	0.06	0.3560	13.55	13.49	0.94	0.20	0.0970
Suprailiac skinfold thickness	8.66	8.08	0.03	0.51	0.3290	9.9	9.44	0.27	0.40	0.5450
Body density	3.63	3.79	−1.05	0.16	0.2320	4.01	3.90	0.31	0.39	0.1740
Percentage of body fat	16.68	18.25	−0.25	0.41	0.4210	18.88	21.67	−1.34	0.91	0.4320

Note Figures besides HH and HL, in parentheses, are sample sizes for treatment and control groups, respectively tation period (irrespective of whether mother had fasted or not) to onset of diseases at midlife. The sample size is much larger in the case of the latter, the proxy for foetal starvation is different and the phenomenon being studied is different. Nevertheless, the results of this subsection question the advisability of relying on short-term nutritional shocks in the form of Ramadan fasting as a proxy for foetal starvation, ignoring long-term nutritional intakes. It is a finding that is in line with meta-reviews suggesting that Ramadan fasting does not adversely affect birth weight and that there is insufficient evidence regarding potential effects on other perinatal outcomes (Glazier et al. 2018).

3.3 *Facing Economic Adversity*

In the third step of our analysis, we test whether persons exposed to foetal starvation are able to adapt better to conditions of scarcity and economic hardship than those who are not exposed to foetal starvation, using a double-difference approach. Results are given in Table 5.

Among male respondents, ATE is statistically significant for BMI (10% level), mid-upper arm circumference (10% level) and skinfold thickness in the subscapular region (5% level). Among females, ATE is high for skinfold thickness in tricep region (5% level). There seems some evidence indicating that exposure to foetal starvation may better prepare respondents to face a decline in economic status.

Table 5 Average treatment effects for male and female respondents—by exposure

Anthropometric outcomes	Male ($n_T = n_C = 17$)			Female ($n_T = 15$; $n_C = 5$)		
	ATE	*t*-ratio	Prob.	ATE	*t*-ratio	Prob.
BMI	4.04	1.77	0.08	0.90	0.41	0.69
Mid-upper arm circumference	3.53	1.90	0.06	−0.12	−0.06	0.95
Biceps: skinfold thickness	1.61	1.00	0.32	3.03	1.05	0.30
Triceps: skinfold thickness	3.95	1.54	0.13	10.41	2.00	0.05
Subscapular skinfold thickness	5.74	2.14	0.04	7.43	1.49	0.15
Suprailiac skinfold thickness	1.14	0.53	0.60	1.65	0.32	0.75
Body density	0.28	1.30	0.20	0.45	1.09	0.28
Percentage of body fat	2.02	0.72	0.48	1.34	−0.30	0.76

Note Figures in parentheses, besides male and female, are treatment and control group sizes

4 Conclusion

To sum up, the FOH argues that exposure to foetal starvation may increase disposition to develop diabetes mellitus, hypertension, cardiac ailments, nephrological diseases, etc. in midlife. The underlying mechanism is the growth of a "thrifty" phenotype that prepares the organism to survive in nutrition-deficit environments, but results in maladaptation if the organism is subsequently exposed to prosperous conditions. The PAR theory extends this idea to suggest that the foetus receives signals of the expected environment in which it will grow up. If the signals predict the future environment correctly, then the organism is suited to its local environment and will have optimal outcomes. If, however, signals are faulty and wrongly predict the future environment, then outcomes will be adversely affected as the organism is maladjusted to its local environment.

The present study attempts to contribute to the literature by testing for PAR in a chronically underdeveloped locality. Our study finds some evidence in support of PAR, particularly among male respondents. Mismatching of the predicted and actual environment appears to affect body fat levels, reducing measures to values slightly above the low levels. This may have serious long-term health consequences in midlife as the chances of suffering from non-communicable diseases will increase. However, there is scope to extend the study by undertaking a study with a larger sample size, which would enable controlling for more variables.

Even the limited evidence produced by this pilot study has important implications for maternal and child health—an important component of Sustainable Development Goals adopted by the United Nations in 2015. The current study shows that individuals exposed to foetal starvation may become programmed to function *only in nutrition-deficit conditions*. Subsequent prosperity affects their anthropometric outcomes adversely, which may increase the probability of developing lifestyle diseases in midlife. A sharp rise in non-communicable diseases in developing societies has been reported in various studies (World Health Organization 2018). It implies that, unless we ensure proper nutrition to pregnant women, their offsprings are likely to be programmed in such a manner that they will become susceptible to non-communicable diseases and unfit to reap the benefits of growth and development. This calls for concerted attempts to guarantee adequate nutrition during pregnancy to ensure that growth leads to an inclusive society.

Designing such strategies, however, is not easy. Nutritional interventions designed to increase birth weight have had only modest success (Kramer 1993; Rasmussen 1999). Even if one assumes that association between birth weight (BW) and cardiovascular diseases (CVD) is causal, increasing birth weights will improve cardiovascular outcomes marginally (Joseph and Kramer 1997). Such findings imply that "although studying factors in early life has offered some new ways of thinking about the origins of CVD, increasing BW is likely to be much less effective in reducing mortality than modifying the traditional risk factors observed during adulthood" (Rasmussen 2018: 87). Public health policies to increase birth weight has costs; they may expose infants to the risk of being macrosomic at birth or developing cancer

(Rasmussen 2018). Further, recent studies have argued that birth weight is only a poor proxy for a process or processes that affect the foetus. Improving our understanding of how and when genetic programming occurs, therefore, is an important precondition to design intervention strategies during pregnancy to reduce the development of the chronic disease.

Acknowledgements The assistance provided by members of Sundarban Unnayan Niketan, and Koumi Dutta and Rabia Mondol, during the survey is gratefully acknowledged. Sukumar Sarkar and Antara Dhar entered the data. Probal Chaudhuri (Theoretical Statistics and Mathematics Unit, Indian Statistical Institute Kolkata) made some useful suggestions regarding the methodology. We are also grateful to study participants for having borne with our probing. Responsibility for any remaining errors lies with authors.

Consent for Publication Respondents had given their consent to the data being used for academic purposes after concealing their individual identities.

Funding The study was funded by the Indian Institute of Technology, Kharagpur. The role of funding body was restricted to providing the finance; it did not play any role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

Availability of Data and Material The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate The study is based on a primary survey eliciting history, socio-economic characteristics and anthropometric measures of respondents and their children. There was no intervention. Hence, clearance was not required from the ethical committee. However, informed written consent, witnessed by an independent person, was taken before administering the questionnaires.

References

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. *Journal of Political Economy, 114,* 672–712.

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives, 25,* 153–172. https://doi.org/10.1257/jep.25.3.153.

Almond, D., & Mazumder, B. (2011). Health capital and the prenatal environment: The effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics, 3,* 56–85. https://doi.org/10.1257/app.3.4.56.

Alwasel, S. H., Harrath, A., Aljarallah, J. S., Abotalib, Z., Osmond, C., Al Omar, S. Y., et al. (2013). Intergenerational effects of in utero exposure to Ramadan in Tunisia. *American Journal of Human Biology, 25,* 341–343. https://doi.org/10.1002/ajhb.22374.

Banerjee, A., Duflo, E., Postel-Vinay, G., & Watts, T. (2010). Long-run health impacts of income shocks: Wine and phylloxera in nineteenth-century France. *Review of Economics and Statistics, 92,* 714–728. https://doi.org/10.1162/REST_a_00024.

Barker, D. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal, 301*, 1111. https://doi.org/10.1136/bmj.301.6761.1111.

Bateson, P. (2001). Fetal experience and good adult design. *International Journal of Epidemiology, 30,* 928–934. https://doi.org/10.1093/ije/30.5.928.

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., et al. (2004). Developmental plasticity and human health. *Nature, 430,* 419–421. https://doi.org/10.1038/nature02725.

Blackburn, G. L., Phillips, J. C., Morreale, S. (2001). Physician's guide to popular low-carbohydrate weight-loss diets. *Cleveland Clinic Journal of Medicine, 68*, 761, 765–766, 768–769, 773–774.

Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology, 30,* 92–101. https://doi.org/10.5334/irsp.82.

Durnin, J. V. G. A., & Womersley, J. (1974). Body fat assessed from the total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition, 32,* 77–97.

Edgington, E. S. (1964). Randomization tests. *The Journal of Psychology, 57,* 445–449. https://doi.org/10.1080/00223980.1964.9916711.

Edgington, E. S. (1969). Approximate randomization tests. *Journal of Psychology, 72,* 143–149.

Fall, C. H. D., Stein, C. E., Kumaran, K., Cox, V., Osmond, C., et al. (1998). Size at birth, maternal weight, and type 2 diabetes in South India. *Diabetic Medicine, 15,* 220–227.

Glazier, J. D., Hayes, D. J. L., Hussain, S., D'Souza, S. W., Whitcombe, J., Heazell, A. E. P., & Ashton, N. (2018). The effect of Ramadan fasting during pregnancy on perinatal outcomes: A systematic review and meta-analysis. *BMC Pregnancy and Childbirth, 18*, 421.

Gluckman, P. D., & Hanson, M. A. (2007). Developmental plasticity and human disease: Research directions. *Journal of Internal Medicine, 261,* 461–471. https://doi.org/10.1111/j.1365-2796.2007.01802.x.

Gluckman, P. D., Hanson, M. A., & Spencer, H. G. (2005a). Predictive adaptive responses and human evolution. *Trends in Ecology & Evolution, 20,* 527–533. https://doi.org/10.1016/j.tree.2005.08.001.

Gluckman, P. D., Hanson, M. A., Spencer, H. G., & Bateson, P. (2005b). Environmental influences during development and their later consequences for health and disease: Implications for the interpretation of empirical studies. *Proceedings of the Royal Society B: Biological Sciences, 272,* 671–677. https://doi.org/10.1098/rspb.2004.3001.

Gluckman, P. D., Hanson, M. A., Beedle, A. S., & Spencer, H. G. (2008). Predictive adaptive responses in perspective. *Trends in Endocrinology and Metabolism, 19,* 109–110. https://doi.org/10.1016/j.tem.2008.02.002.

Godfrey, K. M., & Barker, D. J. (2001). Fetal programming and adult health. *Public Health Nutrition, 4,* 611–624.

Government of West Bengal. (2008). *District human development report: South twenty four Parganas*. Kolkata: UNDP.

Hales, C. N., & Barker, D. J. P. (2001). The thrifty phenotype hypothesis: Type 2 diabetes. *British Medical Bulletin, 60,* 5–20. https://doi.org/10.1093/bmb/60.1.5.

Haymann, D. J. (1962). Thalidomide and congenital abnormalities. *The Lancet, 1,* 45.

Hobel, C., & Culhane, J. (2003). Role of psychosocial and nutritional stress on poor pregnancy outcome. *The Journal of nutrition, 133,* 1709S–1717S.

Joseph, K. S., & Kramer, M. S. (1997). Should we intervene to improve fetal growth? In D. Kuh & Y. Ben-Schlomo (Eds.), *A life course approach to chronic disease epidemiology* (pp. 277–295). Oxford: Oxford University Press.

Kasuya, E. (2001). Mann-Whitney U-test when variances are unequal. *Animal Behaviour, 61,* 1247–1249.

Kavehmanesh, Z., & Abolghasemi, H. (2004). Maternal Ramadan fasting and neonatal health. *Journal of Perinatology, 24,* 748–750.

Kelly, E. (2009). *The scourge of Asian Flu: In utero exposure to pandemic influenza and the development of a cohort of British children*. IFS working papers, No. 09, 17. London.

Kramer, M. S. (1993). Effects of energy and protein intakes on pregnancy outcome: An overview of the research evidence from controlled clinical trials. *American Journal of Clinical Nutrition, 58,* 627–635.

Kumaran, K., Fall, C. H. D., Martyn, C. N., Vijayakumar, M., Shier, R., & Stein, C. (2000). Blood pressure, arterial compliance, and left ventricular mass: No relation to small size at birth in south Indian adults. *Heart, 83*, 272–277.

Kunto, Y. S., & Mandemakers, J. J. (2018). The effects of prenatal exposure to Ramadan on stature during childhood and adolescence: Evidence from the Indonesian Family Life Survey. *Economics of Human Biology, 19,* 29–39.

Kunto, Y. S., & Mandemakers, J. J. (2019). The effects of prenatal exposure to Ramadan on stature during childhood and adolescence: Evidence from the Indonesian Family Life Survey. *Economics & Human Biology, 33,* 29–39.

Landro, L. (2010). Healthy living for two. *Wall Street Journal*.

McCay, C. M., & Crowell, M. F. (1934). Prolonging the life span. *The Scientific Monthly, 39*(5), 405–414.

Mccay, C. M., Crowell, M. P., & Maynakd, L. A. (1935). The effect of retarded growth upon the length of life span and upon the ultimate body size. *Journal of Nutrition, 10,* 63–79.

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 1635–1645. https://doi.org/10.1098/rstb.2007.0011.

Nelson, R. E. (2010). Testing the Fetal Origins Hypothesis in a developing country: Evidence from the 1918 Influenza Pandemic. *Health Economics, 19,* 1181–1192. https://doi.org/10.1002/hec.1544.

Neuhäuser, M. (2002). Two-sample tests when variances are unequal. *Animal Behaviour, 63,* 823–825. https://doi.org/10.1006/anbe.2002.1993.

Paneth, N., & Susser, M. (1995). Early origin of coronary heart disease (the "Barker hypothesis"). *BMJ (Clinical Research ed.), 310,* 411–412.

Rasmussen, K. M. (1999). Causes of intrauterine growth restriction. In M. C. McCormick & J. E. Siegel (Eds.), *Prenatal care: Effectiveness and implementation* (pp. 153–174). Cambridge: Cambridge University Press.

Rasmussen, K. M. (2018). The "Fetal Origins" hypothesis: Challenges and opportunities for maternal and child nutrition. *Annual Review of Nutrition, 21,* 73–95.

Ravelli, A. C., van der Meulen, J. H., Michels, R. P., Osmond, C., Barker, D. J., Hales, C. N., et al. (1998). Glucose tolerance in adults after prenatal exposure to famine. *Lancet (London, England), 351,* 173–177.

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioural Ecology, 17,* 688–690.

Savitri, A. I., Amelia, D., Painter, R.C., Mohammad Baharuddin, Roseboom, T. J., Grobbee, D. E., & Uiterwaal, C. S. P. M. (2018). Ramadan during pregnancy and birth weight of newborns. *Journal of Nutritional Science, 7*.

Savitri, A. I., Yadegari, N., Bakker, J., van Ewijk, R. J. G., Grobbee, D. E., Painter, R. C., et al. (2014). Ramadan fasting and newborn's birth weight in pregnant Muslim women in The Netherlands. *British Journal of Nutrition, 112,* 1503–1509.

Schwandt, H. (2017). *The lasting legacy of seasonal influenza: In-utero exposure and labor market outcomes*. COHERE discussion paper No. 5/2017. Odense.

Seckl, J. R., Roberts, P., Wild, A. E., Barker, D. J. P., Ravelli, A. C. J., Schroeder-Tanka, J. M., et al. (1997). Glucocorticoids, feto-placental 11 beta-hydroxysteroid dehydrogenase type 2, and the early life origins of adult disease. *Steroids, 62*, 89–94. https://doi.org/10.1016/s0039-128x(96)00165-1.

Seckl, J. R., & Holmes, M. C. (2007). Mechanisms of disease: Glucocorticoids, their placental metabolism and fetal "programming" of adult pathophysiology. *Nature Clinical Practice Endocrinology & Metabolism, 3,* 479–488. https://doi.org/10.1038/ncpendmet0515.

Smithells, R. W. (1962). Thalidomide and malformations in Liverpool. *The Lancet, 1,* 1270–1273.

Stanner, S. A., Bulmer, K., Andrès, C., Lantseva, O. E., Borodina, V., Poteen, V. V., & Yudkin, J. S. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *BMJ (Clinical Research ed.), 315*, 1342–1348.

Stein, C. E., Fall, C. H. D., Kumaran, K., Barker, D. J. P., Osmond, C., & Cox, V. (1996). Fetal growth and coronary heart disease in South India. *Lancet, 348,* 1269–1273.

Stein, C. E., Kumaran, K., Fall, C. H. D., Shaheen, S. O., Osmond, C., & Barker D. J. P. (1997). Relation of fetal growth to adult lung function in South India. *Thorax, 52*, 895–899.

Taborsky, B. (2006). The influence of juvenile and adult environments on life-history trajectories. *Proceedings. Biological Sciences, 273,* 741–750. https://doi.org/10.1098/rspb.2005.3347.

Van Abeelen, A. F. M., Elias, S. G., Bossuyt, P. M. M., Grobbee, D. E., Van Der Schouw, Y. T., Roseboom, T. J., & Uiterwaal, C. S. P. M. (2012). Famine exposure in the young and the risk of type 2 Diabetes in adulthood. *Diabetes, 61,* 2255–2260.

Van Ewijk, R. (2011). Long-term health effects on the next generation of Ramadan fasting during pregnancy. *Journal of Health Economics, 30*, 1246–1260. https://doi.org/10.1016/j.jhealeco.2011.07.014.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29,* 350–362. https://doi.org/10.2307/2332010.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika, 34,* 23–35. https://doi.org/10.2307/2332510.

Wells, J. C. K. (2006). Is early development in humans a predictive adaptive response anticipating the adult environment? *Trends in Ecology & Evolution, 21,* 424–425. https://doi.org/10.1016/j.tree.2006.05.006.

World Health Organization. (2018). *Noncommunicable diseases country profile 2018*. Geneva: World Health Organization.

Yajnik, C. S. (2004). Early life origins of insulin resistance and type 2 diabetes in India and other Asian countries. *The Journal of nutrition, 134,* 205–210.

Ziaee, V., Kihanidoost, Z., Younesian, M., Bateni, F., Kazemianfar, Z., & Hantoushzadeh, S. (2010). The effect of ramadan fasting on outcome of pregnancy. *Iranian Journal of Pediatrics, 20,* 181–186.

Inequality and Well-Being

Sustainable Development and the Problem of Economic Inequality

Anup Kumar Sinha

Abstract Sustainable development is a complex concept, and most descriptions of the process involve aggregate measures of wealth and capital stock, along with some problems of aggregation and valuation of these stocks. Sustainable development essentially seeks to reduce the inter-generational inequality by trying to ensure that the current generation does not consume and exhaust too much resources such that future generations have a difficult time leading a life that is at least as good as that of the current generation. An important macroeconomic feature of any economy is the extent of inequality in income and wealth experienced by the current generation. The intra-generational inequalities constrain sustainable development. Few who have too much can afford to waste resources and still bequest adequate amounts to their next generation. The many who have too few often struggle and overuse resources, thereby creating environmental stress. The more the inequalities the greater the environmental problems, making sustainable development that much more difficult. Hence, the reduction of intra-generational inequalities is an integral part of the process of sustainable development. This paper discusses some views on inequalities and the alternative ways of addressing a key problem of sustainable development.

Keywords Sustainable development · Inequality · Intra-generational inequality · Inter-generational inequality · Crude look at the whole · Radical ecology

1 Introduction

Sustainable development is a complex concept. It draws on a number of branches of knowledge beginning with the physical and biological sciences, to ethics and political economy, down to the details of economic policy formulations and the tricky challenge of forecasting the future. In all these distinct aspects that converge to a

A. K. Sinha (✉)
Heritage Business School, Kolkata, India
e-mail: aksinha@iimcal.ac.in

Indian Institute of Management Calcutta, Kolkata, India

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_21

knotty problem, the concept of economic inequality is fundamental in defining the core issue of sustainable development. Arguably, the most widely used definition is the one that is attributed to the Brundtland Commission Report titled Our Common Future (1987) that described sustainable development as development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It is obvious from this definition that sustainable development has to do with inter-generational inequality. The current generation may not 'eat up' productive resources that leaves the future generations unable to meet their needs and live like we do. The current generation then has a decision problem to use resources so that everyone living now may have a decent life, while leaving enough resources for the future. Would inequality in the current generation have any bearing on this? It would. The very rich may have a tendency to waste resources since they have enough and more to spare. On the other hand, the very poor may have to depend on natural resources like land and forests and rivers where overuse of these resources could despoil nature and leave a less productive environment for future generations. Since nature is shared by all living beings and the natural environment is commonly owned, any sustainable solution has to be global. It is impossible to have sustainable development in one part of the world and a terrible despoliation of the environment in another part. This is yet another form of equality that sustainability entails.

In both cases, extreme wealth and extreme poverty reduce resources left for posterity. Less economic inequality would save more resources and contribute to a more sustainable future. It can also be easily appreciated that inequality is to be found at two interrelated levels—at the level of individuals within a nation and inequality across nations. This is not to claim that the problem is only about inequality. There are issues in defining 'needs' of a particular generation. Similarly, there could be differences in understanding what exactly are the needs to be sustained; incomes or wealth, or utilities or nature itself which provides us with food and about everything else that make up the world. However, tackling inequality becomes the core of any strategy of development that attempts to address issues of sustaining economic development over time and across generations.

In this paper, we first take a look at how sustainable development can be defined in an operational form where decision-making would be possible. Next, we take a look at using those definitions and indicate how dynamic decision-making may be quite thorny and our attitude to inequality come into play. Finally, we discuss what kinds of changes are necessary to make human society sustainable not only for people inhabiting the planet, but also for all living beings.

2 Sustainable Development as Non-decreasing Comprehensive Wealth

What is to be sustained such that inter-generational well-being is non-decreasing? One widely accepted approach is to ensure that society's stock of wealth is non-diminishing (Dasgupta 2001) so that at least an equivalent stream of income can be generated from it. Capital is supposed to be substitutable so that if one kind of machine or material could be substituted for another. Similarly, knowledge could be replaced too, say from using oil to drive a car, to nuclear energy powered electric batteries. Ecologists and scientists would be quick to point out that not all capital is fungible as economists quite often presuppose (Neumeyer 1999). One can hardly think of substituting fresh water, or clean air, or the fertility of the top soil, or the cyanobacteria that form the base of oceanic food webs. Hence, one may think of a set of substitutable economic and social capital (institutions, rules of functioning) and a distinct non-substitutable set of natural capital. Sustainable development would imply a non-decreasing stock of *both* the sets of capital.

Mere bequests of a stock of non-decreasing capital, including natural capital, is clearly not enough. All these forms of capital could be maintained while having a terribly unequal distribution of power and wealth where political or corporate elites might keep wealth to themselves holding down a dominated and powerless populace. In such a case of absence of freedom, the provision of basic needs for everybody of the current generation would be unattained. Hence, the best way to view sustainability from an anthropocentric perspective is to ensure a non-diminishing measure of human well-being that includes not only income and wealth, but also basic capabilities such as health, education, political voice, natural capital, and the freedom to choose one's lifestyle from an expanding set of functionings. Intra-generational equality is as important as inter-generational equality for development to be made sustainable.

Sustainable development is distinguished from the usual considerations of economic development in terms of ensuring that development is not just a one-off change in the state of affairs of a society. Rather, it has to be seen as a *process* that can be replicated over time and space for future generations of people who will inhabit the earth (Sinha 2012, 2013; Martini 2012). Indeed, sustainable development is essentially a critique of thinking about development as mere economic growth accompanied by improvements in the average income and standard of living attained by a given population. It is more about a fair distribution of resources and access to productive resources, across generations, keeping within the bounds of the planetary natural resource constraints.

Acceptable as it may seem at first glance, the social solution may be difficult to arrive at, and even more difficult to implement. First of all, it entails a basic understanding of the role of Nature in the process of economic activities, and especially the kinds of constraints this role might throw up. The second problem is agreeing about what exactly is to be bequeathed to future generations. Would it be some subjective notion of utility or satisfaction measured with the help of a social welfare function, or some more tractable concept of well-being? The third issue is about how much importance we attach to the well-being of future generations of yet unborn people (Koopmans 1960). Does one treat them to be as important as we consider ourselves, or does one discount their well-being only because they are born at a later date? Finally, depending on how one resolves the first three questions, one has to agree upon a time path of resource allocation and the resultant social well-being from those resources accruing to successive generations.

In solving the dynamic resource allocation problem, there are bound to be many complications too. People living in a society normally prefer to consume things in the present time as opposed to waiting for the future. This is referred to as time preference or impatience and is usually considered in economic theory to be a subjective choice exercised by the individual. How would the social rate of time preference be chosen? Is there a well-defined method of arriving at this number? The social planner chooses some number that could be low implying that society (assumed to live forever) would consider today's consumption as important as tomorrow's consumption (consumption being taken as a rough and ready measure of social well-being). On the other hand, a large amount of poverty and current levels of material deprivation might compel a planner to treat the consumption in the here and now to be more important than the future, especially the distant future. An implication of this would mean having more to consume now. The opportunity of productive investment, on the other hand, implies that if one sacrifices and saves for the future, the action will be rewarded with a positive rate of return on investment, usually captured in the rate of interest as the reward for waiting. This, in turn, would imply society would try to consume a little less now and save for a future attracted by the rewards earned from saving. Finally, the nature of the social well-being function with some reasonable properties such as diminishing marginal gains would imply a smooth distribution of income over time. This is referred to in economic theory as the 'aversion' to inequality. If marginal gains are diminishing, then a tiny amount of consumption (again taken as a rough indicator of social well-being) withdrawn from the future and re-allocated to the present would imply that the loss (of future consumption) would be less than the gain (in current consumption). The dynamic decision problem could exert pulls in three different directions. Society's choice would depend on the *combination* of values chosen for the allocation of resources over time with the values of time preference and the aversion to inequality playing a critical role in determining dynamic outcomes.

There are complicated ethical choices too, even when an individual allocates personal resources over time, say the lifetime of the individual, and the terminal bequests to be left for the next generation. When it comes to an entire society taking a decision to ensure that development is sustainable in the sense of inter-generational well-being is chosen in such a fashion that it is non-decreasing (Dasgupta 2001), the complexity is compounded. How does society ethically choose a set of numbers for planning its allocation of resources as a representative of a large set of individuals each of whom has a particular ethical ranking of states of affairs (current and in the future)? Obviously, the role of the public policy planner becomes complicated. For instance, it would of great convenience for the planner, if she knew that an overwhelmingly large number of persons (whom she is supposed to represent) had a strong ethical preference for guaranteeing a non-diminishing level of well-being for future generations, even if it entails making current sacrifices for the purpose. The complexity of sustainable development does not stop there. Even if all these issues could be resolved or a consensus arrived at, the pathway to sustainability would entail some sacrifices to be made by the current generation of people living on the planet.

What would be the quantum of this sacrifice and how would the cost of making the sacrifice be shared across individuals and nations.

3 Creating Public Policies from Scientific and Economic Knowledge

In the nineteenth and the twentieth centuries, science was understood to be exact knowledge with its laws and precise predictions. Scientific knowledge has increased enormously and has the ability to indicate the possibility of problems and disasters like the effects of increasing pollution, or the effects of biodiversity loss. However, with the rise of knowledge two things changed. The first is the ability to view reality as a more complex system where the sub-systems cannot be broken down as in linear systems. The sub-systems interact with one another and affect the system as a whole. The second aspect of rising knowledge is the ability to view future possibilities but only as possibilities where objective probabilities cannot be assigned. Yet in many cases (for instance, climate change), the future possibilities are understood with deep uncertainty, where predictions about time frames, extent of damage or costs are all subject to large degrees of error. Science, in this sense, provides fuzzy knowledge. On the other hand, if society does not do anything to mitigate future problems, then the costs can turn out to be unacceptably high, even endangering life on earth. Society has to make an ethical choice, and if it needs to act, the action has to be, comprehensive, decisive, and with shared global responsibility. Social responses cannot afford to be hazy and divergent beyond a narrow band. In the past two centuries, it was just the opposite. While science was precise, the social responses were subjective and diverse. In the current era of what is now described as post-normal science, public policy must be generated from a constant inter-action with scientific knowledge as it emerges (however fuzzy it might be), and construct social choices that are feasible.

We are normally used to breaking up complex systems into more manageable simpler sub-systems and studying them to understand the whole. However, in non-linear complex systems it is difficult since the sub-systems are constantly interacting with each other affecting the system itself. When contemplating about the future of human society long term issues such as economic well being, health, political systems, military and diplomatic issues, environment and climate; all are considered important and all are interrelated. All these issues go beyond the individual or the nation state into global or planetary dimensions. These are not decomposable into isolated silos that can be studied by the specialist. This world *problematique* of the twenty-first century needs a more comprehensive knowledge that is not in the ken of any one branch of specialization. Hence, the importance of inter-disciplinarity which starts with a crude look at the whole sometimes referred to as CLAW (Miller 2016). This requires a different approach to understanding the planet and its inhabitants. The geo-sphere is important in the sense that an understanding of it gives us the way important minerals and resources such as water or air or soil are formed and what life support systems signify. The bio-sphere helps us understand the web of life and its intricate interconnectedness, where human beings are just one species out of innumerable ones. The bio-sphere affects the geo-sphere as the soil's fertility is maintained by bio-geo-chemical cycles like the carbon, phosphorus and nitrogen cycles. Geo-sphere can modify the bio-sphere by changing habitat conditions or forcing migration as terrains and climate, along with the availability of food and water alter. Finally, there is the anthropo-sphere which includes human society and all activities around it. This is supposedly the most disruptive as we have changed nature to a very large extent and our use (or as most now believe) overuse of natural resources, and damage done to the environment has had irreversible effects, on both the geo-sphere and the bio-sphere.

If we are to ensure that we can sustain life and human society on earth in the complex system we need to preserve, we need to bestow a non-decreasing comprehensive wealth to future generations. The planning problem is theoretically solvable, but in reality it will need a number of changes in how we do things and above all, how we think about things. For instance, as described earlier we are not used to thinking about history as the history of the planet where human beings are just one critical cog in a giant machine. We think of history as a narrative of humans on earth and consider our domination as limitless and glorious. We are not trained to think in terms of complex systems. We break complexity into simpler parts. That strategy, we argued, is likely to give us erroneous understanding of the whole. However, a crude look at the whole of what sustainability entails gives us some actionable points as it were for societies to engage in with debate and discussions that could inform public policy.

The first concern is that we need to think about ways and means to reduce inequality over time but also in the here and now. Yet we know that eradication of poverty and deprivation in the current generation will require growth and development. The natural way for this is to adopt the business as usual path that has led us into the morass of environmental degradation. We do not know how, and what cost we can decouple growth from high energy use and resource extractions. This is a big challenge that will require disruptive changes. Population growth and demographic transitions is the second area of concern as far as public policy is concerned. How do we tackle the problem of too many people on the planet given technology and the institutions that govern our lives? The earth's carrying capacity has been stretched where scientists believe we are using not only the renewable interest of the environmental capital, and we are also diminishing the stock. However, can we change lifestyles so that we consume less? Who should consume less? Which people are surplus? However, changing population growth patterns and rates is tricky since these decisions are private while the adverse impact is one that hurts all. If economic development reduces population growth rates, as has been observed for the societies that are now materially affluent, then economic growth in the poor nations of the world today will further despoil the environment and make development unsustainable. We yet do not have an alternative model of development.

The third area of concern is that we need a social and institutional transformation that will be more concerned with the well-being of all and can cope with conflict.

New attitudes must also yield some national space to global forums that can make sustainability concerns consistent across nations and making some form of compliance mandatory. This leads us to the fourth area of concern. Social and institutional changes would be impossible without an ideological change where myopic self-centred interests are reconciled with a bigger planetary consciousness which informs our decisions and the choices we make.

The final area of concern, which is arguably the easiest to address, is the need for a new way of looking at technological innovations that reduce materiality of production and help the transition to a low carbon economy, and a set of economic policies that quickly identifies and addresses the internalization of external costs that damage the natural environment. Improvements along this last area of concern are already discernable. These concerns are not independent of each other.

4 Some Alternative Pathways to Reduce Inequality

The areas of concern we discussed above have different proponents who prioritize importance of these areas differently. In a broad sense, these different approaches can be traced to those who put primacy of technology and the human genius, those who put emphasis on ethics and values that would trigger and facilitate the transformations necessary, and finally the rational school that believes that humans will perforce realize that the most wise thing to do is save the planet since without that their own existence could be in jeopardy.

4.1 Technological Primacy

There is widespread belief, especially amongst social scientists who do not actually design new technologies that scientific changes will help decouple growth from higher energy intensity and higher use of material resources. The current problems of environmental degradation, biodiversity loss and climate change are a necessary phase that human society must perforce pass before a greener sustainable world is built. Indeed, there have been a lot of technological changes that are greener and more ecologically efficient. However, by the end of the century primary energy use is predicted to grow threefold. If this turns out to be true, would it lead to a lower use of material inputs in production? Over history from the hunter-gatherer stage to agriculture to the modern industrial phase, this has never actually occurred. Materiality of production and waste creation has systematically increased, sometimes dramatically rapidly as in the past two hundred years. History at least does not support this expectation of a turnaround in the impact of technology. A purely technology led solution may be ruled out. So much of the current world economic order is based on growth where even the rich nations get enormously worried if growth declines. The emerging economies are almost exclusively focussed on the robustness of their growth path and poor nations quite rightly so aspire to grow out of widespread poverty and deprivation.

4.2 Ethics and Values

A second area of the debate revolves round the primacy of values, justice, rights and the natural world. Here, the argument is that human beings have to first build their lives around a cultural identity with nature and other living beings, and this must be reflected in the political and ethical systems of society, in the construction of social order, in religious understandings and spiritual relationships. This will lead to a transformation that will make the world more sustainable and less unequal. In this view, there are non-technological limits to growth when growth impinges upon the rights of others, including other living beings, whether now or in the future and whether the problems arise nearby or in a far away geography. This culture, in short, is expected to be universal and shared by all. It clearly calls for a revolution in lifestyles and values that may be difficult to attain, especially in the short span of a few decades. There are some problems with this approach which explains why it might take more time than expected. Human beings have for centuries put material comfort and well-being as far more important than the mental state they live in. A materially poor life where the person experiencing the poverty is happy is indeed rare. It might happen for a few people especially in religious settings of a hermitical existence, but it is very difficult to think of the overwhelming majority of people to be transformed in any short span of time. The second problem is that if this was to be engineered through social policy or education, there would be a clash between these attempts and a fundamental human value of freedom especially the individual's freedom to choose.

4.3 Rational Choice

A third approach in the debate about what is to be done is the reliance on rationality resembling the enlightened human intellect. This is particularly strong in European and Western thought where there is a deep belief that the rational side of human thought would resolve the complications of values and ethics in the construction of identity. However, rational thought has seen the development of universalistic monolithic institutions of capitalism which dominate the economic and technological structures. This has been referred to as the totalitarian aspect of enlightenment. There is a hope that enlightened self-interest will make the human mind understand the lock-in and help transcend that condition, driven by human agency and will power. Science and scientific evidence play a role here. The warning is sounded based on the new evidence that science has brought forth in the last two hundred years. It is only recently we know why the dinosaurs became extinct, how old the earth is, about ice ages, about where the sun obtains its energy from and how chemistry works. There was no knowledge about the biology of genes and epigenetic theories. There was no theory of evolution and the big history of the planet was unknown. This new knowledge rather than a faith in technology or values can be demonstrated to convince the rational human mind that something has to change and change pretty fast if a disaster has to be avoided.

This expectation is also doubtful. Whether this evidence will actually help transcend the older beliefs about the world, ideologies, and above all selfish short term material interests is questionable. The bigger challenge is how to integrate this knowledge of science (however imprecise and probabilistic) into slow-moving societal belief systems and trigger a faster than usual response in terms of changed actions and preferences.

4.4 A Crude Look at the whole

It is important to begin with a crude look at the whole where science and culture have to come together to find solutions and trigger changes. This can be done by looking at earth system processes which is developing as a scientific, networked, collective, social global subject that is attempting to order its understanding of the world with all its complexities. The features of this new knowledge building attempts include a highly developed comprehensive earth systems science using complex computer simulation as an important tool for making projections of the geo-sphere bio-sphere and anthropo-sphere and their interactions. A second key feature is a giant earth observation system which provides empirical links between the earth's past and future. Finally, there is a globally networked system of communication, negotiation and goal setting that is essential to manage the large number of processes in a consistent manner. Together these features might contribute towards the building of a global consciousness that might steer policy and action towards sustainability. If, however, time is of the essence, then waiting for socially acceptable change may take too long and some catastrophic event might set in even earlier than that. Therefore, it might be necessary to provide active management or engineering to bring certain changes that can trigger others. For instance, an emphasis on new technology and the reduced use of energy in production and consumption, or the creation of new institutions for managing the environment can easily be started through conscious efforts without waiting for a global consensus. Such changes must obviously pay attention to design, but the designs themselves must be deeply embedded in a value-based analysis when weighing the consequences and adverse effects of these changes.

As far as changing the way we think about human societies and their relationship to the planet, we need to create and develop what is being referred to as the Big History of the planet (Christian 2004; Harari 2015; Spier 2005). This big history project puts together three time scales through which we normally view the earth and its inhabitants and puts them together. The first is the very large geological–astronomical scale of time followed by the biological time scale of life on earth and finally the human time scale when human beings evolved from the big apes. This would help in a more nuanced understanding how human beings along with other forms of life interact with the environment and how the environment itself evolves. The construction of such views is not without their dangers. Societal visions of nature and the natural have often clouded human interactions with the natural environment. It requires capable science that is embedded in culture and deeply reflective. This is enormously significant since projections of the past as well as the future are fuzzy and imprecise while our knowledge of ecological systems is also incomplete. Any radical changes in technology will require fundamental social transformations too. There are also large differences in the underlying assumptions about human behaviour that historians, economists, sociologists and anthropologists make. Yet the hope is that with the crude look at the whole humanity will overcome the fragmentary perceptions of the world and be able to construct mental images where the whole is more than the isolated parts of our knowledge.

4.5 Radical Ecology

There is another group of thinkers who have been arguing that private-property-based market capitalism would have to be replaced by a new set of institutions which would reflect an alternative socio-economic order of things (Foster 2002, 2009). Proponents of this radical ecology assert that the fundamental dynamic of capitalism is to accumulate wealth and increase consumption. It works on the principle of self-interest, and not social considerations. Hence, the time horizon of all self-interested decision-making that is hardwired into the capitalist system would itself be an insurmountable challenge. This is definitely true for households and individuals. As far as businesses are concerned, they might have a planning horizon which is greater than the life span of an individual and are usually guided by the current situation of competition and demand in the market. In essence, any form of private property, self-interest-based capitalism would usually place a higher value on the present than on the distant future. It has a built-in tendency to create inequalities in outcomes as far as wealth and income are concerned. High levels of inequality would generate tendencies to waste by the rich and over exploitation of the environment by the very poor. Putting it starkly, the philosophy of capitalism does not contain in its vocabulary the word 'enough'.

Some form of planning would be necessary for reducing extreme inequalities which trigger unsustainable use of resources. This planning need not imply large government or big bureaucracy; rather, it would entail a large set of guidelines for resource use and business processes. Further, it would be essential for local communities and businesses to adhere to these guidelines for which appropriate incentive mechanisms must be put in place. The capitalist system of production and exchange is a continuous reworking of society's relationship with the biophysical world. This process of continuous transformation degrades the natural environment through depletion of resources and creation of wastes—referred to as the metabolic rift. No system that is essentially based on private property can reverse this damage done to nature.

4.6 Marx and Nature

The classical Marxist critique of capitalism was built around the inherent instability of capitalist production systems and the consequent struggle between capitalists and workers as two distinct classes. Marx and Engels (Engels 1940; Marx 1959/1844; Parsons 1977) were well aware that capitalism necessarily despoils nature, but focussed on the primary social conflict between two classes of people. Radical ecologists (Jones 2011) point out that in the current stage of capitalist development, the primary contradiction lies in the relationship of capitalism's logic of change and nature's own rules of the biophysical world. The radical ecologists therefore believe that it is unlikely that capitalism can be reformed to deal with its relationship with the natural world such that a more harmonious evolution could take place. This is particularly so because congruence with nature requires a different type of property system which would be governed by a complex and multi-layered system of usufruct rights, customs and obligations where nature would be looked at as shared common property. This sharing would have to be understood not only as one amongst human beings but amongst all other living beings too.

If one has to argue that capitalism cannot be reformed enough to define a sustainable relationship with nature, then new institutions and property relations would have to be defined and created. Such radical social changes would not come about without opposition from certain sections of business that stand to gain by utilizing nature for private profit. Some Marxists (Gramsci 1971) have argued that the state and the capitalist class, even in the face of economic crisis, can control and mitigate opposition from the working class. This is typically done, partly through coercion and use of force, and partly through a more subtle creation of consent, for the prevalent order of things. In the case of the contradiction of capitalism's inherent logic of accumulation with the degradation and depletion of nature, there is no way that consent can be manufactured in nature. If the relationship with nature becomes essentially one of coercion, then, as is becoming increasingly obvious to scientists, nature can be quite cruel in its reaction and retaliation. Indeed, human beings living in society are not fully aware of the kind of severity with which nature might hit back. The crux of this argument is that there are no feedback and automatic corrective mechanisms in the capital–nature relationship unlike the capital–labour contradiction.

4.7 Inequality as the First Priority to be Addressed

The first task when thinking about sustainability as a systemic problem that is hard to break down into fragmented parts is the need to tackle inequality. A widely accepted solution for inequality is economic growth, but that might directly conflict with sustainable development, at least the type of growth that informs the imagery of development—energy-intensive material affluence. Economic growth also promotes mobility and enlarges the ecological footprints of individuals as well as communities. Economic growth alone, in the age of global mobility, does not necessarily reduce inequality and make the world more sustainable on its own. There are a host of other social policies required to redistribute the fruits of growth in a more equitable fashion. Moreover, economic globalization has eroded the power of nation states to handle these social policies in as effective a manner as before. It is also true that the environmental setting in which economic growth is to take place (as at least a necessary condition for reducing inequality) is already quite alarming. In the United Nations Millennium Ecosystem Assessment (2005) was the first check on global health of nature. It estimated that 60 % of the planets ecosystem services have been degraded. These services are not only fundamentally important for human well-being, but are critical for poor communities and a key feature of the planets capacity to adapt to climate change. In the absence of concerted global action to fight inequality and its associated ills, sustainable development will remain on a collision course with business as usual and the pattern of growth that emerges from it.

In the absence of the transformations we have discussed, especially the reduction of contemporaneous inequality, there could arise environmental crises sooner than many scientists expect. In that event, two distinctly different social responses could come about. In the past, there have been instances when big crises have united people and nations by creating a common reason for action. Effective leadership has often been able to engage a larger population to greater social trust and collective action like the Great Depression did in the making of the welfare state. More recent events, such as global health problems of, pandemics have fostered much greater cooperation and action than ever before in fighting disease. However, crisis can lead to conservative responses too. International migration has often been perceived as a threat to the local populace and that has led to insularity and exclusion and often increased racism.

5 Conclusion

We have argued that inequality is germane to understanding sustainable development, and its reduction is of paramount importance in creating a pathway to sustainability. However, to be able to do that we require a more nuanced analysis of sustainability as a complex system which cannot be usefully broken up into independent parts. Unless it is understood that the planetary system is bigger than its constituent parts, any effort to reduce inequality and move to a sustainable world will be well-nigh impossible. We have argued that a nuanced understanding of the earth system has to begin with a crude look at the whole and realize that interactions between the environment, society and ecology are constantly taking place. To change and transform the business as usual model of growth and development, we must understand the links between the geo-sphere, bio-sphere and the anthropo-sphere. Once we understand that (however imperfectly) we need a series of radical transformations that entail better management of the environment, better understanding of science and technology as well as a far better understanding of how to change social institutions and belief systems. They need to change from being parochial and local to one where there is consciousness of the large framework of planetary history. Social change based on fuzzy predictions of science requires technological change, changes in value systems and an open approach to observing and analysing scientific evidence and projections. These three pathways of transformation would serve to reduce inequality as an essential element of moving towards sustainability. The core of that challenge is to address both inter-generational inequality and intra-generational inequality. Attempting both simultaneously is a big challenge since the instruments of attaining them may conflict with one another. Economists know this as the trade-off between efficiency and equity. In the arena of sustainable development, the trade-off is between justice now and justice later. It is more complex and more planetary in its aspects than economics as a discipline can handle by itself.

References

Christian, D. (2004). *Maps of time*. Sacramento: University of California Press.

Dasgupta, P. (2001). *Human wellbeing and the natural environment*. New Delhi: Oxford University Press.

Engels, F. (1940). *Dialectics of nature*. Moscow: Progress Publishers.

Foster, B. J. (2002). *Ecology against capitalism*. New York: Monthly Review Press.

Foster, B. J. (2009). *The ecological revolution: Making peace with the planet*. New York: Monthly Review Press.

Gramsci, A. (1971). *Selections from the prison notebooks*. New York: International Publishers.

Harari, Y. N. (2015). *Sapiens: A brief history of humankind.* Penguin Books.

Jones, A. W. (2011). Solving the ecological problems of capitalism: Capitalist and socialist possibilities. *Organization & Environment, 24*(1), 54–73.

Koopmans, T. (1960). Stationary ordinal utility and impatience. *Econometrica, 28,* 287–309.

Martini, V. (2012). *Economic theory and sustainable development: What can we preserve for future generations?* London: Routledge.

Marx, K. (1959/1844). *Economic and philosophical manuscripts of 1844.* Moscow: Progress Publishers.

Miller, J. H. (2016): A *crude look at the whole: The science of complex systems in business, life and society*. Basic Books.

Neumayer, E. (1999). *Weak versus Strong Sustainability: Exploring the limits of two opposing paradigms*. Northampton MA: Edward Elgar.

Parsons, H. L. (Ed.). (1977). *Marx and Engels on ecology*. Westport, CT: Greenwood Press.

Sinha, A. (2012). Sustainable development and governance: Complex processes and unpredictable outcomes. *Society and Management Review, 1*(1).

Sinha, A. (2013). Sustainability: Ethics and the future. *Journal of Human Values, 19*(2), 113–126.

Spier, F. (2005). *Big history and the future of humanity.* New York: Wiley.

United Nations. (2005). *Millenium ecosystem assessment.*

World Commission on Environment and Development. (1987). *Our common future*. Oxford: Oxford University Press.

Informality and Disaster Vulnerability: A Preliminary Exploration

Lopamudra Banerjee and Snehashish Bhattacharya

Abstract In this paper we argue that the vulnerability that arises from structural conditions of production and employment in the informal economy may also exacerbate the conditions of disaster vulnerability if and when an extreme natural event occurs. We further argue that given the heterogeneity within the informal spaces of production and exchange, heterogeneity may also arise in disaster experiences of people operating in these spaces. More specifically, the risk of experiencing exposure to adverse disaster conditions is reduced for people that are advantaged in the social relations of production and employment. We illustrate these arguments with evidences from Indonesia, a country where the informal sector is a major presence in the economy and natural disasters are frequent occurrences. We draw upon the third and fourth waves of the Indonesian Family Life Survey datasets for our study. In the light of our findings, we argue that 'normal' or everyday conditions of economic and social lives of people engaged in informal work formatively affect, and may even determine, their incidental vulnerability to disaster shocks. We maintain that extreme natural events get actualized as disaster phenomena only in the presence of specific social and structural phenomena. We posit, furthermore, that control over productive resources as well as over labor processes may play a fundamentally important role towards reducing people's disaster vulnerability.

The original version of this chapter was revised: The incorrect abstract has been replaced with the updated one. The correction to this chapter is available at https://doi.org/10.1007/978-981-13-9981-7_26

L. Banerjee (✉)
Bennington College, Vermont, USA
e-mail: lbanerjee@bennington.edu

S. Bhattacharya
South Asian University, New Delhi, India
e-mail: snehashish@econ.sau.ac.in

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_22

1 Introduction

The working populations engaged in the informal economies are vulnerable in a dual sense of the term. One form of vulnerability arises from what Bourdieu (1998) saw as *précarité* in everyday life of informal labor; the other form haunts periodically and incidentally, as people operating in this economy experience unexpected disruptions in their normal economic activities. It is plausible that the endogenous structural conditions of employment that make people vulnerable in their everyday economic life would also shape their vulnerability to sudden occurrences of adverse exogenous shocks. Our paper explores this idea.

We consider three propositions here. The first is in regard to how *everyday* vulnerability may *manifest* in normal economic conditions of people with informal employment. It is well recognized in the literature that the nature and extent of vulnerability are different from that experienced by people with formal and secured conditions of employment (Kalleberg 2009). But, within the informal economy, there may be further differences in people's experiences. This heterogeneity may surface since the informal economy, in itself, is a heterogeneous entity (Chakrabarti 2016). There are multiple segments within this economy, each with its own operational structures, economic opportunities, and earning possibilities. Accordingly, a heterogeneity in living standards is likely for people in various strata of an informal workforce. We pay specific attention to people's consumption expenditure and their asset conditions. We make the rather unexceptional assumption that consumption is an important indicator of a person's ordinary material quality of life, and her asset conditions can signal her relative prospects for economic mobility. Assets also indicate her relative ability to deal with any unexpected disruptions in livelihood conditions. We posit: differences will exist in the average level of consumption and asset ownership across various strata of informal working population. The differences will not be just a matter of individual characteristics of agents but are shaped by systemic factors.

Our second and third propositions in the paper are in regard to vulnerability that is *incidental* to unexpected occurrences of adverse events. Our specific interest is in covariate shocks; and, as shock-generating events, we consider natural extremes like floods, droughts, or earthquakes for our study. While these events may be exogenously generated and non-discriminative, if there is segmentation within the informal economy, then, it is possible that the relative risk of experiencing "exposure" may be different across different strata of the informal workforce. We posit: the relative risks will be higher for people in certain stratum of an informal workforce that are also in positions of greater vulnerability in everyday life given their livelihood conditions. In an extension of our second proposition, we posit, third: the outcomes of "exposure" to extreme events in people's economic lives will also be different for different strata of the informal working population. We assess the adverse outcomes as deviations in consumption levels for the "exposed" people from the pre-shock to the post-shock period.

We explore our propositions in terms of a household survey dataset from Indonesia, a country where informal workforce is a major presence in the economy and where covariate shocks like extreme events in nature are recurrent occurrences. We base our study on the third (2000) and fourth (2007) waves of the Indonesian Family Life Survey (henceforth, IFLS3 and IFLS4, respectively). We purposefully consider a household (and not a labor market- or firm-level) panel survey, as the data allows us to associate people's initial employment conditions in informal sector with their standards of living in their regular everyday lives and as well as when they experience shock of exposure to natural extremes.

We structure our paper thus. In Sects. 2 and 3, we set the context of our study. In Sect. 2, we elaborate on our notions of *everyday* and *incidental* forms of vulnerability; and, in Sect. 3, we describe the segmentation within an informal economy and postulate why vulnerability conditions may differ among people operating in these various segments. The evidential support of our arguments is presented in next three sections of this paper. In Sect. 4, we introduce our data. In Sect. 5, we concentrate on our first proposition and examine the heterogeneity in consumption expenditures and asset conditions across different segments of our studied informal economy. In Sect. 6, we concentrate on our second and third propositions. We start by examining if and how the relative risk of experiencing the shock of exposure to natural extremes may vary across the strata of informal working population. Then, we examine if post-shock deviations in consumption differ across these strata. We summarize our conclusions in Sect. 7 of the paper.

2 Informal Work and Vulnerability

Broadly defined, an informal economy is an economic space where production and sale of commodities are carried out by households (or individuals) that own unincorporated private enterprises, which operate on partnership and proprietary basis and employ less than ten workers in total (NCEUS 2007). The distinctive feature of such enterprises in this economy is that their owners themselves usually participate in the labor process along with one or more family members. An extensive body of evidences has now established that the economic life of people engaged in informal work is marred by an importunate riskiness in their livelihood conditions.[1] "Vulnerability" is invoked in this literature as a catch-all term to capture this lived experience of riskiness for people.[2]

We imagine two aspects of this vulnerability. The *everyday* form of vulnerability indicates a situation of living and working without stability or safety net. The situation is endogenously generated for people engaged in any form of informal enterprise, given the relative insecurity of livelihood conditions, insufficient and ever-fluctuating income from informal work, and the lack of welfare protection and insurance against joblessness and income fluctuations. In the industrialized world of the Global North, this form of vulnerability is associated with the absence of so-called standard employment relations and the retreat of the welfare state (Kalleberg

[1] See Cook and Kabeer (2010) for a discussion on the issue in the specific context of Asian economies.
[2] See, for example, the discussions in Chambers (2006).

and Hewison 2013). The vulnerability conditions are further exacerbated as rights and entitlements of informal workers are eroded away and they lose control over the labor process (Cranford et al. 2003). In the less-industrialized Global South, *everyday* forms of vulnerability are the *normalized* state of being which impedes people from maintaining a minimum acceptable quality of life and constraints their economic mobility.[3] In some studies, the question of *everyday* forms of vulnerability is linked to whether or not people are able to earn an above-poverty level of income given their occupation and employment status (Webster et al. 2008). In other studies, *everyday* vulnerability is described in terms of (lack of) economic opportunities. A main thesis explored in these studies is that the people with unstable, insecure, and uncertain conditions of livelihood are also poorer in terms of their prospects to attain a higher level of consumption, asset formation, or income mobility than the people with stable and secured conditions of employment (Agarwala 2013).

Together with this *everyday* form of vulnerability, the working population in an informal economy also experiences vulnerability of another form periodically and unwittingly. This vulnerability is incidental to the unexpected occurrences of exogenous shocks that can throw into disarray the economic activities of many informal firms at the same time and de-arrange income flows of many informal workers simultaneously and in a sudden manner. Extreme events in nature (for example floods, droughts or earthquakes) are often cited as examples of such covariate shocks. The riskiness in livelihood conditions that arise in periods of these covariate shocks may be identified as *incidental* form of vulnerability.

While people with secured and stable conditions of employment may also, of course, experience exposure to extreme events in nature, empirical evidences suggest that people with more precarious conditions of livelihood are at greater risks of experiencing, first, exposure to shock-generating events, and second, adverse outcomes of exposure on consumption levels and other indicators of material wellbeing. A number of reasons have been explored to explain this difference. In one strand of the literature, people's livelihood conditions have been associated with their relative access to material and nonmaterial resources that enable them to fend off shock-generating events. Wisner et al (2004), for example, note that vulnerability in times of natural disasters is greater when people lack the wherewithal to prepare for and cope with these adverse events, and, that the relative lack of resource access can be traced back to the initial conditions of their employment. Similar arguments are also presented in Cutter (2006), Hewitt (1997) and Tierney (2007). Other strands in the literature have associated people's occupation and employment status to their access to social network and social capital (Adger 2006; Woolcock and Narayan 2000). It is argued that the people with precarious employment in their everyday life also have less durable network of access to information and power, which, in turn, constraints their effort to deal adequately with unexpected fluctuations in livelihood conditions in times of exigencies.

---

[3]The International Labour Organization (2006:1), for example, estimates that "in 2005, 84 percent of workers in South Asia, 58 percent in South-East Asia, 47 percent in East Asia… did not earn enough to lift themselves and their families above the US$2 a day per person poverty line."

While "vulnerability" is an inexact concept and does not render itself easily to empirical analysis, it is not difficult to imagine that *everyday* vulnerability in employment would leave a mark on people's normal consumption and asset conditions. It is also conceivable that *everyday* vulnerability would mold people's relative resiliency in exceptional times of covariate shocks and forge the ways in which *incidental* form of vulnerability manifests. We base our study in this paper on these premises.

3 Heterogeneity Within the Informal Economy

Neither the manifestations of *everyday* vulnerability in people's normal conditions of living nor the displays of *incidental* vulnerability in exceptional times of covariate shocks may, however, be uniform across all agents operating within the informal sector. The variations may arise since the informal economy is in itself a heterogeneous entity. Within the mainstream understanding, the informal economy has been conceptualized as a complex structure segmented into "upper" and "lower" tiers (Fields 2005). The "upper tier" comprises of owners of informal enterprises, all of whom are, by definition, self-employed, and some of them may employ hired wageworkers in their enterprises, while others do not. The "lower tier", in contrast, comprises of informal workers, who are all employees within the informal enterprises, and many of whom are hired against wages, while others are family workers of enterprise owners and work under income-sharing arrangements. It is estimated that 35–85% of non-agricultural employment in Asia consists of this "lower-tier" informal workforce, while the corresponding numbers for Africa is 40–97%, and for the Latin America-Caribbean region 30–75%.[4]

A fuller perspective on segmentation of an informal working population is available when we examine the social relations of production and employment within informal enterprises. These relations can be derived from two basic criteria: ownership of the enterprise, particularly the means of production (Wright 1978), and control over the labor power of others, i.e., control of the processes of production, appropriation, and distribution of surplus (Resnick and Wolff 2006). The twin criteria generate a fourfold stratification within the informal workforce. Based on the first criterion, we may separate the owners of informal enterprises from the informal workers. This preliminary stratification corresponds well with the segmentation of informal workforce into "upper" and "lower" tiers. Based on the second criterion, we may further stratify the agents operating in the "upper" tier into a class of self-employed informal petty producers (and traders) who do not employ any hired wageworkers but may run their enterprises with one or more family members, and a class of owners of informal establishments who do. In the absence of capital-wage labor relations, the petty producers seem to operate in a *noncapitalist* space of economic activities. The informal establishments, in contrast, are more akin to *capitalist* micro-entrepreneurial enterprises (Bhattacharya 2017; Bhattacharya and Kesar 2018). Differences in employment relations within an informal enterprise also

[4]See the discussions in Sudarshan and Unni (2003), Blunch et al (2001) and Fields (2005).

allow us to separate out the informal workers in the "lower" tier further into two strata—workers hired against wages in informal establishments, and, the unpaid family workers who work under income-sharing arrangements in the enterprises. Important differences exist between these two strata of informal workers given the presence or absence of kinship relations with the owner–employer of enterprises and given the nature of their (informal) employment contracts.

We can, thus, imagine four strata of working population active in the informal economic space: owners of larger informal establishments employing wage labor (akin to small-scale capitalist enterprises), self-employed owners of smaller informal enterprises without wage labor (akin to petty commodity production units), wageworkers hired in establishments, and non-wage family workers. For notational simplicity, we will identify the population members belonging to these strata, respectively, as: employers, petty producers, hired workers, and family workers. Agents belonging to different strata differ from each other in terms of their relative control over productive resources, labor processes, surplus produced, and power to make economic decisions within an enterprise. The stratification thus reflects the relative positions of advantage or disadvantage of agents operating within the informal economy.

While people belonging to lower strata of informal workforce would be better-off in terms of their earning possibilities and other livelihood conditions if they can move to a higher stratum, economic mobility across the strata is not easy. There are structural barriers to free mobility within the economy. These barriers give rise to segmentation of informal working population. Presumably, heterogeneity in experiences of *everyday* and *incidental* forms of vulnerability is also brought about by this segmentation.

In the mainstream literature, the structural barriers to mobility within informal economy are often described in terms of the differences in volume of initial endowments of productive resources, including physical assets, financial capital, and human capital (indicated by education levels) of agents (Fields 2005; Cox 1987). Several works from across the world—e.g., from Côte d'Ivoire (Günther and Launov 2012), India (Unni 2005), Cameroon (Nguimkeu 2014), Chile (Amuedo-Dorantes 2004), and Egypt (Radchenko 2014), suggest similar results. It is also speculated that differences in so-called entrepreneurial talent among agents bring about different outcomes for people operating in different segments of the economy (Cunningham and Maloney 2001).

When seen through the lens of social relations of production and employment, a richer image of this structural barrier becomes apparent. The barrier appears not just to be a matter of the volume of productive resources employed (or the capital intensity or the size of enterprise), but, more importantly, how the production and labor processes are controlled within the enterprises. Recent literature on informal economy in India, for example, highlights the constraints to mobility within the "upper" tier of the economy. The studies emphasize, in particular, that the pathways toward growth and expansion veer to be different between the smaller informal enterprises owned and operated by petty producers and the larger informal establishments run mostly with hired workers (Harriss-White 2012; Bhattacharya 2017). The informal establishments operate at much bigger scale and tend to be technologically more advanced, more dynamic and productive than the smaller *noncapitalist* enterprises. The establishments are also more likely to have the favorable characteristics that aid a firm to retain a larger proportion of earnings (or surplus) than the *noncapitalist* enterprises and to grow over time (Bhattacharya and Kesar 2018). The presence of constraints and structural barriers to upward economic mobility for petty producers perpetuates segmentation within the "upper" tier of the informal economy.

Such barriers may also exist within the "lower tier". Inferences drawn from case studies show that difference in conditions of living exists between wageworkers and non-wage family workers in the "lower tier". The presence or absence of kinship relation that a worker may have with the enterprise owner appears to play an important role in erecting the barrier. Differences in the contractual nature of work may also bring about a separation between informal workers hired against wages and the family workers laboring under income-sharing arrangements.

It is plausible that the segmentation of informal economic space would bring about a variation in how *everyday* and *incidental* forms of vulnerability *manifest* in the economic life of people operating in this space. In particular, the segments of workforce that are advantaged in the social relations of production and employment within this space would fare relatively better in terms of their ability to manage the importunate riskiness in *everyday* life, as well as in terms of their power to contend with *incidental* forms of vulnerability in periods of exogenous covariate shocks. We empirically explore this postulation in the following sections.

4 Household Survey Data from Indonesia

Our empirical evidences are drawn from Indonesia. The informal economy in the country contributes more than one-third of its gross domestic product (GDP) and provides employment for a vast majority of its workforce (Van der Loop and Andadari 2010). Some of the key characteristics of informal employment in the country are lack of protection from non-payment of wages, retrenchment without notice or compensation, unsatisfactory occupational health and safety conditions, and absence of social benefits such as pensions, sick pay, and health insurance. Needless to say, these conditions spawn *everyday* vulnerability in economic life of the workforce.

At the same time, Indonesia happens to be at risk to multiple hazards and the economy is particularly prone to covariate shocks of natural disasters. Over the past 30 years, there have been, on an average, 289 significant natural disasters per year, with an average annual death toll of approximately 8000. The disasters include flooding, earthquakes, landslides, tsunami, volcanic eruptions, and cyclones. Costs during major disaster years reach 0.3% of national GDP and as high as 45% of GDP at the provincial level (GFDRR 2011). These extreme natural events are likely to set off the conditions of *incidental* forms of vulnerability for population in the country.

We draw upon the Indonesia Family Life Survey (IFLS) dataset for our analysis. IFLS is a continuing longitudinal socioeconomic and health survey representative of households living in 13 of the nation's 26 provinces and is one of the few large-scale longitudinal surveys available for a developing nation. We start by considering the third wave of the survey (henceforth IFLS3). IFLS3 was fielded in 2000, three years after the economic crisis hit Indonesia, and covered 10,400 households. Our study is based on a subsample of 7525 households for which all relevant variables are recorded. Once we recognize the households in terms of employment categories of their heads, we divide them along the axes of formal and informal sectors, and then further categorize the informal households in terms of the four tiers described in the previous section, based on the location of the household heads in production and employment relations: employers, petty producers, hired workers, and unpaid family workers. The variables used for our study include, first and foremost, employment category of adult head of a household, the household head's gender and educational achievements, per capita level of consumption expenditure of the household on food and nonfood items per month, and monetary value of productive assets owned. We examine the differences in median consumption and asset levels of households. We also compare the indicators of normal living standards of the informal households with those engaged as employees in the formal sector of the economy (either as a public servant or in the private sector). We seek to draw inferences from this analysis on how *everyday* states of vulnerability manifest for people in different strata of employment.

In the second stage of our study, we focus on how heterogeneity in *incidental* states of vulnerability manifest, and examine if employment status in a prior period can explain relative risks of experiencing shock of "exposure" to an extreme event at a subsequent period. We consider the data on employment category for household heads as recorded in IFLS3 and obtain the data on whether the household was exposed to a disaster event that occurred between the third and fourth waves of IFLS. The fourth wave of the survey (henceforth IFLS4) was fielded in 2007. In the third stage of our study, we focus on the outcome of exposure to a disaster event and examine if prior employment status of household head (recorded in IFLS3) may explain the relative risks and the extent of deviations in consumption expenditure of household (recorded in IFLS4) that experienced "exposure".

5 Between-and Within-Sector Heterogeneity

Our main concern in this paper is with heterogeneity within the informal economy. But before proceeding to explore the within-sector differences, we enquire into the between-sector differences in labor force in terms of key attributes of heads of households surveyed in IFLS3. We abide by the more mainstream approach and examine if the differences in individual characteristics—for example, gender, geographic location, and educational achievements—may allow us to differentiate between the people operating in the formal and the informal sectors of the Indonesian economy. Table 1 describes these characteristics in broad terms.

Column 2 in Table 1 presents the count data of households in our studied population. It appears that informal economy supports a vast section of the working population. 93% of the surveyed households (7033 out of 7525 households) had their heads engaged in various segments of the informal economy. Two features of

Table 1 Headcount of households in various strata of informal and formal workforce

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8
Sector	Occupation categories	Total number of households in various occupation strata	Female	Urban	Education level			
					Less than primary level of schooling	Primary school level	Middle school level	High school (or higher) level
Informal	Employers	94	49	62	87	1	1	5
	Petty producers	6307	525	2939	1691	368	251	3997
	Family workers	501	56	252	162	26	21	292
	Hired workers	131	18	55	37	5	2	87
Informal	Total	7033	648	3308	1977	400	275	4381
Formal	Salaried employees	492	270	306	118	29	20	325
Population	Total	7525	918	3614	2095	429	295	4706

these households are worth noting. First, Column 3 in Table 1 indicates that among the heads of households in our population, while only 9% of those participating in the informal economy were female, 55% among the salaried employees in formal sector are female (respectively 648 and 270 individuals). Second, Column 4 in the table, in turn, indicates that 47% of households in the informal sector were located in urban areas, while 62% of households in the formal sector were urban (respectively 3308 and 306 households).

The more curious feature is, however, in terms of educational attainments of these people. By and large, the levels of attainment show a similar pattern for people engaged in formal and in informal work. For example, 28% of the informal workforce is with less than primary school-level of education, while 62% had secondary or higher level of education; the corresponding numbers for salaried employees in formal sector are 24 and 66% (see columns 5 and 8 of Table 1). Indeed, there are many more people with high school degree in our studied population than with primary school degree. Even more confounding feature is the educational distribution among the people engaged in various tiers of the informal economy. We find that among the employers, i.e., among the people in the higher strata of the "upper" tier of the informal economy, 93% have less than primary level of schooling, while only 0.05% has secondary or higher level of education. Whereas among the petty producers, i.e., among the people at the lower strata of the "upper" tier, 27% have less the primary education, while 63% have high school or higher level of education. Similarly, in the "lower" tier, much higher proportions of family workers and hired workers have higher education than that for the employers.

These features are very unlike the ones described in the literature for informal economies across the world (see, for example, Günther and Launov 2012; Unni 2005; Amuedo-Dorantes 2004; Cunningham and Maloney 2001). Studies on informal workforce typically show that the proportion of "less" educated vis-à-vis "more" educated people are higher in informal sector than in formal sector of employment and higher in upper strata of the informal economy than in the lower strata. Based on these results, personal attainment of schooling is often held as an important determinant of people's chances for finding secure and stable conditions of employment and their chances of economic mobility. The idea is, however, not validated in our studied population. One reason for this may be that 12 years of schooling is compulsory for all people in Indonesia and is provided free of charge at public schools. While the quality of schooling may of course differ among groups, education is accessible to all till high school. It is possible that in the context of Indonesia, an individual's attainment of schooling may not act as barrier to entry to formal sector or to the higher strata of the informal sector, but other structural factors may be at play.

We shift our attention to other within-sector differences in attributes of people engaged in various tiers of informal work. Almost 90% of informal working population in our studied data are petty producers (6307 individuals), while a little more than one percent are owners of establishments that employ hired workers (94 individuals). The gender composition of these owners of enterprises is interesting. Given the prevalent socioeconomic norms of patriarchy, the women-headed households and enterprises are generally expected to perform worse than the male-headed ones

(Bhattacharya and Kesar 2018). Hence one would expect the enterprises at the higher strata of the informal economy to be comprised more of male-headed enterprises than female-headed ones. However, we find that only 8% of the petty producers are female, while as much as 52% of the employers are female. This, as well, is in sharp contrast to the situation in many other countries in the Global South, where women, if they happen to own enterprises in the informal sector, are more likely to be concentrated in petty production, rather than in larger production units employing wageworkers.[5]

How do material conditions of living vary and vulnerability conditions manifest for these people engaged in various tiers of informal work? Table 2 allows us to describe these conditions in terms of consumption expenditures (on average) and asset situations of households belonging to various strata.

There is a sizeable difference in consumption expenditures (on average) of households operating in informal and formal sectors of economy, with the median expenditure in the latter class of households being about 19% higher than the former (see column 2 of Table 2). Also, the median value of productive assets owned by the latter is higher than the former, although there is a large degree of variance in asset ownership among the salaried employees in formal sector (see columns 4 and 5 of Table 2). These results are similar to those obtained for other informal economies in the Global South. One can speculate that this lower access to consumption and asset is symptomatic of what Bourdieu (1998) called *précarité* in everyday life of informal labor.

We expect an initial level of differentiation between the "upper" vis-à-vis the "lower" segments of informal workforce would exist in terms of consumption levels and asset ownership. A further differentiation may exist within the upper tier, i.e., among the petty producers and the owner–employers of larger informal establish-

Table 2 Monthly consumption expenditure (per capita) and asset value (productive assets) of households in various strata of informal workforce

Sector	Column 1	Column 2	Column 3	Column 4	Column 5
	Occupation categories	Consumption expenditure		Assets	
		Median	Variance	Median	Variance
Informal	Employers	195,000	2.39E+10	30,900,000	1.86E+16
	Petty producers	114,813	3.58E+10	12,500,000	8.29E+15
	Family workers	106,633	2.55E+10	10,300,000	1.35E+16
	Hired workers	95,633	4.68E+10	312,500	1.95E+14
Informal	Total	109,546	3.07E+10	11,400,000	1.0895E+16
Formal	Salaried employees	130,420	3.07E+10	12,900,000	3.48E+15
Population	Total	114,722	3.48E+10	13,065,008	8.71E+15

[5]It is, however, possible that for many of the female-headed informal establishments, the actual decision-making power in the enterprise might be with male members of the households, given the patriarchal social relations.

ments, as well as within the lower tier, i.e., among the family workers and the hired workers. The results in Table 2 provide some interesting insights in this regard.

First, median consumption levels for employers and petty producers, indeed, tend to be higher than that for the family workers and hired workers as expected, with owners of informal establishments having significantly higher consumption expenditure than petty producers on an average (see column 2). Second, the unpaid family workers have higher level of consumption expenditure than the hired wageworkers in informal sector, although there is a large degree of variance in consumption expenditure for the latter class of workers (see columns 2 and 3). Third, median value of productive assets owned by employers in informal economy is by far greater than that of the rest of the population: it is 2.5 times the median value of productive assets owned by petty producers, three times that owned by family workers, and almost hundred times that owned by hired workers (see column 4). While the distribution of median values of productive assets conforms to the stratification of the informal workforce along expected lines, a striking feature is the magnitude of difference between different strata.

Table 3 presents cross-classification of our studied population in terms of occupation and asset ownership. Our asset groups are population quartiles derived according to the distribution of asset values of surveyed households.

Columns 2, 5, 8, and 11 of Table 3 present the count data of households in asset quartiles 1 through 4 (from the poorest to the richest). While the petty producers are almost evenly distributed across various strata, a vast majority of the employers (68%) are concentrated in asset stratum 3. Family workers and hired workers are also evenly spread across the asset strata. In contrast, a large section of salaried employees in formal sector (47%) is in the highest asset stratum.

Certain observations on consumption expenditure of households in various tiers of workforce can be made. First, no matter the occupation category, median expenditures of households tend to rise as we climb up the asset ladder. In other words, median consumption expenditures increase monotonically with median values of productive assets for all households. Second, when we compare between-group consumption expenditures for each asset stratum, we find that the per capita expenditures of households belonging to the "upper" tier of informal economy is, on an average, greater than that of the households belonging to the "lower" tier for all the asset strata. The only exceptions are family workers in asset stratum 2, who have higher per capita expenditure petty producers in the same stratum. Otherwise, the hierarchy between occupation segments is strictly maintained in terms of average consumption expenditures for all asset strata in the informal economy, with hired workers faring the worst in all strata.

Certain broad inferences can be derived from our findings above. Consumption level and asset ownership tend to be higher for households in situations of relative advantage given their employment strata in the informal economy. These are the households for whom *everyday* experiences of vulnerability are relatively less. Educational attainment, however, seems not to differ across the strata. One crude implication of the result might be that access to higher living standards in terms of

Table 3 Monthly consumption expenditure (per capita) of households in various asset strata and strata of informal workforce

	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13	Column 14
Sector	Occupation categories	Asset stratum 1 (poorest)			Asset stratum 2			Asset stratum 3			Asset stratum 4 (richest)			Total number of households in occupation strata
		Total number of households	Consumption expenditure		Total number of households	Consumption expenditure		Total number of households	Consumption expenditure		Total number of households	Consumption expenditure		
			Median	Variance		Median	Variance		Median	Variance		Median	Variance	
Informal	Employers	17	101,830	1.36E+10	10	108,600	2.10E+10	64	113,820	1.65E+09	3	274,446	1.16E+11	94
	Petty producers	1392	90,022	1.94E+10	1618	92,495	1.85E+10	1682	110,746	3.82E+10	1615	232,888	5.58E+10	6307
	Family workers	109	86,855	4.84E+10	129	94,221	9.69E+09	130	100,008	2.83E+10	133	178,175	1.71E+10	501
	Hired workers	38	47,608	3.56E+09	30	76,229	7.02E+09	27	79,167	2.89E+10	36	148,517	1.10E+11	131
	Total	1616			1787			1842			1788			7033
Formal	Salaried employees	61	79,696	8.28E+09	75	93,429	9.84E+04	126	100,746	2.08E+10	230	177,926	4.21E+10	492
Population	Total	1677			1862			1968			2018			7525

consumption is associated with higher levels of material resource (productive assets), but not with nonmaterial resource (education), at least in our population.

6 Within-Sector Differences in Disaster Vulnerability in the Informal Economy

Our second and third propositions in the paper are with regard to *incidental* forms of vulnerability. We imagine that this form of vulnerability emanates from people's relative inability to secure for themselves the conditions that enable them to resist sudden disruptions in their economic life, or at least recover quickly from the adverse effects of these disruptions. We consider incidences of natural extremes (e.g., volcanic eruptions, earthquakes, and tsunamis) as examples of such disruptive events. We focus on disaster incidences in Indonesia between third and fourth waves of IFLS for our study. IFLS4 chronicles whether or not a surveyed household had experienced exposure shocks to these disasters that had occurred in the between-survey years. We find that a total of 537 households have experienced shocks (about 7% of our studied population in IFLS4), of whom 93% (498 households) were reported to have been engaged in informal work in IFLS3 before the incidences of these disasters. This finding provides us with at least an initial empirical knowledge that an association exists between people's prior conditions of employment insecurity and their subsequent conditions of *incidental* vulnerability.

Our concern is, however, with the heterogeneity in experiences of exogenous shock across various segments of the informal economy. The premise based on which we proceed is a familiar one: risk realized in times of disasters within working populations in informal economy would be relatively less for households that are relatively less vulnerable, and it would be more for households that are not. Using the data from IFLS4, we proceed to carry out a simple exercise to measure risk realized within various strata of informal workforce.

We consider two dimensions of risk associated with natural extremes. The first entails physical exposure to unfavorable, or even dangerous, conditions that arise in one's natural physical environment in times of natural extremes. The second entails loss of something of value, for example, consumption, from exposure to the hazardous conditions. The differences in the two notions of risk can be illustrated as follows: risk of flood exposure for a household would be its probability of experiencing the rising water levels (or an overflow), given that it is located in an area affected by flooding. Risk of loss in consumption from flood exposure would be the probability for the household of experiencing a shortfall in consumption level, given that it had experienced the disruption effect of the shock of flood exposure. We consider the risk realized within the population of hired workers in informal economy of Indonesia as the benchmark category and measure the risk realized within the classes of employers, petty producers, and family workers relative to this benchmark. For this analysis, we borrow a technique commonly applied in, for example, epidemiological studies.[6] The technique can be formally described thusly.

Consider a discrete indicator variable of disaster exposure with J levels $(\Phi = \phi_1, \ldots, \phi_J)$, with ϕ_1 corresponding to the lowest category of exposure. Also consider the stratifying factor Ψ that segments the informal economy in K segments $(\Psi = \psi_1, \ldots, \psi_K)$. Let ψ_1 corresponds to the lowest stratum and ψ_K corresponds to the highest stratum in population. The cross-classification of Φ and Ψ yields JK levels in population. The risk of disaster exposure at level $\Phi = \phi_j$ for households that belonged to stratum $\Psi = \psi_k$ in a prior period (before the disaster incidence) is simply given by the conditional probability, $Pr\left(\Phi = \phi_j | \Psi = \psi_k\right)$. And, the risk for households belonging to stratum $\Psi = \psi_k$ relative to stratum $\Psi = \psi_1$ is

$$R_k = Pr\big(\Phi = \phi_j | \Psi = \psi_k\big) / Pr\big(\Phi = \phi_j | \Psi = \psi_1\big),$$

where $j = 1, \ldots, J$; $k = 1, \ldots, K$.

Given our data, disaster exposure Φ has $J = 2$ levels $(\Phi = \phi_1, \phi_2)$ in the studied population of 7525 households in IFLS, with the lowest category, ϕ_1, simply being "non-exposure" (baseline level) and the highest category, ϕ_2, being "exposure". As mentioned earlier, there are 537 households in the population category ϕ_1, and the rest are in category ϕ_2. In our data, Ψ, on the other hand, has $K = 4$ levels $(\Psi = \psi_1, \ldots, \psi_4)$, representing the four tiers of informal workforce, with, for example, ψ_1 consisting hired workers employed against wages and ψ_4 consisting employers who own informal establishments. Table 4 describes the risk of disaster "exposure" realized within various strata of households with their heads engaged in informal economy. The results lend credence to our second proposition in the paper.

Column 4 in Table 4 presents the count data of "exposed" households in each stratum in informal as well as in formal segments of the Indonesian economy, and column 5 describes the corresponding risk of "exposure" for a randomly chosen member of the strata. We find the risk of experiencing an exposure shock to disaster

Table 4 Relative risk of exposure to disaster shocks for households in various strata of informal workforce

Sector	Occupation categories	Total	Disaster exposed	Risk (incidence rate)	Relative risk (vis-à-vis hired workers)
Informal	Employers	94	3	0.032	0.380
	Petty producers	6307	444	0.070	0.838
	Family workers	501	40	0.080	0.951
	Hired workers	131	11	0.084	1.000
	Total	7033	498	0.071	0.843
Population	Total	7525	537	0.070	0.850

[6] See the applications, for example, in Kraemer et al. (1997) and Lachin (2009).

conditions is highest if a household had its head employed as a hired informal worker prior to the disaster incidence, and risk is lowest for employers owning informal establishments.

Column 6 in Table 4 describes the relative risk of "exposure" for households in different strata vis-à-vis the hired workers. Relative risk presents us with a way to quantify the magnitude of association between the location of a household in a prior period in a stratum of informal workforce and its subsequent chances of experiencing exposure to covariate shocks. For example, when we find relative risk is less than 1, our interpretation is that the chance of experiencing "exposure" is higher for a household headed by a hired worker ($\Psi = \psi_1$) than other types of households ($\Psi = \psi_k, k \neq 1$). As column 6 shows, the chance is lower for all categories of households in informal economy when compared to the hired workers. The smaller the relative risk, the more strongly is it associated with a lower chance of "exposure". The chance is lowest for employers, followed by petty producers and family workers.

In other words, the pattern of risk profiles of households in various strata of informal economy reflects the embedded hierarchy between the "upper" and the "lower" tiers, as well as that within each tier. It is, indeed, curious that natural disasters, which are, conventionally, examined in the literature as random, exogenous shocks, with their impacts realized in a determinate system in an arbitrary manner (Benson and Clay 2004; Toya and Skidmore 2007; Barro and Ursua 2008; Noy 2009) produce a pattern of risk of disaster "exposure" along the lines of the embedded hierarchies in the informal economy. This problematizes the asocial and asystemic understanding of disasters as exterior phenomena. Rather, it points toward an alternative plausible conjecture that extreme natural events get actualized as disaster shocks in specific ways in the presence of specific structural and material conditions of people's lives and their locations along the economic axes.

In the particular context of the informal economy in Indonesia, a further conjecture regarding the relationship between risks of disaster exposure and material conditions of households can be made. Our second proposition in this paper relates the *incidental* forms of vulnerability to the locations of households and individuals in various strata in the "upper" or "lower" tiers of the informal economy. However, we have found in Sect. 5 above that some of the usual markers of hierarchy among these strata, like gender and education levels of heads of households, do not strictly mirror the linear hierarchy. Hence, at the surface, it seems unlikely that the differences in risks of exposures can be solely attributed to the variations in these characteristics among the workforce in different strata of the informal economy. Rather, the characteristic that reflects the hierarchy among the strata, i.e., the average value of productive assets that individuals in each strata control, might play a major role in shaping conditions of security and stability for the individuals. Thus, the control over productive resources might be one of the key indicators of relative vulnerability of workforce in the informal economy in Indonesia.

In examining the relative risk of loss for households from exposure to the covariate shock of natural extremes, we examine consumption deviations for households with "disaster exposure". This brings us to our third proposition in this paper, which is: for the "exposed" households, risk of consumption loss would be relatively lower for those in a position of relative advantage within the informal economy prior to the shock. To explore this proposition, we start by focusing on only the subpopulation of 537 "disaster exposed" households identified in IFLS4. We examine consumption expenditure (per capita, given the household size) of the households in IFLS4 and look back at their per capita consumption expenditures recorded in IFLS3. Using these two data series, we define for each household i a continuous variable d_{i1} that measures consumption deviation, where

$$d_{i1} = c_{i0} - c_{i1} \tag{1}$$

and c_{i0} is the consumption expenditure in 2000 and c_{i1} is that in 2007 (deflated by 2000-level prices) for the household. By definition, c_{i0} and c_{i1} are, respectively, the expenditures in pre- and post-disaster periods. While many factors may bring about a deviation in households' consumption expenditure over time, we consider a rather imprecise method to examine if one of these factors might be the occupation category of the household head in a prior period. Given the series on d_{i1}, we consider two categories in our data: loss is "observed" ($D = 1$) and "not observed" (baseline level) ($D = 0$). The complete cross-classification of D, Φ, and Ψ yields $2 \times JK$ levels in data. Risk of consumption loss for households with disaster exposure at level $\Phi = \phi_j$ and belonging to employment stratum $\Psi = \psi_k$ can be expressed as the conditional probability $Pr\ (D = 1 \mid \Phi = \phi_j,\ \Psi = \psi_k)$. The relative risk for these households vis-à-vis the hired workers ($\Psi = \psi_1$) is

$$R_{j|k} = Pr\left(D = 1|\Phi = \phi_j, \Psi = \psi_k\right)/Pr\left(\mathrm{D} = 1|\Phi = \phi_j, \Psi = \psi_1\right). \tag{2}$$

Column 5 in Table 5 describes absolute value of risk, and column 6 describes the relative risks realized within the different classes of informal working populations.

Table 5 Relative risk of consumption loss from exposure to disaster shocks for households in various strata of informal workforce

Sector	Occupation categories	Disaster exposed	Experience loss in consumption	Risk (incidence rate)	Relative risk (vis-à-vis hired workers)
Informal	Employers	3	0	0.000	0.000
	Petty producers	444	213	0.480	0.880
	Family workers	40	18	0.450	0.825
	Hired workers	11	6	0.545	1.000
	Total	498	237	0.476	0.872
Population	Total	537	260	0.484	0.888

It appears that if a household had experienced disaster exposure, then its chance of experiencing consumption loss would be greatest (0.545) if it had belonged to the class of hired workers prior to this exposure. This might be due to the fact that, given the casualized nature of their employment relations, they are more likely to lose their jobs when disaster shock negatively affects their work environments. This chance of consumption loss, in contrast, is nonexistent, if a household had belonged to the class of employers, at least in our studied dataset. It should, however, be noted that one cannot seek to derive any general inference about the vulnerability conditions of the employers given that only three employers in the dataset were exposed to disaster. Also, when $R_{j|k}$ is less than 1, the interpretation is that the risk of loss is higher for hired workers than any other class of households, even when both classes of households experienced disaster exposure at the jth level ($\Phi = \phi_j$). We find this to be true for households in all the other three strata in the formal economy.

However, the unusual result is that relative risk of consumption loss is less for family workers (0.825), who belong to the "lower" tier of the informal economy (and who own, on an average, a lower value of productive assets), than for petty producers (0.880), who belong to the "upper" tier. Following the body of literature that takes off from Sanyal (2007), we may consider the following hypothesis to explain this odd finding. The petty producers do not have any clear demarcation between the enterprise, i.e., the space of production, and the household, i.e., the space of consumption, and the economic survival of households is intricately linked to the reproduction of enterprises (Bhattacharya and Kesar 2018). This enables as well as impels them to deal with the disruptive effects of disaster shocks in a different way from that of the employers (who maintain a degree of separation between the household and the enterprise) and the family workers or hired workers (who do not have any control and decision-making power in the enterprise). In the face of disaster exposure, a petty producer, driven by the logic of reproduction of the enterprise, may curtail household consumption (thus recording a higher risk of consumption loss from exposure) and use a higher proportion of the retained earning from the enterprise in productive purposes, thereby ensuring the survival of the enterprise. This is, usually, not possible for individuals in the "lower" tier of informal economy.

Broadly speaking, we interpret the results on unevenness in risk distribution within informal economy as indicative of the unevenness in people's ability to attend to shock-generating events in their life. Stated differently, the heterogeneity within informal economy is manifested in terms of how risk is realized in a heterogeneous manner across various strata of informal workforce.

7 Some Broad Conclusions

In this paper, we have presented a preliminary exploration of the possible links between conditions of informality and disaster vulnerability of populations that derive their livelihoods in the informal economy. We have posited that *everyday* forms of vulnerability manifest in people's economic lives differentially according to their different locations in the formal and informal segments of the economy. We have further hypothesized that *incidental* forms of vulnerability that are related to risks of "exposure" to disaster shocks and of suffering from consumption loss due to such exposures would be heterogeneous for the informal workforce depending on their locations in different strata of the informal economy.

Given the entrenched heterogeneity among the people within the informal economy in terms of their positions in labor and production processes, the vulnerability conditions for informal workforce are also likely to be heterogeneous. There is an embedded hierarchy between the "upper" and "lower" tiers of the informal economy and the various strata within each tier in terms of relative control over production and labor processes. This hierarchy is usually manifested in the individual characteristics of the workforce in each tier and stratum in terms of their educational attainments, access to productive assets, etc., which in turn affect, at least partly, their conditions of vulnerability.

Based on the above understanding, we have put forward three elementary propositions in the paper and provide some illustrations using longitudinal household survey data from Indonesia, where a vast proportion of workforce is engaged in informal economy and which is prone to experiencing frequent natural disaster shocks.

We find that individuals in the formal sector are able to attain higher levels of consumption expenditures, asset ownership and education than those in the informal sector. Yet, the expected patterns of distribution of individual characteristics between the formal and the informal economy, as well as between different strata within the informal economy, do not always uniformly hold for Indonesia. Specifically, within the informal economy, we find that individuals in the "upper" tier are much better-off in terms of consumption expenditures that those in the "lower" tier, though we do not find any discernible hierarchy between the strata in terms of educational attainments of individuals. However, the distribution of average values of productive assets owned by individuals in the formal economy, as well as in various strata of the informal economy, follows a strict hierarchy, with large differences between the average values of assets for each segment.

Further, we find that the pattern of *incidental* vulnerability of individuals in different tiers and strata of the informal economy to disaster shocks, as understood in terms of their risks of "exposure" to shocks and their risks of experiencing consumption losses due to such exposure, is closely associated to their locations in the economy. The risk of exposure falls as one moves from "lower" to "upper" tiers, as well as when one moves up the strata from hired workers to family workers to petty producers and finally to employers. However, we find that while risk of consumption loss due to exposure is highest for hired workers, the relative risk (vis-à-vis hired workers) for petty producers is higher than that for family workers.

This exploration helps us to put forward two preliminary arguments for consideration and for further investigation. First, we argue it is imperative to move away from the standard economic understanding of natural disasters as random, exogenous shocks, whose effects are realized in the system and materialized in the economic conditions of individuals in an arbitrary manner. Rather it is necessary to explore how extreme natural events get actualized as disaster phenomena only in the presence of specific social and structural phenomena, which, in turn, produces specific conditions of vulnerability. Put differently, it is necessary to provide a *social* theory of natural disasters. Second, we argue "normal" or everyday conditions of economic and social lives of individuals—and the embedded hierarchies, contradictions, and possibilities—formatively affect, and may even determine, the conditions of vulnerability of individuals facing extreme natural events. In this regard, it is conjectured that control over productive resources as well as over labor processes may play a fundamentally important role (along with other cultural, political, and economic aspects) in shaping the vulnerability conditions of individuals and households. We think it would be immensely productive and illuminating if further research is done in these directions in the future.

Acknowledgements We are grateful to Surbhi Kesar and Ethan Muller for their substantial support at various stages of our research.

References

Adger, W. N. (2006). Vulnerability. *Global Environmental Change, 16*(3), 268–281.

Agarwala, R. (2013). *Informal labor, formal politics, and dignified discontent in India*. New York: Cambridge University Press.

Amuedo-Dorantes, C. (2004). Determinants and poverty implications of informal sector work in Chile. *Economic Development and Cultural Change, 52*(2), 347–368.

Barro, R. J., & Ursua, J. F. (2008). Consumption disasters in the twentieth century. *American Economic Review, 98*(2), 58–63.

Benson, C., & Clay, E. (2004). *Understanding the economic and financial impacts of natural disasters*. Disaster Risk Management Series No. 4. Washington, DC: World Bank.

Bhattacharya, S. (2017). Reproduction of noncapital: A Marxian perspective on the informal economy in India. In T. A. Burczak, R. Garnett, & R. McIntyre (Eds.), *Knowledge, class, and economics: Marxism without guarantees* (pp. 346–358). New York and London: Routledge.

Bhattacharya, S., & Kesar, S. (2018). Possibilities of transformation: The informal sector in India. *Review of Radical Political Economics, 50*(4), 727–735.

Blunch, N. H., Canagarajah, S., & Raju, D. (2001). The informal sector revisited: A synthesis across space and time. *World Bank Social Protection Discussion Papers, 119*.

Bourdieu, P. (1998). La Précarité est Aujourd'hui Partout. *Contre-feux* (pp. 95–101). Paris, France: Liber-Raison D'agir.

Chakrabarti, S. (2016). *Inclusive growth and social change: Formal-informal-agrarian relations in India*. New Delhi: Oxford University Press.

Chambers, R. (2006). Vulnerability, coping and policy (Reprinted from *IDS Bulletin, 20*, 1989). *IDS Bulletin*. *Institute of Development Studies, 37*, 33–40.

Cook, S., & Kabeer, N. (2010). *Social protection as development policy: Asian perspectives*. New Delhi: Routledge.

Cox, D. (1987). Motives for private income transfers. *Journal of Political Economy, 95*(3), 508–546.

Cranford, C. J., Vosko, L. F., & Zukewich, N. (2003). Precarious employment in the Canadian labour market: A statistical Portrait. *Just labour, 3,* 6–22.

Cunningham, W. V., & Maloney, W. F. (2001). Heterogeneity among Mexico's microenterprises: An application of factor and cluster analysis. *Economic Development and Cultural Change, 50*(1), 131–156.

Cutter, S. L., & Emrich, C. T. (2006). Moral hazard, social catastrophe: The changing face of vulnerability along the hurricane coasts. *The Annals of the American Academy of Political and Social Science, 604*(1), 102–112.

Fields, G. S. (2005). A welfare economic analysis of labor market Policies in the Harris-Todaro Model. *Journal of Development Economics, 76*(1), 127–146.

Global Facility for Disaster Reduction and Recovery (GFDRR). (2011). Vulnerability, risk reduction, and adaptation to climate change. Available online at: https://www.gfdrr.org/sites/default/files/publication/climate-change-country-profile-2011-indonesia.pdf. Accessed on May 27, 2019.

Günther, I., & Launov, A. (2012). Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics, 97*(1), 88–98.

Harriss-White, B. (2012). Capitalism and the common man: Peasants and petty production in Africa and South Asia. *Agrarian South: Journal of Political Economy, 1*(2), 109–160.

Hewitt, K. (1997). *Regions of risk: A geographical introduction to disasters*. Essex, UK: Longman.

International Labour Office (ILO). (2006). *Realizing decent work in Asia*. Report of the Director-General, 14th Asian Regional Meeting. Geneva: ILO.

Kalleberg, A. L. (2009). Precarious work, insecure workers: Employment relations in transition. *American Sociological Review, 74*(1), 1–22.

Kalleberg, A. L., & Hewison, K. (2013). Precarious work and the challenge for Asia. *American Behavioral Scientist, 57*(3), 271–288.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*(4), 337–343.

Lachin, J. M. (2009). *Biostatistical methods: The assessment of relative risks*. New York: John Wiley & Sons.

National Commission for Enterprises in the Unorganised Sector (NCEUS). (2007). *Definitional and statistical issues: Task force report*. New Delhi: NCEUS.

Nguimkeu, P. (2014). A structural econometric analysis of the informal sector heterogeneity. *Journal of Development Economics, 107,* 175–191.

Noy, I. (2009). The macroeconomic consequences of disasters. *Journal of Development Economics, 88*(2), 221–231.

Radchenko, N. (2014). Heterogeneity in informal salaried employment: Evidence from the Egyptian Labor Market Survey. *World Development, 62,* 169–188.

Resnick, S., & Wolff, R. (2006). *New departures in Marxian theory*. London and New York: Routledge.

Sanyal, K. (2007). *Rethinking capitalist development: Primitive accumulation, governmentality and post-colonial capitalism*. New Delhi: Routledge.

Sudarshan, R. M., & Unni, J. (2003). Measuring the informal economy. In R. Jhabvala, R. M. Sudarshan, & J. Unni (Eds.), *Informal economy centrestage: New structures of employment* (pp. 19–38). Sage. New Delhi: New Delhi.

Tierney, K. J. (2007). Businesses and disasters: Vulnerability, impacts, and recovery. In H. Rodríguez, E. Quarantelli, & R. Dynes (Eds.), *Handbook of disaster research* (pp. 275–296). New York: Springer.

Toya, H., & Skidmore, M. (2007). Economic development and the impacts of natural disasters. *Economics Letters, 94*(1), 20–25.

Unni, J. (2005). Wages and incomes in formal and informal sectors in India. *Indian Journal of Labour Economics, 48*(2), 311–319.

Van der Loop, T., & Andadari, R. K. (2010). *Social security for informal economy workers in Indonesia: Looking for flexible and highly targeted programmes*. Jakarta: International Labour Organization (ILO).

Webster, E., Lambert, R., & Bezuidenhout, A. (2008). *Grounding globalization: Labour in the age of insecurity*. Oxford, UK: Blackwell.

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2004). *At risk: Natural hazards, people's vulnerability and disasters*. London and New York: Routledge.

Woolcock, M., & Narayan, D. (2000). Social capital: Implications for development theory, research, and policy. *The World Bank Research Observer, 15*(2), 225–249.

Wright, E. O. (1978). Race, class, and income inequality. *American Journal of Sociology, 83*(6), 1368–1397.

Estimation of Median Incomes of the American States: Bayesian Estimation of Means of Subpopulations

Hee Cheol Chung, Gauri Sankar Datta and Jerry Maples

Abstract The Fay–Herriot model (J Am Stat Assoc 74:269–277, 1979) is an immensely popular model in small area estimation of means of some characteristics for many related subpopulations. This model regresses the subpopulation mean on a set of auxiliary variables to borrow strength from other subpopulations. The Fay–Herriot model uses a fixed regression coefficient vector but only a random intercept term to account for variability that cannot be explained by the non-random mean function. In some applications, the non-random regression coefficient often does not adequately account for the variability of the subpopulation means. To accommodate extra variability, we consider an extension of the random intercept model by treating some of the regression coefficients also as random. This model is referred to as the random regression coefficient model. For the flexible random regression coefficient model, we allow the intercept and some of the regression coefficients to randomly vary across states with a suitable normal distribution. We use a suitable noninformative prior for all the model parameters to conduct our Bayesian analysis. We establish propriety of the resulting posterior density function and generate Monte Carlo samples from this distribution to get point estimates and associated measures of accuracy of these estimates. We also construct relevant credible intervals of the small area means as another measure of uncertainty for the point estimates. The method is illustrated to predict four-person family median incomes for 1989 for the US states.

Keywords Area-level model · Current population survey · Empirical Bayes · Fay–Herriot model · Hierarchical Bayes · Noninformative Bayes · Propriety of posterior density · Random regression coefficient model · Small area estimation · Stein's shrinkage estimation

---

H. C. Chung (✉) · G. S. Datta
Department of Statistics, University of Georgia, Athens, GA 30602, USA
e-mail: heecheol.chung25@uga.edu

G. S. Datta
e-mail: gauri@uga.edu

G. S. Datta · J. Maples
Center for Statistical Research and Methodology, US Census Bureau, Washington, DC 20233, USA
e-mail: jerry.j.maples@census.gov

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_23

1 Introduction

Small area estimation has become an important topic in survey sampling. Both the public and the private sectors during the last forty years felt an acute need to produce reliable estimates for many important economic, health, or social characteristics for various segments of a population. For example, the US Department of Education is required to produce reliable estimates of the poverty rates of school-age children (5–17 years old) for all the counties. The US Department of Health and Human Service is in need to accurately estimate health insurance access for different population subgroups defined by age, race, sex, and geography.

Sample surveys conducted to produce estimates of economic characteristics are usually targeted to achieve desired accuracy (such as a margin of error for the estimate) of the estimates for the entire population or domain at the national level. The data available from a survey for sub-national domains (or sub-domains which are subsets of the surveyed population) can be small or even nonexistent for some domains. Consequently, sub-national estimates that are calculated based on the corresponding sub-national sample only, if there is any, are subjected to a large sampling variability. These estimates, referred to as the "direct" estimates in the small area estimation literature, often have so large variability that they cannot be reliably used to make official policy decisions.

Populations for which estimates of characteristics are needed, but do not have adequate samples to produce reliable direct estimates, are called small areas or small domains. Small areas are often formed by demographic and/or geographic division of the population. More precise estimates for a common characteristic of many small areas can be constructed by "borrowing strength" from the other direct estimates and related auxiliary variables. Borrowing of information from the direct estimates of other small areas and related auxiliary variables to produce reliable "indirect" small area estimates is pursued through shrinkage estimation. A shrinkage estimator is obtained by shrinking a direct estimate toward another estimate of that characteristic, formed by using data on auxiliary variable and direct estimates from the other small areas.

Stein (1956) in a seminal paper introduced shrinkage estimation of a multivariate normal mean vector, where each component is measuring similar characteristic for one of the groups. Stein (1956) and James and Stein (1961) explicitly showed under sum of squared error loss that suitable shrinkage estimator of the mean vector can be obtained which is more precise than the direct or the standard estimator under appropriate balanced conditions.

Shrinkage estimators became immensely popular when in a series of articles Efron and Morris put forward an interesting empirical Bayes (EB) interpretation (see, e.g., Efron and Morris 1973). Their EB interpretation of the shrinkage estimators provided a transparent and heuristic justification of the domination result established by James and Stein (1961). Moreover, since the EB method can be used to develop shrinkage estimators even when the balanced conditions needed for the explicit domination are not realized, it increases the utility and popularity of shrinkage estimators in applications.

Small area estimation methodology benefited immensely from Stein's shrinkage estimation to develop reliable small area statistics by borrowing strength from the direct estimates of the other areas and appropriate auxiliary variables available for all the areas. A part of the variability of the population small area means is explained through some regression model based on the auxiliary variables. The regression model connects the direct estimates from the survey data with the auxiliary variables to construct indirect "synthetic regression estimates" of the small area means used in shrinkage estimation.

In multi-level models, such as small area estimation models, as an alternative to EB method, a hierarchical Bayes (HB) method is well suited to produce estimates of multiple subpopulation means, which measure similar population characteristics across areas. An HB method is really a Bayesian method, where the prior distribution is expressed in multiple steps. To be specific, the prior distribution of the related parameters (e.g., the subpopulation means) is expressed in multiple or hierarchical steps. In step 1 of the prior distribution hierarchy, conditional on some common parameters, a joint distribution of the true population means is specified, and in the second step a prior distribution on the common parameters is specified. This hierarchy can continue beyond two steps. If all these prior distributions are available, we can pursue an HB method to estimate the related parameters or any function of them.

Often, in many applications, the main difficulty is to specify the prior distribution on the common parameters appearing in the last stage of an HB model. While Bayesians may use a vague prior at this stage, if necessary, other statisticians estimate these parameters from the data and use them to construct empirical Bayes (EB) estimators. While the shrinkage estimators based on the EB method may have explicit expressions, the HB estimators usually lack such expressions and are computed numerically. However, both the HB and the EB estimators benefit from shrinking the direct estimators in producing more accurate estimators for all the parameters, assessed via frequentist risk (based on the sampling distribution of the direct estimators as given by the sampling part of the Fay–Herriot model below) resulting from the sum of squared error loss.

In a pioneering article, Fay and Herriot (1979) suggested an important extension of Stein's shrinkage estimation using Efron and Morris's EB proposal for estimation of small area characteristics. Many authors subsequently developed the fully Bayesian (that is, HB) version of the Fay–Herriot model (see, e.g., Datta et al. 1991, 1996, 2005). After explaining its components, we consider a more general small area estimation model which also includes random regression coefficients. While the original Fay–Herriot model is based on the assumption of constant variability of the small area population means, this general model is more appropriate if the small area means have unequal variances. Note that the Fay–Herriot model is a special case of the random coefficient model. In Sect. 3, we compare these two models by applying these to estimate four-person family median incomes for all fifty American states and the district of Columbia. We conduct a small simulation study in Sect. 4 to further our comparison of these two models. Section 5 provides a few concluding remarks. We added an appendix to provide a proof of the propriety of the posterior probability density function for the random regression coefficient model that resulted from noninformative improper prior density for the model parameters.

2 The Fay–Herriot Model and the Random Regression Coefficient Model

Suppose there are m small areas, and for a characteristic Y, we want to estimate θ_i, which we interpret as the mean of the characteristic Y for the ith small area. Suppose data from a survey have been summarized, and $Y_1, \ldots, Y_m$ are the summary statistics that we refer to as direct estimates of θ_i, $i = 1, \ldots, m$. Many direct estimates do not meet the accuracy requirement due to their small sample sizes. Often in small area estimation, in addition to the direct estimates, data are also available for auxiliary variables that are related to the study variable Y. To develop more reliable estimates of the population small area means $\theta_1, \ldots, \theta_m$, Fay and Herriot (1979) proposed the following hierarchical model for EB prediction of θ_i, $i = 1, \ldots, m$.

(I) **Sampling model**: $Y_i|\theta_1, \ldots, \theta_m, \beta, \sigma_v^2 \overset{\text{ind}}{\sim} N(\theta_i, D_i)$, $i = 1, \ldots, m$;

(II) **Linking model**: $\theta_i|\beta, \sigma_v^2 \overset{\text{ind}}{\sim} N(x_i^T \beta, \sigma_v^2)$, $i = 1, \ldots, m$,

where the D_i's are known sampling variances of the direct estimators Y_i's and the $x_i = (1, x_{i2}, \ldots, x_{ip})^T$'s are $p \times 1$ vectors of covariates associated with Y_i. The regression parameter $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ and the model error variance σ_v^2 are the model parameters. This model can be expressed as a mixed linear model by writing

$$Y = X\beta + v + e,$$

where $e = (e_1, \ldots, e_m)^T \sim N(0, D)$, which is independent of $v = (v_1, \ldots, v_m)^T \sim N(0, \sigma_v^2 I_m)$. Also, $Y = (Y_1, \ldots, Y_m)^T$, $D = \text{diag}\,(D_1, \ldots, D_m)$, and $X = [x_1, \ldots, x_m]^T$.

The above hierarchical model can also be written as

$$Y_i = \theta_i + e_i, \ \ \theta_i = \beta_{i1} + \sum_{j=2}^{p} \beta_j x_{ij}, \quad i = 1, \ldots, m, \tag{1}$$

$$\Rightarrow Y_i = \beta_{i1} + \sum_{j=2}^{p} \beta_j x_{ij} + e_i, \ \ i = 1, \ldots, m,$$

where the e_i's, defined earlier, are sampling errors of direct estimators with $e_i \overset{\text{ind}}{\sim} N(0, D_i)$, $i = 1, \ldots, m$ and the θ_i's are expressed as a sum of a *random* intercept β_{i1} and a linear function of the covariates $x_{i2}, \ldots, x_{ip}$ with *non-random* regression coefficients $\beta_2, \ldots, \beta_p$. The random intercept accounts for error in explaining the

θ_i's by the linear regression on the covariates. The Fay–Herriot model assumes that β_{i1}'s are independently and identically distributed as $N(\beta_1, \sigma_v^2)$.

In their application of shrinkage estimation of per capita income for small places, Fay and Herriot (1979) used the EB approach. Assuming the model parameters β and σ_v^2 were known, based on the hierarchical model in (I) and (II) above, Fay and Herriot (1979) first obtained the Bayes estimators of $\theta_i, i = 1, \ldots, m$. The Bayes "estimator" of θ_i, given by

$$\theta_i^B = Y_i - \frac{D_i}{D_i + \sigma_v^2}(Y_i - x_i^T \beta),$$

shrinks the direct estimator Y_i to the regression function $x_i^T \beta$. However, since the model parameters, β and σ_v^2, are unknown, these Bayes estimators of $\theta_1, \ldots, \theta_m$ could not be used. To obtain usable version from these estimators, they estimated the model parameters from the marginal distribution of $Y_1, \ldots, Y_m$. They used these estimates in place of β and σ_v^2 in the Bayes "estimators" of $\theta_1, \ldots, \theta_m$, resulting in the EB estimators of $\theta_1, \ldots, \theta_m$.

As an alternative to EB estimation, some practitioners use HB estimation. To deal with the unknown model parameters, a hierarchical Bayesian assigns a prior distribution to them. The resulting posterior distribution of the unknown model parameters is used to integrate them out from the Bayes estimators, described above. The HB estimation method explicitly accounts for the estimation error of the model parameters. However, the plug-in EB method suffers from underestimation of associated measures of uncertainty of the EB estimators since the method does not automatically account for estimation error of the model parameters. Extensions to EB have attempted to incorporate the extra uncertainty due to estimating the parameters (Morris 1983). It is well documented in the literature that EB predictors are also the empirical best linear unbiased predictors (EBLUPs) of the small area means. Accurate approximation of the mean squared error (MSE) of the EBLUPs and estimation of the MSE have been extensively discussed in Rao and Molina (2015, Chap. 5); see also Prasad and Rao (1990), Datta and Lahiri (2000), and Datta et al. (2005).

While the HB method facilitates use of prior information for the model parameters, in many applications no specific proper prior distribution may be available. Even in such cases, the method is found to be beneficial by using vague or diffuse or noninformative priors. Due to a rapid growth in computing capacity in the past few decades, implementation of HB methods has been relatively easily achieved. As a result, the HB methods, even with diffuse priors, are found to be viable for many complex scientific studies. In a parallel development to model-based estimation in survey sampling, the HB methods have also gained popularity in small area estimation.

We now introduce the HB version of the Fay–Herriot (FH) model based on a class of noninformative priors that has been extensively used for the model parameters.

(I) **Sampling model**: $Y_i|\theta_1, \ldots, \theta_m, \beta, \sigma_v^2 \overset{\text{ind}}{\sim} N(\theta_i, D_i)$, $i = 1, \ldots, m$;
(II) **Linking model**: $\theta_i|\beta, \sigma_v^2 \overset{\text{ind}}{\sim} N(x_i^T\beta, \sigma_v^2)$, $i = 1, \ldots, m$,
(III) **A popular noninformative improper prior density function for β and σ_v^2**:

$$\pi(\beta, \sigma_v^2) = (\sigma_v^2)^{-\alpha} \text{ for } \beta \in \mathbb{R}^p, 0 < \sigma_v^2,$$

where $1 - (m - p)/2 < \alpha < 1$.

The above noninformative prior density function does not integrate to 1, unlike any regular probability density function (pdf). Thus, it is termed as an improper prior, and it is required to verify that the resulting posterior pdf will be a proper pdf; that is, it will integrate to 1 after being appropriately normalized. Adapting arguments of Theorem 2 in Datta and Smith (2003), the resulting posterior pdf of β, σ_v^2 is a proper pdf provided $1 - (m - p)/2 < \alpha < 1$. The prior corresponding to $\alpha = 0$, which is a uniform prior for the model parameters, is a popular choice. We use this prior in our illustrative application.

In the FH model in (1), the random intercept term is intended to account for variability among the θ_i's that is left unexplained by the covariates. It allows for a constant variability of the θ_i around its mean $x_i^T\beta = \beta_1 + \sum_{j=2}^p \beta_j x_{ij}$. In that spirit, it is conceivable that the variability of the θ_i's around the regression line may not be constant, but it may also be dependent on covariates. This leads to the random regression coefficient (RRC) model. For unit-level small area estimation models, Prasad and Rao (1990) suggested the RRC model (see also Datta and Ghosh 1991). These random coefficients can be viewed as similar to interaction effects between the covariates and the small areas. For the area-level data, we propose below a RRC model, which can be viewed as a generalization of the FH model. Suppose $q - 1$ ($q \le p$) covariates have random regression coefficients in the mean function of θ_i. Without loss of generality, let $x_{i2}, \ldots, x_{iq}$ be these $q - 1$ covariates. We introduce the HB version of the RRC model below.

(I) **Sampling model**: $Y_i|\theta_1, \ldots, \theta_m, \beta, \Sigma_v \overset{\text{ind}}{\sim} N(\theta_i, D_i)$, $i = 1, \ldots, m$,
(II) **Linking model**: $\theta_i|\beta, \Sigma_v \overset{\text{ind}}{\sim} N(x_i^T\beta, z_i^T\Sigma_v z_i)$, where $z_i = (1, x_{i2}, \ldots, x_{iq})^T$, $i = 1, \ldots, m$ and Σ_v is a $q \times q$ positive definite matrix,
(III) **A noninformative prior for the model parameters**:

$$\pi(\beta, \Sigma_v) = g(\Sigma_v), \tag{2}$$

where g is a suitable function defined below. We introduce a noninformative prior on the model parameters below. Let $\Psi = \Sigma_v^{-1}$ be the precision matrix and define

$\tau_{jk} = \psi_{jk}/(\psi_{jj}\psi_{kk})^{1/2}$, where $\Psi = \{\psi_{jk}\}_{1 \le j,k \le q}$. Then, $\tau_{11}, \ldots, \tau_{qq}$ are all 1. Let $T = \{\tau_{jk}\}_{1 \le j,k \le q}$. We propose a noninformative prior given by

$$\pi(\beta, \psi_{11}, \ldots, \psi_{qq}, \tau_{12}, \ldots, \tau_{1q}, \ldots, \tau_{q-1,q}) = \prod_{j=1}^{q} \psi_{jj}^{-a_j/2} I(T \text{ is p.d.}), \qquad (3)$$

where the a_j's (>2) are suitably chosen and $I(\cdot)$ is the indicator function. Here, $\beta \in \mathbb{R}^p$, and p.d. stands for positive definite. We note that the precision matrix is parameterized in terms of its diagonal elements and τ_{jk}, $1 \le j < k \le q$. We prove the following theorem in the appendix.

Theorem 1 *For $a_1 > 2, \ldots, a_q > 2$ and under some upper bound conditions on $a_1, \ldots, a_q$ (noted explicitly in the proof), the posterior pdf corresponding to the noninformative prior in (3) is a proper pdf.*

Remark 1 We note that the HB FH model is a special case with $q = 1$ and $a_1 = 4 - 2\alpha$.

In this paper, we pursued an HB approach for the RRC model. However, a frequentist approach to this problem via EB approach or a mixed model formulation can be pursued. Frequentist estimation of the variance parameters may not be easy, as it is not obvious how to define $q(q+1)/2$ estimating equations for the components of Σ_v. Also, the maximum likelihood (ML) approach may present a hurdle since the likelihood (residual or profile likelihood) function may be quite flat, which would make convergence of the ML computing rather slow.

3 Four-Person Family Median Income Estimation

In this section, we consider an application of estimating four-person median incomes by states, i.e., median income of four-person households for fifty states and Washington, D.C. The US Department of the Health and Human Services needs accurate estimates of medians for the state distributions of income of four-person families to implement a cash welfare program to provide energy assistance benefits to low-income American families. For over twenty years through the mid-1990s, the U.S. Census Bureau continued to calculate these state-level estimates annually using the Current Population Survey (CPS) data. The statewide samples available from the CPS to compute direct estimates of state characteristics are not sufficiently accurate for some states due to smallness of available samples.

Let Y_i be the direct estimate of 1989 four-person median income of the ith state. We consider two auxiliary variables x_{i2} and x_{i3}, where x_{i2} is the median income for 1979 collected from the 1980 census and x_{i3} is the adjusted census median income. Adjusted census median incomes are obtained by utilizing 1979 and 1989 per capita incomes (PCIs) from the Bureau of Economic Analysis of the US Department of

Commerce such that $x_{i3} = (\mathrm{PCI}_{i,1989}/\mathrm{PCI}_{i,1979})x_{i2}$, $i = 1, \ldots, 51$. Thus, we have $x_i = (1, x_{i2}, x_{i3})^T$, $i = 1, \ldots, 51$. Here $p = 3$, and we take $q = p$. Then, the model parameters are given by β (a 3×1 vector) and Σ_v (a 3×3 p.d. matrix). Since the direct estimates and covariates are large values, we standardize the direct estimates and scale auxiliary variables to prevent numerical overflow (underflow). We fit the HB FH model and the HB RRC model to the data using rstan (Stan Development Team 2018). We consider two prior pdfs for the FH model, $\alpha = 0, 0.75$, and for the RRC model we consider two prior pdfs with $a_j = 2.1, 2.5, j = 1, 2, 3$ (close to their lower bound for propriety). We denote results for these models as FH_0, $\mathrm{FH}_{0.75}$, $\mathrm{RRC}_{2.1}$, and $\mathrm{RRC}_{2.5}$, respectively. Posterior means of the model parameters are given in Table 1. As we can see from the left panel of Fig. 1, the four models result in very similar posterior distributions in that a large portion of their 95% credible intervals overlap for the common parameters. Also, we can see that the posterior means of $\sigma_{v,22}$ and $\sigma_{v,33}$ are close to 0, and their lower bounds of the intervals get almost 0. This indicates that regression coefficients do not vary much across the states. Figure 2 illustrates 95% credible intervals of the predictions under FH_0 and $\mathrm{RRC}_{2.5}$. Red (blue) lines and circles (triangles) represent 95% credible intervals and posterior means of the four-person family median incomes under the FH_0 ($\mathrm{RRC}_{2.5}$) model. Black diamonds are the true median incomes for 1989 collected from the 1990 census. Both models underestimate the median incomes of California and Delaware, and the credible intervals miss (fall below) the true values.

We compare both the models according to several criteria. We first compare posterior standard deviations which are illustrated in the right panel of Fig. 1. The $\mathrm{RRC}_{2.5}$ model shows smaller posterior standard deviations than those of the FH_0 model in 31 states. However, for the states of Alaska, Connecticut, and Maryland, the FH_0 model has much smaller posterior standard deviations. On an average, posterior standard deviation of the $\mathrm{RRC}_{2.5}$ model is 1.3% smaller than the FH_0 model. We also compare the average length (AL) of 95% credible intervals and several deviation criteria. For the true median income θ_i for the ith state, let $\hat{\theta}_i$ be any estimate of θ_i, $i = 1, \ldots, 51$. We compute average absolute deviation (AAD) $51^{-1} \sum_{i=1}^{m} |\hat{\theta}_i - \theta_i|$, average squared deviation (ASD) $51^{-1} \sum_{i=1}^{51} |\hat{\theta}_i - \theta_i|^2$, average absolute relative deviation (AARD) $51^{-1} \sum_{i=1}^{51} |(\hat{\theta}_i - \theta_i)/\theta_i|$, and average squared relative deviation (ASRD) $51^{-1} \sum_{i=1}^{51} \{(\hat{\theta}_i - \theta_i)/\theta_i\}^2$. These deviation criteria are listed in Table 2. We see a clear improvement for the model-based estimators over the direct estimators. However, there is only a slight improvement for the RRC models over the FH mod-

Table 1 Posterior means of the model parameters

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_{v,11}$	$\hat{\sigma}_{v,22}$	$\hat{\sigma}_{v,33}$	$\hat{\rho}_{12}$	$\hat{\rho}_{13}$	$\hat{\rho}_{23}$
FH_0	−0.037	0.030	0.791	0.184	–	–	–	–	–
$\mathrm{FH}_{0.75}$	−0.042	0.032	0.789	0.160	–	–	–	–	–
$\mathrm{RRC}_{2.1}$	−0.041	0.056	0.774	0.142	0.009	0.015	0.053	0.007	−0.014
$\mathrm{RRC}_{2.5}$	−0.039	0.085	0.758	0.133	0.035	0.044	0.130	−0.021	−0.074

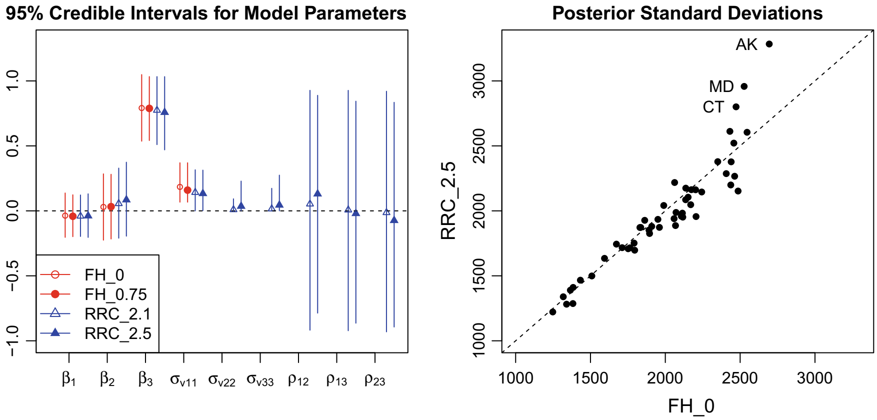


Fig. 1 Ninety-five percentage credible intervals for model parameters (left) and posterior standard deviations of θ_i's (right) under the FH_0 and $\mathrm{RRC}_{2.5}$ models

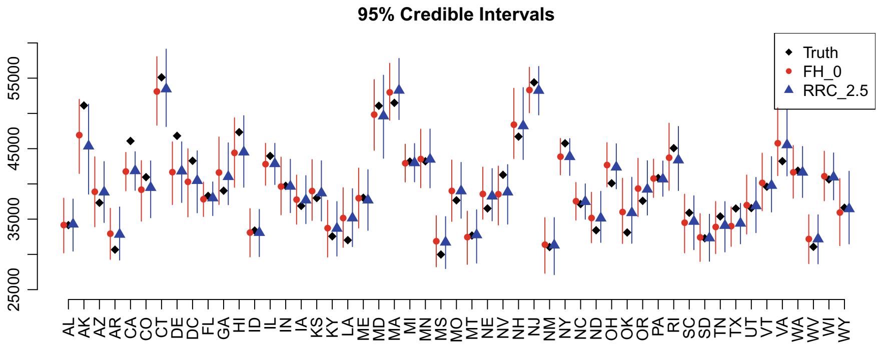


Fig. 2 Ninety-five percentage credible intervals under the FH_0 and $\mathrm{RRC}_{2.5}$ models

els. Also, we have seen from Table 1 that estimated random effect variances are very small except for the random intercept, $\hat{\sigma}_{v,11}$. This indicates that regression parameters do not vary much across the states. Thus, in this example, it is reasonable to use the FH model since the gain from using the more complicated model is rather limited.

4 A Simulation Study

To investigate the effectiveness of the RRC model further, we conducted a simulation study. In our simulation, we used the data for the covariates from the median income application in Sect. 3. To reduce computational burden, in our simulation we used $q = 2$ (instead of $q = 3$ that was in the data analysis) and $p = 3$. That is, we used only one random regression coefficient, corresponding to the most significant covariate

Table 2 Model comparison criteria

Estimate	AAD	ASD	AARD	ASRD	AL
Direct	2928.824	13811122	0.074	0.008	11424.650
FH_0	1507.303	3682226	0.038	0.002	7856.558
$FH_{0.75}$	1438.580	3386631	0.036	0.002	7614.605
$RRC_{2.1}$	1405.772	3310202	0.035	0.002	7621.411
$RRC_{2.5}$	1435.107	3614879	0.035	0.002	7815.770

Table 3 Model parameters for simulation

	β_1	β_2	β_3	$\sigma_{v,11}$	$\sigma_{v,22}$	ρ_{12}
FH	−0.037	0.027	0.793	0.185	–	–
RRC	−0.044	0.070	0.755	0.062	14.706	0.033

found in the analysis. Specifically, we set $z_i = (1, x_{i3})^T$. We considered two scenarios; data are generated from the FH model and the RRC model, respectively.

In order to simulate from a model that will closely mimic the reality, we need some knowledge about model parameters for each model. To suitably choose these model parameter values to reflect the reality, we seek guidance from our data analysis in Sect. 3. Our data analysis showed a marginal improvement of the $RRC_{2.5}$ model over the FH_0 model. This is due to very small variances of the random regression coefficients. In order to have non-negligible contribution of the random regression coefficient in our simulation study, we considered a large value for the variance of the random regression coefficient compared to that of the random intercept term. We first fit the models to the data and obtained the posterior means of the model parameters. We used these values as guidance to our choice of model parameters in the simulation study. In our simulations from the $RRC_{2.5}$ model, we set $\sigma_{v,11} = \hat{\sigma}_{v,22}$, $\sigma_{v,12} = 10\hat{\sigma}_{v,12}$, and $\sigma_{v,22} = 100\hat{\sigma}_{v,11}$, where $\hat{\sigma}_{v,11}$, $\hat{\sigma}_{v,12}$, and $\hat{\sigma}_{v,22}$ are obtained by fitting the $RRC_{2.5}$ model to the four-person median income data with $z_i = (1, x_{i3})^T$. This allows the random regression coefficient to dominate the model error variance. It also preserves the correlation between the random intercept and the random regression coefficient. Finally, we also multiplied the sampling variances by 100 in order to approximately preserve the ratio between model and sampling error variances. The parameter values are given in Table 3. To compare performance of the FH and RRC models, we considered 30 replicated data sets and computed the deviation criteria (introduced in the last section) and overall coverage probabilities of 95% credible intervals by averaging both over the areas and the data sets. Table 4 lists these values.

When data are generated from the FH model, the FH_0 model shows slightly better performance than the $RRC_{2.5}$ model but the differences are not significant. Under the true RRC model, the $RRC_{2.5}$ model outperforms the FH_0 model in all the deviation criteria. Also, it has larger coverage probabilities while the posterior standard deviations are approximately 13% smaller on average, which are not shown

Table 4 Average of various deviation criteria and coverage probabilities

True model	Fitted model	AAD	ASD	AARD	ASRD	Coverage
FH	FH_0	2.072	7.040	1.577	7.236	0.952
	$RRC_{2.5}$	2.077	7.184	1.557	8.825	0.944
RRC	FH_0	2.232	8.993	4.003	92.246	0.936
	$RRC_{2.5}$	2.079	8.391	2.698	27.137	0.959

here. In conjunction with our finding in the real data analysis given in the previous section, the RRC model is as much good as the FH model when all the states share common regression parameters. This is an expected result since the FH is a submodel of the RRC model. The RRC model can fit any data generated under the FH model, although it requires an extra computing effort. However, when the regression coefficients vary across the states, the FH model will be a misspecified model. As a result, the reported posterior standard deviations for the small area means for some areas will be underestimates and for other areas they will be overestimates. Similarly, the credible intervals of the small area means for some areas will be overly short and for some overly long.

5 Concluding Remarks

In this article, we extended the FH model by replacing the constant model error variance assumption by a heteroscedastic variance assumption. The heteroscedasticity that we modeled results from the randomness of the regression coefficients in modeling the small area population means. The heteroscedasticity induced by this assumption models the variance by an appropriate quadratic function of the covariates. We developed a noninformative HB procedure for the RRC model. We applied the FH model and the RRC model to predict four-person family median incomes for the US states. This application and limited simulation study demonstrate the usefulness of the more general model in better prediction of small area means. The RRC model may also be a good check on the standard FH model to see if all areas follow the same regression function (regression plane). Sometimes, there are subsets of areas that may have a different regression relationship with the covariates, such as due to a missing interaction effect (area and covariate). The RRC model can pick up these departures and provide a more robust modeling framework.

Disclaimer This report is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the author(s) and not necessarily those of the U.S. Census Bureau, or the University of Georgia.

6 Appendix

Proof of Theorem 1 In our RRC model, $Y_i = \theta_i + e_i, \theta_i = x_i^T \beta + z_i^T v_i, i = 1, \ldots, m$, where $e_1, \ldots, e_m, v_1, \ldots, v_m$ are all independently distributed with $e_i \sim N(0, D_i)$, $i = 1, \ldots, m,\ x_i = (1, x_{i2}, \ldots, x_{ip})^T,\ p \geq 2,\ z_i = (1, x_{i2}, \ldots, x_{iq})^T,\ q \geq 2,\ \beta = (\beta_1, \ldots, \beta_p)^T$, $v_i \overset{iid}{\sim} N(0, \Psi^{-1})$ and Ψ is a $q \times q$ positive definite precision matrix. We assume that $q \leq p$ and, if necessary, $Y = (Y_1, \ldots, Y_m)^T$, $\tilde{x}_l = (x_{1l}, \ldots, x_{ml})^T$, $l = 2, \ldots, p$ will be standardized with suitable locations and scales. We note that $\Psi = \{\psi_{jk}\}_{1 \leq j,k \leq q}$ and $T = \{\tau_{jk}\}_{1 \leq j,k \leq q}$ are defined as in the main text. We consider the noninformative prior given in (3).

Without loss of generality, we arrange the data set so that the first p observations lead to a nonsingular design matrix X_* given by $X_* = [1_p, \tilde{x}_{*2}, \ldots, \tilde{x}_{*p}]$, where 1_p is a $p \times 1$ vector of ones, $\tilde{x}_{*l} = (x_{1l}, \ldots, x_{pl})^T, l = 2, \ldots, p$. We use these p observations to estimate β by $b = X_*^{-1} Y_*$, where $Y_* = (Y_1, \ldots, Y_p)^T$. Let $Y_0 = (Y_{p+1}, \ldots, Y_m)^T$. Note that $b|\beta, \Psi \sim N(\beta, \Phi(\Psi))$, where Φ is a $p \times p$ positive definite matrix depending on Ψ. Under this model and the prior, the joint posterior pdf of β and $\psi_{11}, \ldots, \psi_{qq}, \tau_{jk}, 1 \leq j < k \leq q$ is given by

$$
\begin{aligned}
\pi(\beta, \psi_{11}, \ldots, \psi_{qq}, \tau_{jk}, 1 \leq j < k \leq q|Y) &= cN(b|\beta, \Phi(\Psi)) \\
&\quad \prod_{i=p+1}^{m} (D_i + z_i^T \Psi^{-1} z_i)^{-1/2} \\
&\quad \times \exp\left\{-\frac{(Y_i - x_i^T \beta)^2}{2(D_i + z_i^T \Psi^{-1} z_i)}\right\} \prod_{j=1}^{q} \psi_{jj}^{-a_j/2}\, I(T \text{ is p.d.}) \\
&\leq cN(\beta|b, \Phi(\Psi)) \prod_{i=p+1}^{m} (D_i + z_i^T \Psi^{-1} z_i)^{-1/2} \prod_{j=1}^{q} \psi_{jj}^{-a_j/2}\, I(T \text{ is p.d.}),
\end{aligned}
\tag{4}
$$

where $Y = (Y_1, \ldots, Y_m)^T$. Integrating both sides of (4) with respect to β, we get

$$
\begin{aligned}
&\pi(\psi_{11}, \ldots, \psi_{qq}, \tau_{jk}, 1 \leq j < k \leq q|Y) \\
&\qquad \leq c \prod_{i=p+1}^{m} (D_i + z_i^T \Psi^{-1} z_i)^{-1/2} \prod_{j=1}^{q} \psi_{jj}^{-a_j/2}\, I(T \text{ is p.d.})
\end{aligned}
\tag{5}
$$

Note that $z_i^T \Psi^{-1} z_i \geq z_{ij}^2 / \psi_{jj}, j = 1, \ldots, q$. From the $m - p$ small areas, select n_j small areas for which $z_{ij}^2 > 0$. We need suitable conditions on $n_1, \ldots, n_q$, where $\sum_{j=1}^{q} n_j = m - p$. By these observations, we have that

$$\prod_{i=p+1}^{m} (1 + z_i^T \Psi^{-1} z_i)^{1/2} \le C(1 + \psi_{11}^{-1})^{-n_1/2} \dots (1 + \psi_{qq}^{-1})^{-n_q/2}. \tag{6}$$

We now use (6) in (5), transform ψ_{jj} to u_j by $u_j = \psi_{jj}^{-1}, j = 1, \dots, q$, and keep τ_{jk}, $1 \le j < k \le q$, fixed. Then, integration of both sides of (6) will lead to

$$\begin{aligned}\int \pi(\psi_{11}, \dots, \psi_{qq}, \tau_{jk}, 1 \le j < k \le q|Y) d\psi_{11}, \dots, d\psi_{qq} \\ \le C \prod_{j=1}^{q} \int_0^\infty u_j^{-2+a_j/2} (1 + u_j)^{-n_j/2} du_j \\ = c^* < \infty \end{aligned} \tag{7}$$

provided $-2 + a_j/2 > -1, -2 + a_j/2 - n_j/2 < -1, j = 1, \dots, q$, i.e., $a_j > 2, n_j > a_j - 2, j = 1, \dots, q$.

Note that c^* is finite and free from τ_{jk}, $1 \le j < k \le q$. Thus, c^* will also be integrable with respect to τ_{jk}, $1 \le j < k \le q$. Consequently, subject to the conditions $a_j > 2, n_j > a_j - 2, j = 1, \dots, q$, the posterior pdf for our RRC model will be proper.

References

Datta, G. S., Fay, R. E., & Ghosh, M. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation. In *Proceedings of the Seventh Annual Research Conference of the Bureau of the Census* (Vol. 7, pp. 63–79). US Department of Commerce, Bureau of the Census.

Datta, G. S., & Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics*, *19*, 1748–1770.

Datta, G. S., Ghosh, M., Nangia, N., & Natarajan, K. (1996). Estimation of median income of four-person families: A Bayesian approach. In D. A. Berry, K. M. Chaloner, & J. K. Geweke (Eds.), *Bayesian analysis in statistics and econometrics* (Chap. 11) (pp. 129–140). New York: Wiley.

Datta, G. S., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, *10*, 613–627.

Datta, G. S., Rao, J. N. K., & Smith, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, *92*, 183–196.

Datta, G. S., & Smith, D. D. (2003). On propriety of posterior distributions of variance components in small area estimation. *Journal of Statistical Planning and Inference*, *112*, 175–183.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association*, *68*, 117–130.

Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, *74*, 269–277.

James, W., & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 361–379). University of California Press.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, *78*, 47–55.

Prasad, N. G. N., & Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, *85*, 163–171.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation*. New York: Wiley.

Stan Development Team. (2018). RStan: The R interface to Stan. *R Package Version*, *2*(17), 3.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 197–206). University of California Press.

Entropy-Based Consumption Diversity—The Case of India

Manisha Chakrabarty and Jayanta Mandi

Abstract In recent years, there has been growing research in analysing the spending diversification of households in applied demand analysis using disaggregated household-level data. Taking cue from Engel's (Die Lebenskosten Belgischer Arbeiter Familien frfther und jetzt, Bulletin de l'institut international de statistique, tome IX, premiere livraison, Rome, 1895) findings that large share of income is spent on basic goods such as food for lower-income decile, the applied demand analysts also observed that with increasing income, there is an increase in spending on other non-food commodities, implying a hierarchical structure of consumption pattern. Evidences also supported positive correlation between household income and the dispersion of household spending both at cross-country-level analysis and at household-level analysis. These findings justify the use of consumption-based measures such as food share (Anand and Harris in Am Econ Rev 84:226–231, 1994) and consumption diversity (Clements et al. in Empirical Econ 31:1–30, 2006; Chai et al. in J Econ Surv 29:423–440, 2014) as indicators of household welfare. In this paper, we attempt to examine the stylized facts of behavioural heterogeneity across disaggregated commodity groups by employing entropy-based Theil's measure. Using National Sample Survey household expenditure data of urban sector of four major states of India for the year 2011–2012, we show the extent to which income (measured through monthly per capita expenditure and thereby controlling household size) and other demographic characteristics such as number of children explain the variation in consumption diversity. We also capture commodity group-wise variations for explaining consumption diversity within commodity group by considering not only inherent characteristics of commodity groups such as income elasticity as control variables but also as random coefficient models varying randomly across commodity groups. The incorporation of between-commodity heterogeneity via random coefficient model is our contribution in this literature on consumption heterogeneity. The random coefficient models establish significant heterogeneity across commodity

M. Chakrabarty (✉)
Indian Institute of Management Calcutta, Kolkata, India
e-mail: mchakrabarty@iimcal.ac.in

J. Mandi
Data Science, Vrije University Brussels, Brussels, Belgium
e-mail: jayanta.mandi@vub.be

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_24

groups, mainly through intercept change, but not so much in income and demographic factors' effect.

Keywords Theil's entropy measure · Random coefficient model · Consumption diversity

1 Introduction

The study of the consumer expenditure and income (the Engel curve) has been at the centre of welfare analysis since the early studies of Engel (1857, 1895), Allen and Bowley (1935) and Prais and Houthakker (1955). Out of these, one could consider Engel's work in 1857 as the first and most famous of all statistical analyses of budgets. But a complete description of consumer behaviour sufficient for welfare analysis requires a specification of both Engel's curve and relative price effects consistent with preference maximization. In addition to the elasticity estimates, the additive complete demand systems provide estimates of welfare indicators such as cost of living indices. An important contribution in the context of demand system estimation was given by Stone (1954), Working and Leser (1943), Christensen et al. (1975), Muellbauer (1976), Deaton and Muellbauer (1980) and Jorgenson et al. (1982). But further analyses of household budget surveys have pointed to more curvature in the income responses than permitted by the above-mentioned demand system equations, and therefore, a more flexible utility maximization consistent systems such as QUAIDS (Banks et al. 1997), QES (Howe et al. 1979), AIDADS (Cranfield et al. 2003) and EASI (Lewbel and Pendakur 2008) were developed in the consumption analysis. Still it has been observed that existing household demand function fails to explain several other features of observed variation in household consumption behaviour (Lewbel 2008) in income dimension. Several such stylized facts about household consumption expenditure pattern have been studied in the literature describing incidence such as the hierarchical structure of demand pattern mostly at aggregate country level (see, e.g., Prais (1952) for the nonlinear curvature in Engel's curve and heterogeneity in consumption for microlevel Dutch data; Theil and Finke (1983) for aggregate data of 30 countries including India for the year 1975; Jackson (1984) for US consumer expenditure data of 1972–1973 by exploring Engel's curve for variety; Pasinetti (1981) for analysing economic growth which relates economic growth to hierarchical expenditure structure; Falkinger and Zweimüller (1996) for aggregate country-level data for 57 countries; Thiele and Weiss (2003) for Germany; Chattopadhyay et al. (2009) for hierarchical demand system estimation using household-level data of West Bengal for the year 1993–1994, and Chai and Moneta (2012)] for British household over the period of 1960–2000 relating expenditure dispersion with income, etc.

The threshold justification explains the fact that poor household who cannot fulfil the threshold level of fundamental needs will not buy more luxury items. Hence, an increase in variety of consumption can be considered to be linked with rise in welfare (Clements et al. 2006; Chai et al. 2014) as it is associated with higher level of income per capita. This idea of hierarchical needs is also backed by psychological research of Canova et al. (2005). One implication of this is that people with higher income display greater diversity in expenditure, and consumption of people with lower income is mainly driven by homogeneous expenditure pattern dominated by food consumption (Chai et al. 2014). This evidence is corroborated by number of studies of cross-country demand analysis (Theil and Finke 1983; Falkinger and Zweimüller 1996; Clements et al. 2006). Interestingly, not only the overall diversity of broad categories of expenditure increases but also diversity of within-group expenditure increases with increase in income. For instance, with the rise in income, there is a shift of consumer's share of expenditure from food items to luxury items and also at the same time a lower-income household may consume only rice within cereal group, but a higher income household will consume more diversified range of products within cereal goods. In essence, the diversity of expenditure represents the well-being of a household which can be explained from the textbook knowledge of utility function describing monotonicity of utility function with respect to consumption expenditure. Hence, increasing mean and variance of the consumption distribution serve as a normative foundation of welfare. Also the empirical evidences validate decreasing food share with increasing income and thereby increase in spending diversity with increasing expenditure on several other disaggregated commodities belonging to non-food categories [see Houthakker (1957) on aggregate basis, Thomas and Strauss (1997) at microlevel]. The empirically supported positive correlation between household income and the dispersion of household spending both at cross-country-level analysis and at household-level analysis justifies the use of consumption-based measures such as consumption diversity (Clements et al. 2006; Chai et al. 2014) as indicators of household welfare because of its association with affluence, measured through increasing income. Yet this spread could also come due to omitted variables such as demographic attributes and or heterogeneity in tastes. We, therefore, attempt to control for demographic attributes such as number of children and household size by considering per capita expenditure,[1] following the literature on demographics-augmented demand system analysis (see Ray 1986).Yet, recent studies have shown ambiguity of such consumption-based welfare measure because of social cost associated with consumption pattern and also due to change in preference pattern coming from expanding possibilities (Shaw and Newholm 2002; Witt 2017). The association of consumption diversity of food with socio-economic status including income and with demographic attributes such as age has also been explored also in medical science (see, e.g., Morseth et al. 2017). Also now the existence of heterogeneity of consumption pattern across commodities is well recognized in the demand system literature (see Banks et al. 1997) which led us to incorporate the commodity-specific characteristics both as control variable and as random component to explain consumption diversity for each good. We, therefore, capture commodity group-wise

[1] We could not consider other demographic attributes such as female headed, location of households, number of earning members because of either very low frequency of such households or absence of data in NSSO consumption expenditure survey. Yet we would like to mention that some unobserved heterogeneity affecting consumption diversity via social network effect or locality effects are captured through the state fixed effects.

variations for explaining consumption diversity within commodity group by considering not only inherent characteristics of commodity groups as control variables but also random coefficient models varying randomly across commodity groups (see Blacklow et al. 2006, for commodity-specific demographic effects in demand system estimation).

Our aim is, therefore, to explain this phenomenon of diversity in household spending behaviour in terms of income, household demographic characteristics such as number of children in households as children plays a crucial role in family purchases (Flurry 2007) and account for commodity-specific characteristics. Our study focuses on India, which can serve as a rich source of data for exploring such behavioural pattern with specific features such as 22% people are living below the poverty line in 2011–2012, the period of our study, and at the same time, India is very much dispersed in terms of consumption habits of people spanning the vast geographical spectrum. Also several studies mentioned above examining the stylized facts of consumption diversity looked into cross-country demand analysis except Chai and Moneta (2012), whereas in this paper, we explore household-level disaggregated consumption data like Chai et al. (2017). But to the best of our knowledge, this is the first time an attempt has been made to investigate the causal impact of income in spending diversity after controlling for several observed such as demographic attributes, income elasticity of commodity and unobserved mediating factors using random coefficient model.

To measure the spending diversity by households, we adopted a Theil (1967) entropy-based measure of diversity. Especially, we fit a random coefficient model with diversity measure as dependent variable and number of children in household, a dummy variable indicating the nature of commodity based on income elasticity, i.e. luxury or necessity and monthly per capita expenditure considered as explanatory variables. It might be true that diversity of consumption across commodity groups is very different in a random way but within-group pattern of diversity across households might be similar. Hence, to examine whether this is indeed the case in order to explain variation in consumption diversity, which is considered to be one of the indicators of household welfare, we consider random coefficient model in this multilevel framework.

The paper is structured in the following way: Sect. 2 describes the data and variables considered in our study; Sect. 3 depicts the entropy measure of consumption diversity considered in this paper along with the econometric model and hypotheses to be tested in our analysis; Sect. 4 presents the results, and finally, in Sect. 5, conclusions are drawn.

2 Data and Variable Description

2.1 Data

Our source of data is the 68th round household consumer expenditure survey conducted by Indian National Sample Survey Organization (NSSO), covering the period from July, 2011 to June, 2012. NSSO has adopted a multistage sampling design where first-stage units (FSUs) are the census villages in the rural sector and urban blocks in the urban sector and ultimate stage units (USUs) are households. The reporting period for collection of data for cereals, pulses, milk, milk products, sugar and salt, and miscellaneous goods is the last 30 days and for edible oil, egg, fish and meat, vegetables, fruits, spices, beverages, processed food and pan, tobacco and intoxicants is last 7 days ending on the day preceding the date of enquiry. For clothing, bedding and footwear items, the reporting period is 365 days. We convert all expenditure to monthly expenditure. Although the sample covers both rural and urban areas, we only consider samples from urban areas to downsize the impact of home-grown produce and also due to the fact that the availability of wide spectrum of goods is limited in rural sector. Also as reported by Government of India (2013), poverty rate in rural sector is much higher (25.7%) as opposed to 13.7% in urban India, and hence, it is less probable to observe consumption diversity in rural sector. The states considered for this study are Maharashtra (MH), Tamil Nadu (TN), Uttar Pradesh (UP) and West Bengal (WB). These four states belong to four different regions, namely west, south, north and east, respectively, of India with varying pattern of consumption style, particularly within the food group. Also, these states account for highest number of surveyed urban household in their respective region. In total, our study covers 12,708 households.

2.2 Variable Description

In the following, we describe the variables considered for our analysis.

State: As mentioned before, we consider four states, and as the consumption behaviour of people varies from state to state, we keep three state-dummy variables to capture the inter-regional variation of consumption. We consider Maharashtra (MH) as the reference state.

MPCE: Needless to say, the consumption behaviour of a household is greatly determined by the income of the household. Households with higher income are expected to buy wide range of products within a category as consumption variety increases consumer utility (Kahn and Wansink 2004). Hence, we use logarithmic value of monthly per capita expenditure as the main determining variable.

Number of children in a household: Existence of children in a household plays a crucial role in family purchases (Flurry 2007; Ray 1986; Lancaster and Ray

1998).With the increase in female labour force participation, children tend to spend more time in out-of-home settings and as result their choice tends to be more disparate than older member of families. Hence, the number of children with age below 15 years is considered another important determining variable.
Commodity: We have considered 14 broad groups of commodities with the details given in the appendix. As Engel's law states that with increase in income the budget share on food decreases, we consider detailed food group and also several other commodity groups belonging to non-food. Within the group "food", we consider several subgroups because Bennett's law (Bennett 1941) states that as income increases, the proportion of the budget spent on "starchy-staples" decreases. Bennett's law reflects a desire for dietary diversity. Hence, in order to capture an even distribution of consumption expenditure across categories, not only we look into food-versus-non-food expenditure, but also looked into several commodity groups within the broad "food" group.

In the non-food group, we do not consider commodity groups such as medical, education, durable goods, house rent and taxes because of specific nature in consumption of these goods. Some of the components within the class of expenditure on education, medical and house rent, and taxes are either compulsory or emergency driven and hence cannot be clubbed with other components. Also, for durable goods, there are repair and second-hand purchase components involved, and durable good consumption does involve an inter-temporal preference pattern, which is beyond the scope of this paper. We also have not considered consumption on tobacco intoxicants due to the fact that the consumption expenditure on this group is mainly driven by habit formation which can be predominantly present among lower-income class and therefore suffers from misreporting/zero consumption (see Sudarshan and Mishra 1999; Subramanian 2004).

The 14 commodity groups considered capture 75% of household total budget on average, as depicted in Table 1 in the result section. We model the between-commodity heterogeneity not only through observed characteristics of commodity groups such as income elasticity but also the unobserved heterogeneity through a random effect (see Maddala et al. 1971), described in the following section, captured by random intercept and random slope terms for all determining variables at the commodity level (see Blacklow et al. 2006, for the economic logic of varying effects of demographics across commodity groups in demand system).

Table 1 Descriptive statistics: mean values

	State (number of observations)	Mean budget share of 14 commodities together (%)	Number of children (sd)	MPCE (sd)
1	Uttar Pradesh (3066)	76.02	1.79 (1.64)	2207.86 (2194.17)
2	West Bengal (2659)	75.91	0.85 (1.07)	2973.84 (2774.96)
3	Maharashtra (3852)	76.39	1.13 (1.21)	3496.83 (3781.81)
4	Tamil Nadu (3131)	72.16	0.89 (1.01)	2722.26 (2201.78)

Income elasticity of the commodity group: The consumption pattern of commodities varies with the nature of commodity which could be partially captured by income responsiveness of the commodities. Heterogeneity of spending patterns across different goods can be represented through different functional form of Engel's curves (see Lewbel 1991; Banks et al. 1997). As we focus on diversity of consumption within several commodity groups where income is the main determining variable, we estimate income elasticity of all 14 commodity groups using the log-quadratic Engel (see Majumder 1992) at the state level and include these characteristics as a fixed effect variable.

For household h, the budget share on commodity i, w_i^h is modelled in terms of per capita total expenditure y_{h} as:

$$w_i^h = \alpha + \beta \log y_h + \gamma (\log y_h)^2 \tag{1}$$

Following the works of Majumder (1992), we calculate the average income elasticity of commodity group i for households within each state as (For details of expressions, see Appendix 2),

$$\eta_i = 1 + \frac{\beta_i + 2\gamma_i \overline{\log y}}{\overline{w_i}} \tag{2}$$

where bar indicates average over households within a state, i.e.

$$\overline{w_i} = \frac{\sum_{h=1}^{H} w_i^h}{\neq H}; \quad \overline{\log y} = \frac{\sum_{h=1}^{H} \log y^h}{\neq H} \tag{3}$$

We classify the commodity groups in two categories—*necessary* and *luxury*—as observed from the data and include one dummy variable accordingly in empirical model to account for any non-random pre-determined variability across commodity groups with necessary as the reference category of commodity. We also interact this dummy variable with our two main predictor variables, namely *MPCE* and number of *children*.

3 Diversity Measure and Random Coefficient Model

3.1 Entropy-Based Diversity Measure

As there exists a range of different approaches to measure spending diversity, we, in this paper, use widely used notion of entropy-based *Theil's* measure of diversity for each household. As a member of the generalized entropy class of inequality measures, Theil's index is a widely popular choice for measuring inequality because it possesses some desirable properties of inequality measures such as principle of transfers, subgroup consistency, subgroup decomposability and large sensitivity to income transfers from poor to rich (Shorrocks 1988; Conceição and Ferreira 2000).

For a household h and a particular commodity group i, with n number of items, if budget share of the items is $w_i^h, i \in [1, n]$ with $\sum_{i=1}^{n} w_i^h = 1 \forall h$, then entropy measure would be:

$$E_i^h = -\sum_{i=1}^{n} w_i^h \log w_i^h \tag{4}$$

3.2 Econometric Model

We consider 14 broad commodity groups; the detailed items contained within these groups are described in the appendix. We also consider number of children in households as households with children are more likely to diversify their expenditure across commodities. We consider three state dummies to control for inter-regional differences in spending pattern with MH as the reference category.

We consider only those commodity groups where the number of items within the group is greater than one as the diversity of a commodity group with only one commodity has no meaning. Then we calculate diversity index for these 14 commodity groups for each household. We consider this as dependent variable varying over households and commodity groups. We consider number of children in household and MPCE as main predictor variables, and state dummies, luxury-dummy variable (indicating whether a commodity is luxury), as control variables. We consider random coefficient model in the multilevel set-up where observations on diversity are recorded for each household within each commodity bracket. The rationale behind including commodity group as random effect variable is the diversity in expenditure within a group may be similar across different households, after controlling for MPCE and other determining variables at the household level, but not across commodity groups which can be purely random even after controlling for commodity's inherent characteristics such as expenditure elasticity. We consider the following simple two-level model to explain the variations of diversity measure—Theil's diversity index (as Y variable, described in Eq. (4) above), with the change in household's monthly per capita expenditure (MPCE). We also control for demographic attributes of household by considering number of children as another control variable and geographical regions such as state of residence. Our full model with random intercept and random slope is described below:

$$\begin{aligned} Y_{ih} = {} & \beta_{00} + \beta_{10} \log(\text{mpce}_h) + \beta_{20} \text{ number of children}_h + \sum_{k=1}^{3} \delta_k \text{ statedummy}_{kh} \\ & + \gamma_1 \text{ luxurydummy}_i + \gamma_2 \text{ luxurydummy}_i * \log(\text{mpce}_h) + \gamma_3 \text{ luxurydummy}_i \\ & * \text{ number of children}_h + v_{0i} + v_{1i} \log(\text{mpce}_h) + v_{2i} \text{ number of children}_h \\ & + E_{ih} \end{aligned} \tag{5}$$

with overall error follows $\varepsilon_{ih} \sim N(0, \sigma^2)$. $h \in H$ indicates number of households, and $i = 1, \ldots, 14$ denotes commodity groups with intercept and slopes are indexed with commodity groups allowing each commodity group to have unique intercept and slope coefficient as given below:

$$\beta_{0i} = \beta_{00} + \nu_{0i}$$
$$\beta_{1i} = \beta_{10} + \nu_{1i}$$
$$\beta_{2i} = \beta_{20} + \nu_{2i}.$$

Errors are independent with variance structure, $\upsilon_{ki} \sim N\left(0, \sigma_k^2\right)$, with $k = 0, 1, 2$.

As the above equation contains a composite error terms which are dependent within each commodity group, the components $\nu_{0i}, \nu_{1i}\nu_{2i}$ are common to every households within the group, heteroscedastic, because $\nu_{0i} + \nu_{1i} \log\left(\text{mpce}^h\right) + \nu_{2i}$ number of children^h vary across households. Hence, we use maximum likelihood method to get estimates which are described in the following result section. For comparison, we also present OLS results without random coefficient model.

The main hypothesis corroborating with the literature (Theil and Finke 1983; Clements et al. 1996; Chai and Moneta 2012) tests the impact of MPCE on diversity and can be hypothesized as.

H1:

$$H_0 : \beta_{10} = 0$$
$$H_1 : \beta_{10} > 0$$

Similarly, the impact of children can be hypothesized as

H2:

$$H_0 : \beta_{20} = 0$$
$$H_1 : \beta_{20} > 0$$

Then, we also test significance of random term by considering three hypotheses if variances of $\upsilon_{0i}, \upsilon_{1i}$ and υ_{2i} are zero or not. From the descriptive statistics (Table 2a–d), we do not observe any consistent pattern of impact of elasticity on diversity measures, though except miscellaneous group,[2] a negative association is observed between luxury category and Theil's measure.

[2]This commodity group is constructed by combining six subgroups total to avoid lot of zero consumption data. If it were constructed using all disaggregated items ranging almost 100 items, this high value of Theil's measure might not be observed.

Table 2 Descriptive statistics of commodity group

	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
(a) Descriptive statistics of Uttar Pradesh						
1	Cereals	0.09 (0.05)	0.01	0.18	N	0.27(0.10)
2	Pulses	0.03 (0.02)	0.00	0.33	N	0.43 (0.19)
3	Milk and milk products	0.11 (0.06)	0.06	0.27	L	0.09 (0.14)
4	Sugar and salt	0.02 (0.01)	0.00	0.37	N	0.22 (0.12)
5	Edible oil	0.04 (0.02)	0.01	0.24	N	0.14 (0.18)
6	Meat–fish–eggs	0.07 (0.05)	0.60	0.21	N	0.05 (0.12)
7	Vegetables	0.06 (0.03)	0.00	0.43	N	0.64 (0.11)
8	Fruits	0.03 (0.03)	0.22	0.09	L	0.15 (0.17)
9	Beverages	0.02 (0.02)	0.00	0.16	L	0.16 (0.18)
10	Packaged food	0.05 (0.04)	0.04	0.30	L	0.31 (0.17)
11	Spices	0.03 (0.02)	0.02	0.58	N	0.70 (0.16)
12	Fuel and light	0.06 (0.04)	0.00	0.40	L	0.31 (0.13)
13	Clothing–bedding–footwear	0.07 (0.03)	0.00	0.34	N	0.63 (0.10)

(continued)

Table 2 (continued)

	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
14	Miscellaneous	0.13 (0.07)	0.00	0.78	L	0.74 (0.11)

	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
(b) Descriptive statistics of West Bengal						
1	Cereals	0.11(0.06)	0.00	0.21	N	0.26 (0.11)
2	Pulses	0.02 (0.01)	0.02	0.24	N	0.31(0.19)
3	Milk and milk products	0.04 (0.03)	0.18	0.32	L	0.13(0.18)
4	Sugar and salt	0.01 (0.01)	0.01	0.39	N	0.30(0.13)
5	Edible oil	0.04 (0.02)	0.02	0.24	N	0.13(0.16)
6	Meat–fish–eggs	0.10 (0.05)	0.08	0.33	N	0.28(0.19)
7	Vegetables	0.07 (0.04)	0.02	0.50	N	0.67(0.14)
8	Fruits	0.03 (0.03)	0.27	0.11	L	0.18(0.19)
9	Beverages	0.02 (0.02)	0.03	0.17	L	0.18(0.19)
10	Packaged food	0.06 (0.08)	0.01	0.34	L	0.35(0.16)
11	Spices	0.03 (0.02)	0.02	0.57	N	0.70(0.15)
12	Fuel and light	0.07 (0.04)	0.00	0.42	L	0.31 (0.12)

(continued)

Table 2 (continued)

	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
13	Clothing–bedding–footwear	0.07 (0.03)	0.00	0.32	N	0.60 (0.09)
14	Miscellaneous	0.14 (0.08)	0.00	0.81	L	0.75(0.12)
	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
(c) Descriptive statistics of Maharashtra						
1	Cereals	0.08(0.04)	0.02	0.27	N	0.39(0.13)
2	Pulses	0.03 (0.02)	0.01	0.37	N	0.49 (0.16)
3	Milk and milk products	0.07 (0.03)	0.03	0.29	N	0.11 (0.14)
4	Sugar and salt	0.02 (0.01)	0.01	0.39	N	0.23(0.12)
5	Edible oil	0.04 (0.04)	0.01	0.18	N	0.02 (0.07)
6	Meat–fish–eggs	0.06 (0.05)	0.43	0.27	N	0.13 (0.18)
7	Vegetables	0.05 (0.03)	0.02	0.53	N	0.73(0.13)
8	Fruits	0.05 (0.03)	0.04	0.13	L	0.33(0.16)
9	Beverages	0.03 (0.02)	0.01	0.18	N	0.21 (0.18)
10	Packaged food	0.05 (0.06)	0.09	0.27	L	0.25 (0.17)
11	Spices	0.03 (0.03)	0.01	0.62	N	0.73 (0.17)
12	Fuel and light	0.07 (0.03)	0.00	0.44	N	0.34 (0.10)

(continued)

Table 2 (continued)

	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
13	Clothing–bedding–footwear	0.06 (0.03)	0.00	0.32	N	0.61 (0.09)
14	Miscellaneous	0.18 (0.08)	0.00	0.81	L	0.77 (0.11)
	Commodity group	Mean budget share (sd)	Proportion of zero consumption	Proportion of items consumed out of total items in the group	Category Necessary (N), Luxury (L)	Average(sd) diversity (Theil's measure)
(d) Descriptive statistics of Tamil Nadu						
1	Cereals	0.06(0.04)	0.05	0.15	N	0.15 (0.13)
2	Pulses	0.03 (0.01)	0.01	0.41	N	0.54 (0.14)
3	Milk and milk products	0.07 (0.03)	0.05	0.24	N	0.06 (0.11)
4	Sugar and salt	0.01 (0.01)	0.00	0.44	N	0.43 (0.16)
5	Edible oil	0.02(0.01)	0.01	0.22	N	0.09 (0.15)
6	Meat–fish–eggs	0.07 (0.05)	0.24	0.28	N	0.17(0.19)
7	Vegetables	0.05 (0.03)	0.01	0.53	N	0.73 (0.10)
8	Fruits	0.04 (0.02)	0.02	0.10	L	0.23 (0.16)
9	Beverages	0.03 (0.03)	0.01	0.19	N	0.21 (0.18)
10	Packaged food	0.05 (0.05)	0.08	0.26	L	0.24 (0.18)
11	Spices	0.04 (0.03)	0.01	0.85	N	0.89 (0.10)
12	Fuel and light	0.05 (0.03)	0.00	0.43	N	0.33 (0.10)
13	Clothing–bedding–footwear	0.05 (0.02)	0.00	0.27	N	0.57 (0.10)
14	Miscellaneous	0.17 (0.07)	0.00	0.82	L	0.78 (0.10)

4 Results

In Table 1, we present mean value of total budget share of 14 commodity groups together, number of children and mean MPCE across four states. The figures imply that UP belongs to the lowest level of income among the four states considered with higher number of children in the family, and Maharashtra is most wealthy among these four states. These findings are also substantiated from Table 2a–d, where the number of luxury items is much less for Maharashtra, the rich state, as opposed to other states like UP.

Table 2a–d describes mean budget share of each of the 14 commodity groups, along with percentage of zero expenditure, income elasticity group and Theil's diversity measure. The variation in consumption expenditure is evident for certain groups of commodities such as milk and milk products, meat–eggs–fish, fruits and also to a certain extent in miscellaneous goods. The varying dietary habit across regions is also evident in the average value of proportion of zero consumption, particularly for the commodity group meat–egg–fish. This is the reason, in addition to considering state fixed effects, we also calculate Theil's diversity measure at state level and income-elasticity dummy is also computed by taking state-level average value of budget shares. The proportion of items consumed is also a descriptive indication of diversity. The higher proportion of items consumed in the food group "spices" and "vegetables" explicitly indicate exclusive dietary pattern of Indian households (predominantly vegetarian and spice based). In the miscellaneous group, we considered subgroup items (described in appendix) within which there are lot of individual items. Hence, high proportion (0.78–0.82) does not necessarily mean all detailed items are consumed by households, instead these values reveal that consumers do not spend even on some of these broad subgroups consisting of entertainment, conveyance, personal article, consumer services, etc.

In Table 3, we present regression results of random coefficient model along with benchmark OLS model. It is quite evident that for hypotheses 1 and 2, null hypotheses are accepted with more strong impact of income on diversity is observed. This is quite consistent with the findings in the literature (see, e.g., Theil and Finke 1983; Clements et al. 1996; Chai and Moneta 2012). We also demonstrated the significance (Chai and Moneta 2012) of demographic attributes, locality and income elasticity of commodities. Also the influence of MPCE is more for luxury commodities as opposed to necessary commodity, though on its own a negative association between diversity and luxury nature of commodities is observed. The rationale behind this negative coefficient is the fact that these commodities are highly priced, and given the level of limited resources of Indian households, households can be expected to afford only a subset of these groups of commodities rather a variety of goods. The impact of MPCE becomes less statistically significant when we have random slope parameter as captured by Model 6. Also, it is seen that proportion of error variance due to random intercept is much high as opposed to variation among households within commodity group, particularly after allowing for variations in MPCE slope parameters. We also have done likelihood test to test significance of between-commodity

Table 3 Main findings on consumption diversity

Model	Model 1 (no dummy)	Model 2 (intercept luxury dummy)	Model 3 (intercept and also interactive dummy with MPCE)	Model 4 (intercept and interactive dummy with both slope variables)	Now random coefficient model	Model 5 (random intercept model)	Model 6 (random intercept and random MPCE coefficient model)	Model 7 (random intercept and random children coefficient model)	Model 8 (full random coefficient model as described in Eq. 4)
Intercept	−0.01(2.18)**	−0.02 (1.72)*	0.11 (10.71)***	0.11 (10.39)***		0.06 (0.97)	0.05 (0.37)	0.06 (0.90)	0.02 (0.20)
MPCE	0.05 (45.04)***	0.05 (45.06)***	0.03 (25.79)***	0.03 (25.06)***		0.04 (53,37)***	0.04(4.28)***	0.03 (53.52)***	0.04 (3.94)***
Children	0.02 (25.79)***	0.02 (25.80)***	0.02 (26.20)***	0.01 (21.54)****		0.01 (35.92)***	0.01 (35.84)***	0.02 (6.43)***	0.02 (5.36)***
Luxury		−0.02 (13.10)***	−0.41 (25.66)****	−0.41 (22.64)****		−0.24 (24.09)***	−0.19(10.49)***	−0.24(23.10)***	−0.13 (6.80)***
Luxury*MPCE			0.05 (24.57)***	0.05 (22.42)***		0.03 (26.49)***	0.03 (11.62)***	0.03 (26.48)***	0.03 (8.46)***
Luxury*children				0.0002 (0.20)		0.004 (5.63)***	0.01(6.47)***	−0.002 (1.433)	−0.004 (2.83)**
TN	0.02 (10.41)***	0.02 (10.42)***	0.02 (9.92)***	0.02 (9.91)***		0.02 (18.23)***	0.02 (18.53)***	0.02 (18.41)***	0.02 (18.91)***
UP	−0.02 (11.66)***	−0.02 (9.32)***	−0.02 (9.27)***	−0.02(9.27)***		−0.03 (23.15)***	−0.03 (24.62)***	−0.03 (23.29)***	−0.03 (24.35)***
WB	−0.001 (−0.39)	0.003 (1.76)*	0.002 (1.066)	0.002 (1.07)		−0.01 (4.46)***	−0.01 (5.53)***	−0.01(5.05)***	−0.01 (5.71)***
Error variance of intercept only ((v_{0i})						0.05 (0.02)[a]	0.20 (0.07)[a]	0.05 (0.02)[a]	0.24 (0.09)[a]

(continued)

Table 3 (continued)

Model	Model 1 (no dummy)	Model 2 (intercept luxury dummy)	Model 3 (intercept and also interactive dummy with MPCE)	Model 4 (intercept and interactive dummy with both slope variables)	Now random coefficient model	Model 5 (random intercept model)	Model 6 (random intercept and random MPCE coefficient model)	Model 7 (random intercept and random children coefficient model)	Model 8 (full random coefficient model as described in Eq. 4)
Error variance of MPCE coefficient (v_{1i})							0.001 (0.001)[a]		0.002 (0.001)[a]
Error variance Children coefficient (v_{2i})								0.0001 (0.00003)[a]	0.0001 (0.0001)[a]
Error variance full model (v_{0i}, v_{1i} and v_{2i})									
Over all variance						0.02	0.02	0.02	0.02
TOTAL OBSERVATIONS = N*#H) = 14*12708	177,912	177,912	177,912	177,912		177,912	177,912	177,912	177,912
Model fitness	0.01483 (Adj R2)	0.01577 (Adj R2)	0.01909 (Adj R2)	0.01909 (Adj R2)		83286.1 (LL)	85,399.4 (LL)	83,675.6 (LL)	85,966.4 (LL)
					LR test = χ^2	Vs model 4= 2.1 + e05***	Vs model 5 = 4226.5***	Vs model 5 = 779.06***	Vs model 6 = 1134.06***

Note T-statistics are reported in parentheses
*, **, *** indicate significance level 0.10, 0.05 and 0.01, respectively
[a]Standard error of variance estimator

effects by comparing this random intercept model with single-level model. The LR test statistic comes out to be very high with p value close to 0.00.[3] The comparison among different random coefficient models shows the superiority of these models as compared to pooled OLS model. Yet, we can affirm that overall conclusions regarding fixed effect variables remain same across our benchmark model and random coefficient models.

5 Conclusion

There exists considerable amount of empirical evidences to support for larger diversification with the increase in income. There are studies in cross-country dimensions as well as at household level, but all these studies are for the developed nation. These studies found positive association between income and dispersion of commodity expenditure mainly captured through number of consumed items. In this paper, we attempt to capture wide variety of household consumption pattern for the developing country India, using Theil's entropy measure for different regions of India with varying degree of level of livings, expressed by mean MPCE level. Also, the case of India is more interesting due to varying food habit which is well demonstrated in this study. The states considered for this study are Maharashtra (MH), Tamil Nadu (TN), Uttar Pradesh (UP) and West Bengal (WB) for the year 2011–2012, the last available year of bigger consumption survey. The choice of the states mainly stems from the fact that these four states belong to four different regions, namely west, south, north and east, respectively, of India, with varying pattern of consumption style, particularly within the food group. Moreover, we also look into other factors which might have significant impact on diversity of consumption even after controlling for main determining variable, i.e., MPCE such as *number of children*.

The contrasting impact of luxury commodity on diversity on its own and when interacting with MPCE should be judged keeping in mind the nature of correlation between luxury commodity and MPCE. As indicated before, higher value of diversity, observed to be significantly associated with higher level of income in empirical studies, hence indicating higher level of household welfare, is also well demonstrated with significant impact of MPCE in our study. Hence, due to the presence of high-valued items in luxury commodity group, people cannot afford to have them with lower level of MPCE (as observed from our descriptive statistics that UP has more number of luxury commodities) and therefore generating negative impact on diversity. Yet the moderating impact of this dummy with respect to MPCE is more pronounced as compared to necessary goods due to innate preferences of human beings for a high quality of goods, which could be made availability with increasing MPCE. This is also supported in findings of Jackson (1984), Drescher et al. (2008).

---

[3] P value is adjusted for such non-standard hypothesis tests as the null hypothesis of zero variance is on the boundary space.

Though there could be other unobserved characteristics at the household level such as taste changes and network effect, we attempt to incorporate demographic attributes such as number of children, place of residence to tackle some of these unobserved variables in order to draw unbiased causal inference of income (proxied as MPCE) on consumption diversity. We also notice product differentiation matter significantly as evidenced through LR statistics from the random coefficient model. But the variations across commodity groups are mainly captured through variations in intercept rather through variations in slope coefficients related to variables MPCE and demographic attributes.

Appendix 1: Description of Items

	Commodity group	Number of items	Item description
1	Cereals	19	Rice—other sources chirakhoi, lawa muri other rice products wheat/atta—other sources maida suji, rawasewai, noodles bread (bakery) other wheat products jowar and its products bajra and its products maize and products barley and its products small millets and their products ragi and its products other cereals cereal substitutes: tapioca, etc.
2	Pulses	12	Arhar, tur gram: split gram: whole moong masur urd peas khesari other pulses gram products besan other pulse products
3	Milk and milk products	6	Milk: liquid (litre) baby food milk: condensed/powder curd ghee butter
4	Sugar and salt	6	Salt sugar, non-PDS sugar—other sources gur candy, misri honey
5	Edible oil	6	Vanaspati, margarine mustard oil, groundnut oil, coconut oil, refined oil (sunflower, soya bean, saffola, etc.), edible oil: others
6	Meat, egg, and fish	7	Eggs (no.) fish, prawn, goat meat/mutton, beef/buffalo meat, pork, chicken, others: birds, crab, oyster, tortoise, etc.
7	Vegetable	17	Potato, onion, tomato, brinjal, radish, carrot, palak/other leafy vegetables, green chillies, lady’s finger, parwal/patal, kundru, cauliflower, cabbage, gourd, pumpkin, peas, beans, barbati, other vegetables

(continued)

(continued)

	Commodity group	Number of items	Item description
8	Fruits	27	Banana (no.), jackfruit, watermelon, pineapple (no.), coconut (no.), green coconut (no.), guava, singara, papaya, mango, kharbooza, pears/nashpati berries, leechi, apple, grapes, groundnut dates, cashew nut, walnut, other nuts, raisin, kishmish, monacca, etc., other dry fruits
9	Spices	11	Ginger (gm), garlic (gm), jeera (gm), dhania (gm), turmeric (gm), black pepper (gm), dry chillies (gm), tamarind (gm), curry powder (gm), oilseeds (gm), other spices (gm)
10	Beverages	10	Tea: cups (no.) tea: leaf (gm) coffee: cups (no.) coffee: powder (gm), mineral water (litre), cold beverages: bottled/canned (litre) fruit juice and shake (litre), papad, bhujia, namkeen, mixture, chanachur chips (gm), pickles (gm), sauce, jam, jelly (gm)
11	Processed food	10	Cooked meals, cooked, snacks purchased, prepared sweets, cake, pastry, biscuits, chocolates, etc., papad, bhujia, namkeen, mixture, chanachur, chips, sauce, jam jelly, pickles etc.
12	Fuel and light	8	Electricity (std. unit), LPG [excl. conveyance], petrol (litre) [excl. conveyance], diesel (litre) [excl. conveyance], candle, match box kerosene—other sources (litre)
13	Clothing, bedding, and footwear	37	Dhoti (no.), saree (no.), cloth for shirt, pyjama, kurta, salwar, etc. (metre), cloth for coat, trousers, suit, etc. (metre), coat, jacket, sweater, windcheater (no.), shawl, chaddar (no.), kurta-pyjama suits: males (no.), kurta-pyjama suits: females (no.) kurta, kameez (no.) pyjamas, salwar (no.) shirts, T-shirts (no.), shorts, trousers, bermudas (no.), frocks, skirts, etc. (no.), blouse, dupatta, scarf, muffler (no.), lungi (no.), baniyan, socks, other hosiery and undergarments, etc.(no.), gamchha, towel, handkerchief (no.), headwear, belts, ties (no.), knitting wool (gm), bed sheet, bed cover (no.), rug, blanket (no.), pillow, quilt, mattress (no.), cloth for upholstery, curtains, tablecloth, etc. (metre), mosquito net (no.), leather boots, shoes leather sandals, chappals, etc., other leather footwear rubber/PVC footwear other footwear
14	Miscellaneous	6 subgroups	Entertainment, toilet articles, mini-durables, household consumables, consumer services, conveyance

Appendix 2: Average Income Elasticity

Algebraic form of log-quadratic demand function:

$$\overline{w_i} = \alpha_i + \beta_i \overline{\log y} + \gamma_i$$

Let p_i and $\overline{q_i}$ be price and average of quantity consumed. Now substituting $\overline{w_i} = \frac{p_i \overline{q_i}}{\bar{y}}$ in the above equation, we obtain

$$p_i \overline{q_i} = \alpha_i \bar{y} + \beta_i \bar{y} \overline{\log y} + \gamma_i \bar{y}$$

Differentiating the above equation w.r.t. $\bar{y}$ yields

$$\frac{\mathrm{d}\overline{q_i}}{\mathrm{d}\bar{y}} = \alpha_i + \beta_i \overline{\log y} + \gamma_i$$

Now recognizing average income elasticity as $\eta_i = \frac{\mathrm{d}\overline{q_i}}{\mathrm{d}\bar{y}} \frac{\bar{y}}{\overline{q_i}} = \frac{\mathrm{d}\overline{q_i}}{\mathrm{d}\bar{y}} \frac{p_i}{\overline{w_i}}$ yields

$$\eta_i = 1 + \frac{\beta_i + 2\gamma_i \overline{\log y}}{\overline{w_i}}.$$

References

Allen, R. G. D., & Bowley, A. L. (1935). Family expenditure: A study of its variation. *Studies in statistics and scientific method, 2*. London, PS King.

Anand, S., & Harris, C. (1994). Choosing a welfare indicator. *The American Economic Review, 84*(2), 226–231 (Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association).

Banks, J., Blundell, R., & Lewbel, A. (1997). Quadratic Engel curves and consumer demand. *Review of Economics and Statistics, 79*(4), 527–539.

Bennett, M. K. (1941). *Wheat studies of the Food Research Institute* (Vols. 12 and 18). Stanford, CA: Stanford University.

Blacklow, P., Cooper, R., Ham, R., & Mclaren, K. (2006). *A regular demand system with commodity-specific demographic effects*. University of Tasmania, Department of Economics and Finance, Working Papers, 2006–06.

Canova, L., Rattazzi, A. M. M., & Webley, P. (2005). The hierarchical structure of saving motives. *Journal of Economic Psychology, 26*(1), 21–34.

Chai, A., & Moneta, A. (2012). Back to Engel? Some evidence for the hierarchy of needs. *Journal of Evolutionary Economics, 22*(4), 649–676.

Chai, A., Kiedaisch, C., & Rohde, N. (2017). The saturation of spending diversity and the truth about Mr Brown and Mrs Jones. Griffith Business School Discussion paper series. 2017–01.

Chai, A., Rohde, N., & Silber, J. (2014). Measuring the diversity of household spending patterns. *Journal of Economic Surveys, 29*(3), 423–440.

Chattopadhyay, N., Majumder, A., & Coondoo, D. (2009). Demand threshold, zero expenditure and hierarchical model of consumer demand. *Metroeconomica, 60*(1), 91–118.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1975). Transcendental logarithmic utility functions. *The American Economic Review, 65*(3), 367–383.

Clements, K. W., & Chen, D. (1996). Fundamental similarities in consumer behaviour. *Applied Economics, 28*(6), 747–757.

Clements, K., Yanrui, W., & Zhang, J. L. (2006). Comparing international consumption patterns. *Empirical Economics, 31*(1), 1–30.

Conceição, P., & Ferreira, P. (2000). *The young person's guide to the Theil index: Suggesting intuitive interpretations and exploring analytical applications*. UTIP working paper.

Cranfield, J. A. L., Eales, J. S., Hertel, T. W., & Preckel, P. V. (2003). Model selection when estimating and predicting consumer demands using international cross section data. *Empirical Economics, 28*(2), 353–364.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American Economic Review, 70*(3), 312–326. JSTOR.

Drescher, L., Thiele, S., & Weiss, C. R. (2008). The taste for variety: A hedonic analysis. *Economics Letters, 101*(1), 66–68.

Engel, E. (1857). Die Productions-Und Consumtionsverhältnisse Des Königreichs Sachsen. *Zeitschrift Des Statistischen Bureaus Des KöniglichSächsischenMinisteriums Des Innern, 8*: 1–54.

Engel, E. (1895). Die Lebenskosten Belgischer Arbeiter Familien frfther und jetzt, Bulletin de l'institut international de statistique, tome IX, premiere livraison, Rome.

Falkinger, J., & Zweimüller, J. (1996). The cross-country engel curve for product diversification. *Structural Change and Economic Dynamics, 7*(1), 79–97.

Flurry, L. A. (2007). Childrens influence in family decision-making: Examining the impact of the changing American family. *Journal of Business Research, 60*(4), 322–330.

Houthakker, H. S. (1957). An international comparison of household expenditure patterns, commemorating the centenary of Engel's law. *Econometrica, 25*(4), 532–551.

Howe, H., Pollak, R. A., & Wales, T. J. (1979). Theory and time series estimation of the quadratic expenditure system. *Econometrica: Journal of the Econometric Society*, 1231–1247.

Jackson, L. F. (1984). Hierarchic demand and the Engel curve for variety. *The Review of Economics and Statistics, 66*(1), 8–15.

Jorgenson, D., Lau, L. J., Stoker, T. M., Basmann, R. L., & Rhodes, G. (1982). The transcendental logarithmic model of aggregate consumer behavior. *Advances in Econometrics*. JAI Press.

Kahn, B. E., & Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research, 30*(4), 519–533.

Lancaster, G., & Ray, R. (1998). Comparison of alternative models of household equivalence scales: The Australian evidence on unit record data. *Economic Record, 74*(224), 1–14.

Lewbel, A. (1991). The rank of demand systems: Theory and nonparametric estimation. *Econometrica, 59*(3), 711.

Lewbel, A., & Pendakur. K. (2008). Estimation of collective household models with Engel curves. *Journal of Econometrics, 147*(2), 350–358.

Maddala, G., et al. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica, 39*(2), 341–358.

Majumder, A. (1992). Measuring income responses: A log-quadratic demand model for consumers in India. *Empirical Economics, 17*(2), 315–321.

Morseth, M. S., Grewal, N., Kaase, I., Hatloy, A., Barikmo, I., & Henjum, S. (2017). Dietary diversity is related to socioeconomic status among adult Saharawi refugees living in Algeria. *BMC Public Health, 17*.

Muellbauer, J. (1976). Community preferences and the representative consumer. *Econometrica: Journal of the Econometric Society*. JSTOR, 979–999.

Pasinetti, L. L. (1981). *Structural change and economic growth*. Cambridge: Cambridge University Press.

Prais, S. J. (1952). Non-linear estimates of the Engel Curves. *The Review of Economic Studies, 20*(2), 87–104.

Prais, S. J., & Houthakker, H. S. (1955). *The analysis of family budgets* (Vol. 4). CUP Archive.

Ray, R. (1986). Demographic variables and equivalence scales in a flexible demand system: The case of AIDS. *Applied Economics, 18*(3), 265–278.

Shaw, D., & Newholm, T. (2002). Voluntary simplicity and the ethics of consumption. *Psychology and Marketing, 19*(2), 167–185.

Shorrocks, A. F. (1988). Aggregation issues in inequality measurement. In *Measurement in economics* (pp. 429–451). Berlin: Springer.

Stone, R. (1954). Linear expenditure systems and demand analysis: An application to the pattern of British demand. *The Economic Journal, 64*(255), 511–527.

Sudarshan, R., & Mishra, N. (1999). Gender and tobacco consumption in India. *Asian Journal of Women's Studies, 5*(1), 84–114.

Subramanian, S. V., Nandy, S., Kelly, M., Gordon, D., & Smith, G. D. (2004). Patterns and distribution of tobacco consumption in India: Cross sectional multilevel evidence from the 1998–9 national family health survey. *Bmj, 328*(7443), 801–806.

Theil, H. (1967). *Economics and information theory*. Technical report.

Theil, H., & Finke, R. (1983). The consumer's demand for diversity. *European Economic Review, 23*(3), 395–400.

Thiele, S., & Weiss, C. (2003). Consumer demand for food diversity: Evidence for Germany. *Food Policy, 28*(2), 99–115.

Thomas, D., & Strauss, J. (1997). Health and wages: Evidence on men and women in urban Brazil. *Journal of Econometrics, 77*(1), 159–185.

Witt, U. (2017). The evolution of consumption and its welfare effects. *Journal of Evolutionary Economics, 27*(2), 273–293.

Working, H. (1943). Statistical laws of family expenditure. *Journal of the American Statistical Association, 38*(221), 43–56.

Trend in Inter-state Disparity of Well-Being in India in the Post-Reforms Period

Asis Kumar Banerjee

Abstract This paper compares the disparities of well-being among sixteen of the most populous states of India in the three years—1991, 2002 and 2012. It finds that while disparity increased between 1991 and 2012 and also between 2002 and 2012, what happened to it between 1991 and 2002 is ambiguous. The major differential contribution of the paper is *methodological*. By taking a Lorenz dominance approach, we show that, in contrast with the findings in the existing literature, those of this paper—both the definitive and the ambiguous ones—are *robust*: they do not depend on the use of any specific inequality index as the measure of disparity. Furthermore, the task of reducing well-being matrices to vectors (which is a part of the exercise) has been handled in a way that seems to be logically more satisfactory than the method followed in the literature. Among other distinguishing features of the paper is the fact that *economic* well-being is measured by asset ownership which, we have argued, is a better indicator of this aspect of well-being than income (which is the measure that is customarily used in this context). Moreover, unlike the existing literature, our time coverage includes the decade of the 2000s.

Keywords Inter-state disparity · Well-being · Multidimensional Lorenz dominance

1 Introduction

Inter-state disparity in well-being in India is a topic on which there is, by now, a voluminous literature. Much of the motivation behind this type of work has come from the well-known literature (that came into being in the 1980s and the 1990s) on the convergence of the growth rates and the levels of income among *countries* over the years (see, for instance, Barro and Sala-i-Martin (1995) and the references cited therein). Because of the obvious socio-economic importance of the need to study regional disparity in India, economists have naturally been interested in answering the corresponding question for the regions (taken to be represented by the states) *within* the country.

A. K. Banerjee (✉)
Calcutta University, Kolkata, India
e-mail: asisbanerjee.cu@gmail.com

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_25

A number of studies on this topic have concentrated on the disparity of *income*. In this paper, however, we shall be concerned with the broader notion of disparity of *well-being*. Dholakia (2003) and Ghosh (2006) found convergence of well-being among the Indian states over the period 1981–2001. Roy and Bhattacharjee (2009) found absolute β-convergence of the Human Development Indices (HDIs) of the states but insignificant σ-convergence.[1] Noorbaksh (2003) concentrated on economic well-being but considered its different aspects (especially income and expenditure) and also considered the rural and the urban sectors separately; he found evidence of divergence. Banerjee and Kuri (2015), however, report convergence of well-being.

Intuitively, disparity in this context is nothing but the *inequality* in the distribution of well-being among the states. The existing literature corroborates this fact by using, for the purpose of estimating disparity, such measures as the coefficient of variation, the standard deviation of the log values of well-being and the Gini coefficient, which are actually *inequality indices*. For this reason in this paper, we shall use the words *disparity* and *inequality* interchangeably.

The present paper is motivated by the fact that in the existing literature, there seem to be conceptual gaps in the *methods* by which disparity of well-being is estimated. We intend to have a re-look at the convergence question in the Indian context with the help of a new method that takes care of these gaps.

There are *two* major problems with the way in which regional disparity of well-being in India has so far been measured. *First*, assuming that at any given point in time the levels of well-being of the states are described by a vector, which inequality index do we use to estimate disparity? There is no uniquely defined index of inequality. While the coefficient of variation, the Gini coefficient, etc., are widely used in practice, they do not nearly exhaust the list of possible indices. In fact, we have an embarrassment of riches here: the number of inequality indices is *infinite* (see, for instance, Cowell 2000; Sen 1997). The problem is that a given pair of well-being vectors may not be ranked (in terms of disparity) in the same way by different inequality indices. In other words, the answer to the question whether disparity decreases between any two given years may depend crucially on which specific inequality index is used. Note that experimenting with a number of *alternative* inequality indices would not be a way out of this impasse since, as stated above, there are an infinite number of such indices and there is nothing in the theory that can help us in choosing any particular finite set of inequality indices for this purpose. It is only on the basis of the researcher's *subjective* preferences that such choices can be made. This problem raises serious doubts as to the *robustness* of the conclusions regarding the time trend of inter-state disparity in India that have so far been reported in the literature.

---

[1]The terminology is borrowed from the literature on convergence between *countries*. There is said to be σ-convergence if the (cross-sectional) dispersion of income levels among countries decreases over time. On the other hand, there is (absolute) β-convergence if countries with *lower* income levels have *higher* growth rates of income. It is known that σ-convergence implies β-convergence. If there are just two countries, the converse is also true. In the general case, however, there can be β-convergence without there being σ-convergence (see, for instance, Martin and Sunley 1998 for a discussion of these issues).

It is well known that there is a circumstance under which this problem would not arise. If it so happens that the well-being vector in Year 2 Lorenz dominates that in Year 1, then we can safely conclude that inequality in Year 2 is less than that in Year 1, no matter which specific index is used for measuring inequality. However, if neither vector Lorenz dominates the other, then it cannot be the case that all inequality indices would give the same inequality rankings. In other words, in that case, there must exist at least two inequality indices that would give contradictory answers to the question whether disparity has decreased (see Foster 1985). In this case, the direction of change in disparity must be considered to be ambiguous. Any definitive conclusion would be unacceptable.

The *second* of the two methodological problems arises from the fact that well-being is actually a multidimensional notion, depending, as it does, not only on income or assets but also on a host of other determinants such as health and education. Hence, a description of the well-being levels of the states in any year should, strictly speaking, be described by a *matrix* (with a given row, for instance, specifying the values of the different determinants for a given state) rather than by a *vector*. Even if we choose a specific inequality index arbitrarily, we would have to compare the degree of inequality of two *matrices*. We would need a *multidimensional inequality index*. However, as in the case of a single dimension, in the multidimensional case also, there is infinity of alternative inequality indices. If we want the conclusions of the exercise of disparity comparisons over time to be robust, we again have to avoid choosing a specific index arbitrarily. We now need a notion of *multidimensional Lorenz dominance*.

The (second) problem can be bypassed if, as is done in HDI comparisons, we reduce the well-being matrix of the states for any given year to a well-being *vector* by aggregating across the columns. A frequently adopted procedure is to take the *arithmetic mean* of the column vectors in the well-being matrix (after normalising the columns to make them unit-free). Once this is done, the multidimensionality problem is avoided: we proceed as if we had vectors rather than matrices to start with. Multidimensional Lorenz dominance would then simply mean Lorenz dominance in the usual (unidimensional) sense applied on the constructed vectors.

However, as we shall see in Sect. 2 below, a multidimensional Lorenz dominance relation (MLDR) obtained in this way would violate one of the basic conditions that one intuitively expects such a relation to satisfy.

The arithmetic mean, of course, is only one of many possible averaging formulas. However, for about 20 years since 1990 (when the UNDP started publishing its Human Development Reports), this was the procedure that was followed in these reports. Since the existing empirical literature on well-being disparity in India is mostly concerned with the time period up to about 2001 and since, for the post-1990 years, the well-being levels of the states are usually taken to be the state HDIs given by the Human Development Reports for India (which follows the UNDP procedure for the country-level HDIs), we have focused our remarks above on the use of the arithmetic mean. Researchers who have included the period *before* 1990 in the time spans of their analyses have also largely followed this particular procedure for calculating the well-being vectors for themselves.[2]

In this paper, too, we shall reduce well-being matrices to vectors. However, for that purpose we shall apply a new procedure that is very similar to the one developed in Banerjee (2016). This will enable us to obtain an MLDR that satisfies all of the properties of such a relation that have been proposed in the theory of multidimensional inequality measurement.

As in the unidimensional case, here too it is possible that, of the two years' matrices, neither multidimensionally Lorenz dominates the other. In such cases, any firm conclusion regarding the direction of change in the degree of regional disparity between these years would again be suspect. We can, however, reach definitive conclusions in cases where there is multidimensional Lorenz dominance.

In order to focus on our task, however, we shall limit the scope of the paper in *other* respects. Two of the major limitations need to be stated explicitly at the outset. (1) Unlike many of the contributions mentioned above, we shall be concerned exclusively with the factual question of convergence or divergence. We do not explore the interrelationships between the trend of regional disparity on the one hand and the trends in various other economic variables (such as the growth rate of the economy, agricultural production and productivity) and policy parameters on the other. Needless to say, these interrelationships are important and they merit detailed study in their own right. However, they lie outside the scope of this paper. (2) We shall be concerned only with convergence of the *level* (i.e. with σ-convergence) of well-being among the states. Thus, the focus is entirely on whether the (cross-sectional) inequality in the distribution of well-being among the states has decreased over time.

Regarding the determinants of well-being, we follow many other researchers in concentrating on health, education and economic well-being. We use life expectancy at birth and the literacy rate as the indicators of health and educational status, respectively. However, so far as economic well-being of a state is concerned, we make a major deviation from the usual practice of using income, usually proxied by Per Capita Net State Domestic Product (PCNSDP), as the indicator. From the point of view of economic theory, *asset ownership* is a better indicator of the economic status of an individual than income. It captures both the ability to finance current consumption and the capacity to bear risks of all types (including those related to health and employment). For a country like India where social security coverage is woefully inadequate, risks and vulnerabilities faced by the individuals are extremely important

[2]From 2010, the UNDP started using the geometric (rather than the arithmetic) mean as the aggregator function. It can be checked, however, that the problem noted in the text in connection with the need to ensure that the MLDR satisfies the basic requirements of such a relation would persist even if the arithmetic mean is replaced by the geometric mean as the averaging formula. In fact, it would also persist for some of the other averaging formulas, for instance, the one based on the idea of the (Euclidean) distance of the actual values of achievements of a state from the "ideal" situation in which the state would enjoy maximal possible values of the different achievements (see Noorbakhsh 1998; Banerjee and Kuri 2015). The theoretical literature also contains yet other suggested aggregation procedures (see, for instance, the class of formulas derived axiomatically in Mishra and Nathan 2014 and the other procedures reviewed there). Most of these, however, have not yet been applied to the study of regional disparity in India.

considerations in judging their well-being. These are not captured by income data. Accordingly, we use the per capita value of assets owned by the average household in a state as the indicator of the state's economic well-being.

We concentrate on the post-Reforms period. We compare the (cross-sectional) inter-state disparities at the three points of time—1991, 2002 and 2012. Choice of these dates was dictated largely by the dates for which assets data were available. Data availability constraints also prevent us from bringing all the states of India within the fold of our analysis. We consider 16 of the most populous states. These states account for more than 90% of the country's population.

Our main finding is that while the direction of movement of regional disparity between 1991 and 2002 is *ambiguous*, there has been an unambiguous *increase* in disparity (i.e. there has been "divergence") both between 1991 and 2012 and between 2002 and 2012.

The finding for the first post-Reforms decade contradicts the findings of those earlier contributions that report either unambiguous convergence or unambiguous divergence.[3] So far as the second decade after the Reforms is concerned, our finding is not comparable with the results in the existing literature since the latter are concerned in most cases with time periods ending in 2001 (or earlier).[4] In this respect, this part of our finding is a net addition to our knowledge regarding the trends of regional disparity in India in very recent times.

More importantly, our finding (including both the definitive and the ambiguous parts of it) is *robust*: it is not dependent on the use of any specific inequality index for measuring disparity.

Section 2 below describes our methodology. The data and the empirical analysis are contained in Sect. 3. Section 4 concludes the discussion.

2 Methodology

Since our method of measuring inter-state disparity of well-being is different from the methods that have so far been applied in the existing literature, we present it here in somewhat greater detail than would otherwise have been appropriate in an empirical paper.

Let n be the number of states. The level of well-being of each state is determined by the amounts of m attributes. Both the numbers n and m will be kept fixed throughout this paper. $M = \{1, 2, \ldots, m\}$ and $N = \{1, 2, \ldots, n\}$ are, respectively, the set of attributes and the set of states. By a *well-being matrix* A, we shall mean an $n \times m$

[3] Results reached by contributions (such as Roy and Bhattacharjee 2009) that only report the absence of evidence in support of the convergence hypothesis and let the matter rest at that, however, come close in spirit to this part of our finding.

[4] Among the contributions in the existing literature, the study by Banerjee and Kuri (2015) covers the longest time span, 1971–2001. Needless to say, we are talking here about the trends of regional disparity of *well-being*. So far as disparity of *income* alone is concerned, there are many studies that cover the 2000s.

matrix whose (p-th row and j-th column) term, a_p^j, is the amount of attribute j in state p for all j in M and for all p in N. For any matrix A, $\mathbf{a}_p$ will denote its p-th row and $\mathbf{a}^j$ its j-th column. We shall assume A to be a positive matrix. The set of all possible well-being matrices will be denoted by **A**.

Since the different attributes of well-being may be measured in different units, in order to ensure that the different columns of a well-being matrix are comparable it is necessary to normalise the entries. There are many ways of doing so. The rule used by the UNDP in constructing HDIs is well known. In this paper, however, we normalise any matrix A by dividing each entry in the matrix by the arithmetic mean of the column containing it (it is easily seen that all columns of the matrix would then sum to n). In order to avoid overuse of symbols, we shall, however, assume that this has been done to start with, i.e. it will be understood that all well-being matrices are normalised.

We shall be concerned with inequality dominance. In the case where there is a single attribute ($m = 1$), the standard notion of inequality dominance is that of Lorenz dominance. For any two positive well-being vectors **a** and **b**, **a** weakly Lorenz dominates **b** if the Lorenz curve of **a** does not lie below that of **b** at any point; **a** strictly Lorenz dominates **b** if, in addition, the Lorenz curve of **a** is above that of **b** at some point. We denote the unidimensional *weak Lorenz dominance* relation by L: for all well-being vectors **a** and **b**, **a** L **b** if and only if **a** weakly Lorenz dominates **y**. P will denote the *strict Lorenz dominance* relation derived from L: **a** P **b** if and only if **a** strictly Lorenz dominates **b**. In other words, P is the asymmetric component of L. The symmetric component of L will be denoted by I. For all well-being vectors **a** and **b**, **a** I **b** if and only if **a** weakly Lorenz dominates **b** and **b** weakly Lorenz dominates **a**. It is easily seen that **a** I **b** if and only if the Lorenz curve of **a** coincides with that of **b** (which, in turn, can happen in our framework if and only if either $\mathbf{a} = \mathbf{b}$ or **a** is a permutation of **b**).

Lorenz dominance is closely related to the notion of Pigou–Dalton (PD) transfers. If the attribute in question is income, a PD transfer is an income transfer from a richer to a poorer state by a positive amount which is less than their initial income difference. The following three statements are equivalent (Hardy et al. 1952; Marshall and Olkin 1979, Chap. 1): (1) **a** strictly Lorenz dominates **b**; (2) **a** *Pigou–Dalton majorises* **b,** i.e. **a** is obtained from **b** by a *finite* sequence of PD transfers; and (3) $\mathbf{a} = \mathrm{Q}\mathbf{b}$ for some bistochastic matrix Q, but **a** neither equals **b** nor is a permutation of **b** (a bistochastic matrix is a non-negative square matrix in which each row as well as each column sums to 1).

For our purposes in this paper, we need a multidimensional version of the notion of Lorenz dominance. With that end in view, we start with the notion of a weak multidimensional inequality dominance relation (MIDR), D, on **A**. For all A and B in **A**, if A D B, this will be interpreted to mean that relative inequality in the distribution of overall well-being in the matrix A is not more than that in B, whatever may be the specific method of measuring the degree of overall inequality. D_P and D_I will denote the asymmetric and the symmetric components of D, respectively, i.e.

for all A and B in **A**, A D_P B if and only if (A D B but not B D A); and A D_I B if and only if (A D B and B D A). D is required to be a reflexive and transitive relation.[5]

In order to be intuitively acceptable, the relation D will have to satisfy a number of conditions. Two of these are of particular importance. To state the first of these, note that we need a generalisation of the concept of Pigou–Dalton majorisation to the multidimensional context. One of the most widely used among such generalisations is the concept of Uniform Majorisation (UM) (see Kolm 1977). For all $n \times m$ matrices A and B in **A**, B is said to uniformly majorise A if B $\neq$ A, B is not a row permutation of A and B = QA for some bistochastic matrix Q. Since B = QA implies, $b^j = Qa^j$ for all j in M, b^j Pigou–Dalton majorises a^j for each j in M; and since the same matrix Q is used to majorise all the columns of A, the majorisation is said to be *uniform* across the attributes. We can now state the condition of UM on the relation D.

Uniform Majorisation (UM): For all A and B in **A** such that B is a UM of A, B D_P A.

All inequality dominance relations are, by definition, concerned with considerations of equity. Some important aspects of such considerations are captured by generalisations of the PD transfer principle such as **UM**. In multidimensional theory, however, there is another relevant aspect of the matter, viz. the pattern of *interrelation* among the distributions of the different attributes and its effect on multidimensional inequality.

For all A in **A** and for all p, q in N, let $\mathbf{a}_p \wedge \mathbf{a}_q$ denote the vector $\{\min(a_p^1, a_q^1), \min(a_p^2, a_q^2), \ldots, \min(a_p^m, a_q^m)\}$ and $\mathbf{a}_p \vee \mathbf{a}_q$ the vector $\{\max(a_p^1, a_q^1), \max(a_p^2, a_q^2), \ldots, \max(a_p^m, a_q^m)\}$.

For all A and B in **A** such that A is not equal to B or a row permutation of B, A is said to be obtained from B by a Correlation Increasing Transfer (CIT) if there exist p and q in N such that

(i) $\mathbf{a}_p = \mathbf{b}_p \wedge \mathbf{b}_q$; (ii) $\mathbf{a}_q = \mathbf{b}_p \vee \mathbf{b}_q$; and (iii) $\mathbf{a}_r = \mathbf{b}_r$ for all r in N—$\{p, q\}$.

We desire an MIDR, D to satisfy the following condition:

Correlation Increasing Majorisation (CIM): For all A and B in **A** such that B is obtained from A by a finite sequence of CITs, A D_P B.

Informally, **CIM** requires that greater correlation among the different columns of the distribution matrix implies greater inequality, irrespective of how inequality is measured. It was introduced in the economic literature by Tsui (1999) in the context of inequality measurement. In the statistical literature, however, it was introduced even earlier by Boland and Proschan (1988). The concept of CIT on which it is based was studied in Atkinson and Bourguignon (1982) and in Epstein and Tanny (1980).

Despite the somewhat technical way in which the condition is stated, it is actually quite intuitive. Consider, for instance, the following example. Let $n = 2 = m$. Let A $= \begin{pmatrix} 4/3 & 1/2 \\ 2/3 & 3/2 \end{pmatrix}$ and B $= \begin{pmatrix} 4/3 & 3/2 \\ 2/3 & 1/2 \end{pmatrix}$. B is obtained by a switch of the entries in the second column of A. It is easily checked that this is a CIT. If it is now asked

[5]*Reflexivity* of D means that if A = B, then A D_I B; *transitivity* means that if A D B and B D C, then A D C.

whether we should consider A to strictly dominate B (i.e. whether A should be judged to display a lower degree of equality as per any measure of inequality), there seems to be strong grounds for an affirmative answer. In A, state 1 has a higher allocation of attribute 1 than state 2. But this is, to an extent, compensated for by the fact that w.r.t. attribute 2, it is state 2 that has the relative advantage. In B, however, the effect of the lower allocation of attribute 1 to state 2 is compounded by the fact that this state faces the same predicament w.r.t. attribute 2, i.e. there is a compounding of inequalities across the attributes. Obviously, multidimensional inequality should be considered to be lower for A than for B.

Apart from these two equity-related conditions, we also require D to satisfy a condition that is more general in nature. It is the condition of Anonymity (**ANON**) which requires that if two well-being matrices A and B are such that A is a row permutation of B. Then A D_I B. It means that, in measuring disparity of well-being among the states, the states can listed in any order, i.e. that it does not matter which state is called state 1, which is called state 2, etc.

Thus, for our purposes in this paper an MIDR, D, is a binary relation on **A** satisfying the conditions **UM**, **CIM** and **ANON**.[6] Finally, since we are interested in obtaining a generalisation of the unidimensional Lorenz dominance relation, L, it is natural to require that the dominance relation reduces to L if there is just one attribute. Therefore, we arrive at the following definition of a multidimensional Lorenz dominance relation (MLDR). An MLDR, L^M, is an MIDR on **A** such that $L^M = L$ if $m = 1$.

The asymmetric and symmetric components of L^M will be denoted by P^M and I^M, respectively. Thus, if A P^M B, i.e. if A L^M B, but $\neg$(B L^M A) where the symbol $\neg$ signifies the negation of the statement following it, then A strictly Lorenz dominates B; if A I^M B, i.e. if both A L^M B and B L^M A, then A and B are equivalent in the multidimensional Lorenz comparison.

Now, there have been many different suggestions in the literature as to how to construct an MLDR (for a review, see Banerjee 2014). Unfortunately, most of these fail to satisfy either **CIM** or **UM**. We shall mention two of these suggestions here. Consider first the weak dominance relation D such that, given any two $n \times m$ matrices A and B in **A**, A D B if and only if $\mathbf{a}^j\ L\ \mathbf{b}^j$ for all j in M. In other words, A weakly dominates B in the multidimensional sense if and only if every column of A Lorenz dominates the corresponding column of B in the usual (unidimensional) sense. While this seems to be an intuitively sensible procedure, on closer scrutiny it is seen to

[6]In the theory of inequality measurement, one also comes across two other general (i.e. non-equity) conditions. One of these is called scalar multiplication (SM). It requires that if D is a (weak) dominance relation on the set of the well-being matrices in their original (i.e. non-normalised) forms, then for any matrix A and for any positive scalar k, A D_I (kA). The other condition is called Population Replication Invariance (PRI) which says that if B is a k-fold replication of the states in A for any positive integer k (i.e. if A is an $n \times m$ well-being matrix and if B is the $nk \times m$ matrix obtained by writing A k times), then, again, A D_I B. However, as stated in the text, under our procedure all well-being matrices are normalised to start with. It is easily seen that SM is, therefore, trivially satisfied by any D. Again, throughout our analysis n is kept *fixed*. PRI also will, therefore, be trivially satisfied.

violate the condition of **CIM**. Consider, for instance, the 2 × 2 matrices A and B that was mentioned above in connection with our discussion of this condition. It is easily checked that $\mathbf{a}^j \ I \ \mathbf{b}^j$ for $j = 1.2$. Hence, A D_I B. Recall, however, that **CIM** requires: A D_P B. Thus, the suggested D is not an MIDR as per our definition. Hence, it is not an MLDR.

Consider now what is essentially the UNDP procedure for calculating the vector of HDIs for countries. Use the arithmetic mean formula to aggregate across the columns of the well-being matrices, and then apply the usual notion of Lorenz dominance. For any two $n \times m$ well-being matrices A and B, A D B if and only if $\left[(1/m)\sum_{j=1}^{m}\mathbf{a}^j\right]L\left[(1/m)\sum_{j=1}^{m}\mathbf{b}^j\right]$. This particular suggestion would satisfy **CIM**. However, it would violate **UM**. For instance, let $\mathrm{A} = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix}$ and $\mathrm{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Note that $\mathrm{B} = \mathrm{QA}$, where $\mathrm{Q} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ is a bistochastic matrix. Moreover, A neither equals B nor is a row permutation of B. Hence, if D is to satisfy **UM**, we must have: B D_P A. Thus, it is required that $\left[(1/2)\sum_{j=1}^{2}\mathbf{a}^j\right]P\left[(1/2)\sum_{j=1}^{2}\mathbf{b}^j\right]$ where P is the asymmetric component of L. However, for the specified A and B, we have: $\left[(1/2)\sum_{j=1}^{2}\mathbf{a}^j\right] = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \left[(1/2)\sum_{j=1}^{2}\mathbf{b}^j\right]$ so that $\left[(1/2)\sum_{j=1}^{2}\mathbf{a}^j\right]I\left[(1/2)\sum_{j=1}^{2}\mathbf{b}^j\right]$ where I is the symmetric component of L. Thus, **UM** is violated.

The question, therefore, arises as to whether there exists any MLDR satisfying all of the conditions that one intuitively expects it to satisfy. To answer this question, for any A in **A**, let A^+ denote the matrix for which the j-th column is the vector

$$\mathbf{a}^{+j} = \left[(1/m)\sum_{k=1}^{m}(\mathbf{a}^j + \mathbf{a}^k)/2\right] \tag{2.1}$$

for all j in M. We shall call A^+ the *correlation-augmented* (or *augmented*, for short) version of A.

Augmented matrices have a simple and intuitive interpretation. First note that in a multidimensional framework such as ours, the contribution of an attribute towards the overall degree of inequality (however measured) cannot be taken to be given by its "direct" (or "own") contribution. The indirect effects of the attribute through its interactions with the other attributes are to be taken into account. For simplicity, it is assumed that, for all A in **A** and for all i and j in M, the indirect contribution of the distribution of the i-th attribute to overall inequality through its interaction with the distribution of the j-th attribute is indicated by the inequality of the distribution given by $(\mathbf{a}^i + \mathbf{a}^j)/2$, the simple arithmetic mean of $\mathbf{a}^i$ and $\mathbf{a}^j$. Now, the total contribution of an attribute to overall inequality is obtained by another round of averaging. It is the arithmetic mean of the contributions made by it through its interaction with *all* the attributes. Therefore, to obtain the total contribution of attribute j to overall inequality, we have to look at the inequality of the vector $\left[(1/m)\sum_{k=1}^{m}(\mathbf{a}^j+\mathbf{a}^k)/2\right]$ which has been denoted in Eq. (2.1) by $\mathbf{a}^{+j}$. Thus, for any A, the augmented matrix A^+ is simply the matrix such that, for any j, the j-th column of A^+ is the vector that summarises the role played by the j-th attribute in determining the degree of overall inter-state disparity prevailing in the country.

Now, for any A in **A** and for all j in M, let $(\mathbf{a}^{+j})^*$ denote the re-arrangement of $\mathbf{a}^{+j}$ in non-decreasing order (with ties broken arbitrarily) and let $\mathbf{a_0}$ be the vector $(1/m)\sum_{j=1}^{m}(\mathbf{a}^{+j})^*$.

Consider now the following relation L^* on the set **A** of well-being matrices.

Definition For all A and B in **A**, A L^* B if and only if $\mathbf{a}_0$ L $\mathbf{b}_0$ where L is the unidimensional weak Lorenz dominance relation.

P^* and I^* will denote, respectively, the asymmetric and the symmetric components of L^*. If A L^* B, we shall say that A weakly Lorenz dominates B (in the multidimensional sense). Similarly, if A P^* B (i.e. if A L^* B but not B L^* A), then A strictly Lorenz dominates B.

It can be shown that L^* is an MLDR, i.e. that it satisfies **UM**, **CIM** and **ANON** (the proof of this assertion is analogous to that of Proposition 4.1 in Banerjee 2016, p. 77).[7]

It is not known whether the conditions stated above uniquely imply this MLDR (i.e. whether there exist other valid MLDRs). However, the fact remains that the existing literature does not contain any *other* relation on **A** that fulfils the defining conditions of an MLDR. For this reason, in our empirical work below we shall apply L^* as our criterion for deciding whether the well-being matrix for Indian states for a given year Lorenz dominates that for another.

3 Data and the Analysis

Armed with the multidimensional Lorenz dominance criterion stated in Sect. 2, we now confront the task of studying the pattern of changes in inter-state disparity in well-being in India in recent years. We compare the cross-sectional disparities prevailing in the three years—1991, 2002 and 2012.

We consider three aspects of well-being: those related to economic well-being, health and education. Moreover, so far as health and education are concerned, we follow the practice of using, respectively, life expectancy at birth and the rate of literacy as the proxy variables. Life expectancy data are available for the years of our interest. An easily accessible source is RBI (2016).[8] Literacy rates for the states for the year 1991 have been taken from the same source. The rates for 2001 and 2011

[7] It also satisfies the conditions of SM and PRI mentioned in Note 6.

[8] In our data source (RBI 2016), life expectancy is defined as "the average number of years that is likely to be lived by an infant exposed to the same mortality conditions until they die". The source gives information on this variable in the form of overlapping 5-yearly moving averages starting from 1991 to 1995. The last 5-year period for which information is available is 2009–2013. For

given there have been used as proxies for the rates of the adjacent years, 2002 and 2012, respectively.

Our choice of the three years of our interest is related more to our choice of the measure of *economic* well-being of a state. The measure that is used most widely for this purpose in the existing literature is income (as indicated by Per Capita Net State Domestic Product (PCNSDP) of a state). In rare cases, per capita expenditure has also been used. It is, however, our opinion that it is per capita *wealth* that is a more reliable indicator of the average economic well-being of the citizens of a state than either income or expenditure. The most important reason for this is that, on the one hand, wealth of an individual has been seen to be at least as good a predictor of current standard of living as (if not a better one than) income and, on the other, wealth also indicates one's capacity to bear various types of risks that one faces in life, for instance, health risks, risk associated with stability of employment, the unpredictability of the monsoon (which is an important issue for the agricultural sector), etc. In a country like India where social security coverage is lamentably inadequate, income or expenditure data do not adequately reflect these vulnerabilities.

However, we need to look at per capita *personal wealth* in a state rather than at per capita wealth in the economy of the state as a whole. *Macroeconomic* indicators of per capita wealth (such as capital stock per capita for the states) would be of little relevance for us.[9] We utilise, instead, the information on the "average values of assets (AVA) per household" in the states of India given in the reports of the 48th, 59th and 70th rounds of the All-India Debt and Investment Survey (AIDIS) conducted by the National Sample Survey Organisation (NSSO) on behalf of the Reserve Bank of India (RBI) in 1991, 2002 and 2012. For our purposes, however, we shall convert AVA per household in a state into AVA *per capita* by dividing the former by the average size of the household in the state. Thus, our indicator for average economic well-being of a state is Per Capita Average Value of Assets (PCAVA) in the households of the state.

We have to take care of two issues relating to comparability of the values of PCAVA for the three years. First, since PCAVA is in monetary units (rupees), we have to consider its values at constant prices across the years. Since an asset price index is

---

each of the three years of our interest, 1991, 2002 and 2012, we have taken the average value of the 5-year period that is centred on a year which is nearest to it.

[9] We mention in passing that even for those who wish to use income as the measure of economic well-being, it may not be a good idea to use PCNSDP as the proxy variable. PCNSDP is a macroeconomic category. Its value is taken from the National Accounts Statistics (NAS). In the Indian economy, it has been known for quite some time now that there is often a large gap between estimates of the rate of change of *per capita consumption* based on NAS and that based on the household consumer expenditure surveys of the NSSO. Macroeconomic data taken from the NAS do not seem to reflect correctly what is happening at the ground level. While time series data on household *income* are still not available from these surveys, there are reasons to suspect that there would be a similar problem with respect to the rate of change of per capita *income* estimated from the NAS. In the present context, this casts doubt on the advisability of comparing between the levels of economic well-being of a state in different years (or between the levels of well-being of different states in a given year) on the basis of macroeconomic categories such as PCNSDP. Using the NSSO figures for average Monthly Per Capita Expenditure (MCPE) of the states may be a better strategy.

not available for India, the usual practice is to use, for this purpose, the Whole Price Index (WPI) or the Consumer Price Index (CPI) as the price deflator. In common with Anand and Thampi (2016), we use the WPI. The figures for all the three years have been expressed in 1991 prices. Secondly, in the 2012 AIDIS, there was a change in the list of items on which information was solicited from the households. All the three surveys collected information on both physical and financial assets. However, in 1991 and 2002, the list of physical assets included land, buildings, livestock and poultry, agricultural machinery, non-farm business equipment, all transport equipment and durable household assets. Household durables, however, were dropped in the 2012 survey. To ensure compatibility, we have followed Anand and Thampi (2016) in re-calculating the 1991 and 2002 figures by excluding household durables.[10]

Our study covers 16 states of India: Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal. The states of Chhattisgarh, Jharkhand and Uttarakhand were created in the early 2000s by culling out some parts of Madhya Pradesh, Bihar and Uttar Pradesh, respectively. However, in order to ensure comparability across the three years of our interest, even for 2002 and 2012, Jharkhand has been clubbed with Bihar, Chhattisgarh with Madhya Pradesh, and Uttarakhand with Uttar Pradesh. With this qualification, it is seen that in all of the three Census years, 1991, 2001 and 2011, these sixteen states together accounted for more than 90% of the population of India. It seems safe to guess that the same would be true of the years 2002 and 2012.

Table 1 presents the values of the three attributes of well-being for the above-mentioned states in 1991, 2002 and 2012. Thus, it gives the (non-normalised) well-being matrices for these years. To normalise the matrix, each numerical entry is divided by the arithmetic mean of the column containing it. The result of this exercise is shown in Table 2.

We denote the normalised well-being matrices in 1991, 2002 and 2012 by A, B and C, respectively. These can be readily read off Table 2.

The next step is to obtain the augmented versions, A^+, B^+ and C^+, respectively, of these three matrices. While these conversions require extensive and tedious calculations, notionally they are straightforward applications of the procedure described by Eq. (2.1) and the accompanying remarks in Sect. 2 above. We, therefore, omit the details of these calculations and report the results. Matrices A, B and C as well as their augmented counterparts are displayed below. In each of these 16×3 matrices, the rows refer to the states (in the same order as in Tables 1 and 2), and the first, second and third columns refer, respectively, to PCAVA, LIFE and LIT.

[10]There was also a change in 2012 in the way in which land and buildings were valued. While in the two earlier surveys, these were valued at prices reported by the interviewees; normative values were used in 2012. We have not found it possible to adjust the figures of the earlier surveys since the normative values for these years do not seem to be available. We believe, however, that since our focus is on the *inequality* of average asset ownership among the states (rather than on the average values of the assets per se), this omission does not seriously affect the qualitative results of the paper.

Table 1 Attributes of well-being in 16 states of India in 1991, 2002 and 2012

	1991			2002			2012		
	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)
Andhra Pradesh	13.3	61.8	44.08	23.9	64.6	60.47	75.5	67.9	67.66
Assam	10.0	55.7	52.89	14.1	58.8	63.25	113.9	63.3	73.18
Bihar	14.9	59.3	38.47	16.5	64.1	48.64	40.3	67.7	64.73
Gujarat	19.3	61.0	61.29	38.2	65.6	69.14	250.2	68.2	79.31
Haryana	56.1	63.4	55.85	79.1	66.1	67.91	1007.4	68.2	76.64
Himachal Pradesh	25.1	64.3	63.86	71.4	69.5	76.48	398.5	71.0	83.78
Karnataka	18.3	62.5	56.04	30.2	65.8	60.47	105.3	68.5	75.60
Kerala	30.5	72.9	89.81	65.5	73.2	90.86	641.6	74.8	93.91
Madhya Pradesh	17.0	54.7	44.19	28.7	59.3	63.98	154.4	63.8	70.74
Maharashtra	21.3	64.8	64.87	35.1	67.5	76.88	928.2	71.3	82.91
Odisha	9.6	56.5	49.09	13.3	60.4	63.08	32.7	64.9	73.45
Punjab	46.0	67.2	58.51	66.6	68.3	69.65	345.5	71.1	76.68
Rajasthan	25.5	59.1	38.55	32.8	64.1	60.41	88.3	67.5	67.06
Tamil Nadu	17.8	63.3	62.66	31.2	66.7	73.45	291.6	70.2	80.33

(continued)

Table 1 (continued)

	1991			2002			2012		
	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)	PCAVA (Rs.'000)	LIFE (Years)	LIT (%)
Uttar Pradesh	23.3	56.8	41.56	28.2	60.5	57.04	92.9	63.8	70.22
West Bengal	13.0	62.1	57.70	20.4	66.8	68.64	126.6	69.9	77.08

Notes
(1) PCAVA = Per Capita Average Value of Assets other than household durables; LIFE = Life expectancy; LIT = Literacy rate
(2) PCAVA for all the years is in 1991 prices
(3) Life expectation in Himachal Pradesh in 1991 was not available in the data source. The displayed figure was arrived at by backward extrapolation from the figures for subsequent years
(4) For 2012, Jharkhand, Chhattisgarh and Uttarakhand are clubbed with Bihar, Madhya Pradesh and Uttar Pradesh, respectively
Sources For PCAVA: calculations based on 48th, 59th and 70th rounds of AIDIS; for LIFE and LIT: RBI (2016)

Table 2 Normalised values of attributes of well-being in 16 states of India in 1991, 2002 and 2012

	1991			2002			2012		
	PCAVA	LIFE	LIT	PCAVA	LIFE	LIT	PCAVA	LIFE	LIT
Andhra Pradesh	0.60	1.00	0.80	0.64	0.99	0.90	0.26	0.99	0.89
Assam	0.45	0.90	0.96	0.38	0.90	0.95	0.39	0.94	0.97
Bihar	0.66	0.96	0.71	0.44	0.98	0.73	0.14	0.99	0.85
Gujarat	0.85	0.99	1.12	1.03	1.01	1.03	0.85	1.00	1.05
Haryana	2.48	1.03	1.02	2.13	1.02	1.02	3.43	1.00	1.01
Himachal Pradesh	1.11	1.04	1.16	1.92	1.07	1.14	1.36	1.04	1.10
Karnataka	0.81	1.01	1.02	0.81	1.01	0.92	0.36	1.00	1.00
Kerala	1.35	1.18	1.63	1.76	1.13	1.35	2.19	1.10	1.24
Madhya Pradesh	0.75	0.91	0.80	0.77	0.91	0.96	0.53	0.94	0.93
Maharashtra	0.94	1.05	1.18	0.94	1.04	1.14	3.16	1.04	1.09
Odisha	0.43	0.92	0.89	0.36	0.93	0.94	0.11	0.95	0.97
Punjab	2.04	1.09	1.06	1.79	1.05	1.04	1.18	1.04	1.01
Rajasthan	1.13	0.96	0.70	0.88	0.98	0.90	0.30	0.99	0.88
Tamil Nadu	0.79	1.03	1.14	0.84	1.02	1.10	0.99	1.03	1.06
Uttar Pradesh	1.03	0.92	0.76	0.76	0.93	0.85	0.32	0.93	0.93
West Bengal	0.58	1.01	1.05	0.55	1.03	1.03	0.43	1.02	1.02

Notes
(1) Each numerical entry in the table is the corresponding entry on Table 1 divided by the arithmetic mean of the column containing it. Hence, the sum of the numbers in each of last nine columns is 16
(2) For 2012, Jharkhand, Chhattisgarh and Uttarakhand are clubbed with Bihar, Madhya Pradesh and Uttar Pradesh, respectively
Source Calculations based on Table 1

Matrix A

$$\begin{pmatrix} 0.60 & 1.00 & 0.80 \\ 0.45 & 0.90 & 0.96 \\ 0.66 & 0.96 & 0.71 \\ 0.85 & 0.99 & 1.12 \\ 2.48 & 1.03 & 1.02 \\ 1.11 & 1.04 & 1.16 \\ 0.81 & 1.01 & 1.02 \\ 1.35 & 1.18 & 1.63 \\ 0.75 & 0.91 & 0.80 \\ 0.94 & 1.05 & 1.18 \\ 0.43 & 0.92 & 0.89 \\ 2.04 & 1.09 & 1.06 \\ 1.13 & 0.96 & 0.70 \\ 0.79 & 1.03 & 1.14 \\ 1.03 & 0.92 & 0.76 \\ 0.58 & 1.01 & 1.05 \end{pmatrix}$$

Matrix B

$$\begin{pmatrix} 0.64 & 0.99 & 0.90 \\ 0.38 & 0.90 & 0.95 \\ 0.44 & 0.98 & 0.73 \\ 1.03 & 1.01 & 1.03 \\ 2.13 & `1.02 & 1.02 \\ 1.92 & 1.07 & 1.14 \\ 0.81 & 1.01 & 0.92 \\ 1.76 & 1.13 & 1.35 \\ 0.77 & 0.91 & 0.96 \\ 0.94 & 1.04 & 1.14 \\ 0.36 & 0.93 & 0.94 \\ 1.79 & 1.05 & 1.04 \\ 0,88 & 0.98 & 0.90 \\ 0.84 & 1.02 & 1.10 \\ 0.76 & 0.93 & 0.85 \\ 0.55 & 1.03 & 1.03 \end{pmatrix}$$

Matrix C

$$\begin{pmatrix} 0.26 & 0.99 & 0.89 \\ 0.39 & 0.94 & 0.97 \\ 0.14 & 0.99 & 0.85 \\ 0.85 & 1.00 & 1.05 \\ 3.43 & 1.00 & 1.01 \\ 1.36 & 1.04 & 1.10 \\ 0.36 & 1.00 & 1.00 \\ 2.19 & 1.10 & 1.24 \\ 0.53 & 0.94 & 0.93 \\ 3.16 & 1.04 & 1.09 \\ 0.11 & 0.95 & 0.97 \\ 1.18 & 1.04 & 1.01 \\ 0.30 & 0.99 & 0.88 \\ 0.99 & 1.03 & 1.06 \\ 0.32 & 0.93 & 0.93 \\ 0.43 & 1.02 & 1.02 \end{pmatrix}$$

Matrix $\mathbf{A}^{+}$

$$\begin{pmatrix} 0.70 & 0..90 & 0.80 \\ 0.61 & 0.84 & 0.87 \\ 0.72 & 0.87 & 0.74 \\ 0.92 & 0.99 & 1.05 \\ 1.99 & 1.26 & 1.26 \\ 1.11 & 1.07 & 1.13 \\ 0.88 & 0.98 & 0.99 \\ 1.36 & 1.27 & 1.50 \\ 0.79 & 0.87 & 0.81 \\ 1.00 & 1.05 & 1.12 \\ 0.59 & 0.84 & 0.82 \\ 1.72 & 1.24 & 1.23 \\ 1.03 & 0.95 & 0.82 \\ 0.89 & 1.01 & 1.06 \\ 0.96 & 0.91 & 0.83 \\ 0.73 & 0,95 & 0.97 \end{pmatrix}$$

Matrix $\mathbf{B}^{+}$

$$\begin{pmatrix} 0.74 & 0.91 & 0.87 \\ 0.56 & 0.82 & 0.85 \\ 0.58 & 0.85 & 0.73 \\ 1.03 & 1.02 & 1.03 \\ 1.76 & 1.20 & 1.21 \\ 1.65 & 1.21 & 1.26 \\ 0.86 & 0.96 & 0.92 \\ 1.59 & 1.27 & 1.38 \\ 0.82 & 0.89 & 0.92 \\ 0.99 & 1.03 & 1.08 \\ 0.56 & 0.84 & 0.84 \\ 1.54 & 1.16 & 1.16 \\ 0.90 & 1.00 & 0.91 \\ 0.91 & 1.00 & 1.04 \\ 0.80 & 0.89 & 0.85 \\ 0.71 & 0.95 & 0.95 \end{pmatrix}$$

Matrix $\mathbf{C}^{+}$

$$\begin{pmatrix} 0.48 & 0.85 & 0.80 \\ 0.58 & 0.85 & 0.87 \\ 0.40 & 0.83 & 0.76 \\ 0.91 & 0.98 & 1.01 \\ 2.62 & 1.41 & 1.41 \\ 1.27 & 1.10 & 1.13 \\ 0.57 & 0.89 & 0.89 \\ 1.85 & 1.31 & 1.38 \\ 0.67 & 0.87 & 0.87 \\ 2.46 & 1.40 & 1.43 \\ 0.39 & 0.81 & 0.82 \\ 1.13 & 1.06 & 1.04 \\ 0.51 & 0.86 & 0.80 \\ 1.01 & 1.03 & 1.04 \\ 0.52 & 0.83 & 0.83 \\ 0.62 & 0.92 & 0.92 \end{pmatrix}$$

As explained in Sect. 2, to decide whether inter-state disparity increased between 1991 and 2012, we have check whether A strictly Lorenz dominates C in the multi-dimensional sense, i.e. whether A P^* C where P^* is the asymmetric component of the relation L^* defined in that section. For that, in turn, we have to check whether it is true that A L^* C but not C L^* A, i.e. that $\mathbf{a}_0\ L\ \mathbf{c}_0$ but not $\mathbf{c}_0\ L\ \mathbf{a}_0$.

Recall the definition of the unidimensional weak Lorenz dominance relation L. For any non-negative n-vector **x,** let **x*** denote the vector obtained by re-arranging, if necessary, the entries in **x** in non-decreasing order (with ties broken arbitrarily). Let x^*_i be the i-th entry in **x***. For any $k = 1, 2,\ldots, n$, let $x_k^{**} = \sum_{i=1}^{k} x_i^*$ Finally, let **x**** be the n-vector whose k-th entry is x^{**}_k. Now, for any two non-negative n-vectors **x** and **y** with the same (positive) mean, **x** L **y** if and only if **x**** $\geq$ **y****. What this means is that **x** L **y** if and only if, for all $k = 1, 2, \ldots, n$, the cumulative value of the variable in question going to the *bottom k* states in **x** is not less than that in **y**.

For any A in **A,** let A* be the matrix whose j-th column, $\mathbf{a}^{*j}$, is $(\mathbf{a}^{+j})^*$ as defined in Sect. 2. Table 3 presents the matrices A*, B* and C*, obtained, respectively, from A, B and C, the three normalised well-being matrices of our interest. In Table 4, we display the vectors $\mathbf{a}_0$, $\mathbf{b}_0$ and $\mathbf{c}_0$ calculated from A*, B* and C*, respectively, as per the procedure described in Sect. 2. Since the entries in the columns of all of the three matrices are in non-decreasing order, so are the entries in $\mathbf{a}_0$, $\mathbf{b}_0$ and $\mathbf{c}_0$. Thus, $\mathbf{a}_0 = \mathbf{a}_0{}^*$, $\mathbf{b}_0 = \mathbf{b}_0{}^*$ and $\mathbf{c}_0 = \mathbf{c}_0{}^*$. Since we are interested in Lorenz comparisons between these vectors, we also display $\mathbf{a}_0{}^{**}$, $\mathbf{b}_0{}^{**}$ and $\mathbf{c}_0{}^{**}$ where the double-starred vectors are as defined above.

Examining the last three columns of Table 4, we see that $\mathbf{a}_0{}^{**} \geq \mathbf{c}_0{}^{**}$ and $\mathbf{b}_0{}^{**} \geq \mathbf{c}_0{}^{**}$. However, neither $(\mathbf{c}_0{}^{**} \geq \mathbf{a}_0{}^{**})$ nor $(\mathbf{c}_0{}^{**} \geq \mathbf{b}_0{}^{**})$. Therefore, $\mathbf{a}_0$ L $\mathbf{c}_0$ and

Table 3 Matrices A*, B* and C*

A*			B*			C*		
a^{*1}	a^{*2}	a^{*3}	b^{*1}	b^{*2}	b^{*3}	c^{*1}	c^{*2}	c^{*3}
0.59	0.84	0.74	0.56	0.82	0.73	0.39	0.81	0.76
0.61	0.84	0.80	0.56	0.84	0.84	0.40	0.83	0.80
0.70	0.87	0.81	0.58	0.85	0.85	0.48	0.83	0.80
0.72	0.87	0.82	0.71	0.89	0.85	0.51	0.85	0.82
0.73	0.90	0.82	0.74	0.89	0.87	0.52	0.85	0.83
0.79	0.91	0.83	0.80	0.91	0.91	0.57	0.86	0.87
0.88	0.95	0.87	0.82	0.95	0.92	0.58	0.87	0.87
0.89	0.95	0.97	0.86	0.96	0.92	0.63	0.89	0.89
0.92	0.98	0.99	0.90	1.00	0.95	0.67	0.92	0.92
0.96	0.99	1.05	0.91	1.00	1.03	0.91	0.98	1.01
1.00	1.01	1.06	0.99	1.02	1.04	1.01	1.03	1.04
1.03	1.05	1.12	1.03	1.03	1.08	1.13	1.06	1.04
1.11	1.07	1.13	1.54	1.16	1.16	1.27	1.10	1.13
1.36	1.24	1.23	1.59	1.20	1.21	1.85	1.31	1.38
1.72	1.26	1.26	1.65	1.21	1.26	2.46	1.40	1.41
1.99	1.27	1.50	1.76	1.27	1.38	2.62	1.41	1.43

Note For $j = 1, 2, 3$, $\mathbf{a}^{*j}$ is the non-decreasing re-arrangement of $\mathbf{a}^{+j}$, the j-th column of the matrix A^+. Similar remarks apply to $\mathbf{b}^{*j}$ and $\mathbf{c}^{*j}$

Table 4 Lorenz comparisons of the matrices A, B and C

a_0	b_0	c_0	a_0**	b_0**	c_0**
0.72	0.70	0.65	0.72	0.70	0.65
0.75	0.75	0.68	1.47	1.45	1.33
0.80	0.76	0.70	2.27	2.21	2.03
0.80	0.82	0.72	3.07	3.03	2.75
0.82	0.83	0.73	3.89	3.86	3.48
0.84	0.87	0.77	4.73	4.73	4.25
0.90	0.90	0.77	5.63	5.63	5.02
0.94	0.91	0.80	6.57	6.54	5.82
0.96	0.95	0.84	7.53	7.49	6.66
1.00	0.98	0.97	8.53	8.47	7.63
1.02	1.02	1.03	9.53	9.49	8.66
1.07	1.05	1.08	10.62	10.54	9.74
1.10	1.29	1.17	11.72	11.83	10.91
1.28	1.33	1.51	13.00	13.16	12.42
1.41	1.37	1.76	14.41	14.53	14.18
1.59	1.47	1.82	16.00	16.00	16.00

Note $\mathbf{a}_0 = (1/3)(\mathbf{a}^{*1}+\mathbf{a}^{*2}+\mathbf{a}^{*3})$ where $\mathbf{a}^{*1}$, $\mathbf{a}^{*2}$ and $\mathbf{a}^{*3}$ are as in Table 3; $\mathbf{b}_0$ and $\mathbf{c}_0$ are analogously calculated. $\mathbf{a}_0$**, $\mathbf{b}_0$** and $\mathbf{c}_0$** are as defined in the text

$\mathbf{b}_0$ L $\mathbf{c}_0$, but neither $\mathbf{c}_0$ L $\mathbf{a}_0$ nor $\mathbf{c}_0$ L $\mathbf{b}_0$. Hence, A L^* C and B L^* C, but neither C L^* A nor C L^* B. Therefore, A P^* C and B P^* C. In other words, we find evidence that inter-state disparity in well-being in India in 2012 was unambiguously greater than that in 2002 and also greater than that in 1991.

However, the comparison between the years 1991 and 2002 yields an ambiguous result. As is seen from the table, we have neither $\mathbf{a}_0^{**} \geq \mathbf{b}_0^{**}$ and nor $\mathbf{b}_0^{**} \geq \mathbf{a}_0^{**}$. For instance, the first entry in $\mathbf{a}_0$**, 0.72, is greater than the corresponding entry, 0.70, in $\mathbf{b}_0$**, but the last but one entry in $\mathbf{a}_0$**, 14.41, is less than the corresponding entry, 14.53, in $\mathbf{b}_0$**. Hence, neither A L^* B nor B L^* A. No definitive conclusion is possible regarding whether or not inter-state disparity increased between these two years.

4 Concluding Remarks

In this paper, we have taken a re-look at the trend in disparity of well-being among the states of India. We have, however, concentrated on the period since the introduction of economic reforms in 1991 and on the sixteen most populous states of the country. We have compared between the (cross-sectional) inter-state disparities in the years 1991, 2002 and 2012. The choice of these dates was dictated by the fact that per capita asset ownership (rather than per capita income) of the residents of a state was used as the indicator of *economic* well-being of the state. For the post-Reforms period, the only authentic sources of data on asset ownership by households in India are the All-India Debt and Investment Surveys which were conducted in these three years. Life expectancy and literacy rate were the other two dimensions of well-being that were considered.

We found that inter-state disparity of multidimensional well-being increased both between 1991 and 2012 and between 2002 and 2012. However, what happened to disparity between 1991 and 2002 is ambiguous.

The major differential contribution of the paper is *methodological*. By taking a multidimensional Lorenz dominance approach, we have shown that, in contrast with the findings in the existing literature, those of this paper—both the definitive and the ambiguous ones—are *robust* in the sense that they do not depend on the use of any specific inequality index as the measure of disparity. Furthermore, we have sought to handle the task of reducing well-being matrices to vectors (which is a part of the exercise) in a way that seems to be logically more satisfactory than the method followed in the existing literature.

References

Anand, I., & Thampi, A. (2016). Recent trends in wealth inequality in India. *Economic and Political Weekly, 51,* 59–67.

Atkinson, A. B., & Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *The Review of Economic Studies, 59,* 183–201.

Banerjee, A. K. (2014). A multidimensional Lorenz dominance relation. *Social Choice and Welfare, 42,* 171–191.

Banerjee, A. K. (2016). Multidimensional Lorenz dominance: A definition and an example. *Keio Economic Studies, 52,* 65–80.

Banerjee, A., & Kuri, P. K. (2015). *Development disparities in India: An enquiry into convergence*. New Delhi: Springer.

Barro, R. J., & Sala-i-Martin, X. X. (1995). *Economic growth*. New York: McGraw-Hill.

Boland, P. J., & Proschan, F. (1988). Multivariate arrangement increasing functions with applications in probability and statistics. *Journal of Multivariate Analysis, 25,* 286–292.

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson, F. Bourguignon (Eds.), *Handbook of income distribution* (pp. 67–166). North-Holland.

Dholakia, R. H. (2003). Regional disparity in economic and human development in India. *Economic and Political Weekly, 38,* 4166–4172.

Epstein, L. G., & Tanny, S. M. (1980). Increasing generalized correlation: A definition and some economic consequences. *Canadian Journal of Economics, 13,* 16–34.

Foster, J. E. (1985). Inequality measurement. In H. P. Young (Ed.), *Fair allocation* (pp. 31–68). Providence, RI: American Mathematical Society.

Ghosh, M. (2006). Economic development and human development in Indian states. *Economic and Political Weekly, 41,* 3321–3329.

Hardy, G. H., Littlewood, J. E., & Poliya, G. (1952). *Inequalities* (2nd ed.). London: Cambridge University Press.

Kolm, S.-C. (1977). Multidimensional egalitarianisms. *The Quarterly Journal of Economics, 91,* 1–13.

Marshall, A. W., & Olkin, I. (1979). *Inequalities: Theory of majorization and its applications*. New York: Academic Press.

Martin, R., & Sunley, P. (1998). Slow convergence? The new endogenous growth theory and regional Development. *Economic Geography, 74,* 201–227.

Mishra, S., & Nathan, H. S. K. (2014). Measuring HDI—The old, the new and the elegant: Implications for multidimensional development and social inclusiveness. LSE Asia Research Centre Working Paper 63.

Noorbakhsh, F. (1998). The human development index: Some technical issues and alternative indices. *Journal of International Development, 10,* 589–605.

Noorbakhsh, F. (2003). Spatial inequality and polarisation in India. CREDIT Research Paper No. 03/16. http://ssrn.com/abstract=482481.

Reserve Bank of India. (2016). *Handbook of statistics on Indian states*. Mumbai: RBI.

Roy, H., & Bhattacharjee, K. (2009). Convergence of human development across Indian states. http://www.igidr.ac.in/pdf/publication/PP-062-22.pdf.

Sen, A. (1997). *On economic inequality*. Oxford: Oxford University Press.

Tsui, K.-Y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Journal of Economic Theory, 67,* 251–265.

Correction to: Informality and Disaster Vulnerability: A Preliminary Exploration

Lopamudra Banerjee and Snehashish Bhattacharya

Correction to:
Chapter "Informality and Disaster Vulnerability: A Preliminary Exploration" in: S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_22

The original version of the book was inadvertently published with an incorrect abstract in the Chapter "Informality and Disaster Vulnerability: A Preliminary Exploration". The book has now been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-13-9981-7_22

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7_26

A Personal Reflection on Professor Sarmila Banerjee

Anjan Chakrabarti

Sarmila Banerjee or Sarmiladi as, she was known to those younger than her, was never formally my teacher. When I was in the Master's programme of Economics Department, University of Calcutta, she was on leave from the University completing her Ph.D. from the University of Florida, Gainesville. We knew her as an expert in Econometrics, a specialization and interest she maintained till the end of her teaching career at the University. As Sarmiladi rejoined the department and I left it to get my higher education and subsequently moved into jobs elsewhere; we never had a chance to meet for the next ten years. My real association with her started from 2000 onwards, when I joined the Economics Department as a faculty and became her colleague. By then, she has developed an interest in Environmental Economics and with time became an authority in it.

As we came to know one other, it was clear that our theoretical approaches and interest did not match. My research orientation in Marxian economics and hers in Neoclassical economics meant that we were poles apart, and this remained true till her retirement. Yet, over a period of time, we developed a friendship for both academic and personal reasons. Digging into the historical repository of the friendship and exchanges, I can only lay down my sense of Sarmiladi's intellectual life as an economist and person.

While she was much senior to me, both of us were witnessing during the early part of this century a fundamental transition in the culture of Indian social science academia, including the economics profession. It is important to recall that we grew up and developed our career when some of our brilliant teachers and later colleagues (which included her husband, the formidable Arup Mallik along with Ajit Chaudhury, Kalyan Sanyal and others) could criss-cross easily between Arrow-Debreu, Marx and Sraffa and some even teach them in different years. This period of great polymaths was before the present era of super-specialization, wherein the pursuit of gaining knowledge of the subject and teaching was given equal, if not more, value than the rate of publications. While oral tradition was emphasized, it was not that these scholars did not write. When they did publish, some more than the others, we knew that it was original and to be noted. At other times, and I have been witness to this on numerous occasions, they just distributed freely very original ideas orally

S. Bandyopadhyay and M. Dutta (eds.), *Opportunities and Challenges in Development*, https://doi.org/10.1007/978-981-13-9981-7

among their students and colleagues. I was among the beneficiaries and, I believe, so was Sarmiladi. All in all, the department became the breeding ground for shaping generations of students, some very well-known today. Despite this tradition among these scholars in the department who still were our colleagues in the early period of this century, both Sarmiladi and myself along with some of our colleagues quickly became mindful of the importance of rate of published research, competition and burden whose presence we could no longer ignore. It is at this point that we started to, quite unwittingly, develop a common objective and space. We were caught in the cusp, neither able to disparage the importance of deep thinking, teaching and distribution of ideas among the students, which were the hallmarks of the older tradition, nor demote the importance of abundant research in written form. It was for us no longer a question of who did what kind of economics but that of how to do economics and how to preserve the quality and presence of the department in the era of fast-paced economic transition, market penetration and disciplinary shift towards research by numbers. When we were still ruminating over the question of *what is to be done* during this period, Sarmiladi was among the first few in our department to aggressively place emphasis on a high-quality research programme in the department, showcasing it and gradually giving it an institutional form. But, she also refused to compromise on the importance of the pursuit of knowledge production and teaching. Possessing a very sharp mind and memory, she went on to become for us the perfect exemplification of teacher and researcher combined into one, backed by a no-nonsense attitude to excellence in both. For good or bad, she just could not take sloppy thinking and a casual attitude to teaching. Similarly, she had contempt for the same when it came to research. For all these reasons, she remained an inspiration for many of us and unofficially acquired a leadership position in the department.

By making the department literally her temple, she immersed herself into maintaining and protecting its sanctity. She was fiercely loyal to it and was renowned for her punishing schedule and absolute dedication, something she wanted her colleagues in the department and her MSc/M.Phil./Ph.D. students to also follow. To be honest, this attitude at times was also a source of friction with her colleagues but in the two decades of my association with her I have not found a single person in the department and beyond who ever doubted her sincerity and dedication. Her total commitment to work and department made people listen to her. This was a respect that was earned through example and not something that she got by virtue of being a senior colleague. While she was loved among the economics fraternity in West Bengal and beyond, which she returned in full amount, it will be wrong to assume that she herself distributed academic respect freely. Her dislike for sloppy and casual thinking meant that she was very choosy with whom she would discuss the matter of economics and exchange ideas and opinions in general. Her ability to separate the chaff from the wheat was something I greatly admired of her.

Because of our historical antecedent and understanding of how the discipline of economics ought to be, both of us maintained a broad interest in social science in general. Notwithstanding our diverse theoretical commitment and specialization, we shared the value of maintaining a heterodox attitude to economics which we wanted our department to imbibe. She believed in giving the students a comprehensive education in economics rather than a one-sided, singular glance at it in the form of Neoclassical economics. Commitment to such an academic culture brought us closer. While we discussed a number of things and at times exchanged course outlines, ideas and research notes for comments (from an external lens as we liked to say), I will only point to one such exchange and collaboration that remains in written form. I am referring to our jointly edited book *Development and Sustainability: India in a Global Perspective* from Springer. It was never a collaboration of convenience but one that we both wanted to be in, with a shared commitment to give something back to the incoming M.Phil./Ph.D. students and young researchers. There was also an underlying curiosity to find out whether and how far two persons from two different orientations in economics, Neoclassical and Marxian, could come to work together in a project of over two years and that too without any compromise on rigour. Rather than becoming an occasion of drifting apart (closeness often does that), the intense work and exchanges during the collaboration in which we were fully invested brought us even closer. This experience also solidified my conviction that Sarmiladi was a neoclassical economist with a heterodox mindset, that was open to ideas and ways of thinking. She did not just believe in it but practised it.

I have no qualms in admitting that as a colleague and friend, I learnt from her about the importance of values of responsibility, work ethic, perfection, camaraderie, loyalty, compassion and patience. In the later period of our friendship in the department, we shared many things. We could do that because of the absolute trust we had developed for each other. She was fiercely independent and yet fully committed to her family and social issues. It was a wonder to me how she could balance the contradictory pressures of home and work with such commitment to both, and that too with a smiling face and absolute grace. At times, I felt that the strands of contradictions in her life would just pull her apart, but quite remarkably she rode all that out. When she retired from the department, which was her life and love for nearly four decades, she left it completely. She has not touched it since. My direct and daily contact with her also ended with that, except for occasional public encounters and a promise to see each other to continue our exchange of ideas, which we have so far failed to realize. Yet, I know that mutual love and respect remains. In fact, despite receiving offers to head this or that place, I understand that she has disconnected herself completely from the discipline of economics and any institutional affiliation. This does not mean that she had given up on the bagful of knowledge gained in all these decades. Instead, she has decided to blend it with other thought worlds and direct her attention towards new frontiers, engaging with social issues through practical engagements of self- and social transformation. I have rarely come across a more courageous person than her. She has not only taken up huge challenges in her life but made it a point to constantly redefine herself and her path when she felt like doing so or when the time demanded her to take the leap. This brings me to the final point.

Rabindranath Tagore brought our attention to two sides/kinds of individual, “The self which displays itself, and the self which transcends itself and thereby reveals its own meaning”. In the academic world I had grown up in, Sarmiladi was one of the last breed of scholars who symbolized the latter. She believed that while the displaying self may have utilitarian benefits, it is utterly hollow in itself. Her post-retirement pursuit is a continuation of her embodied self-reflective and self-transcending attitude to life, a pursuit where revelation and not display is valued. In this context, in the course of our friendship and exchange of ideas, I learnt a secret which I do not mind sharing. Despite her mastery and work in Neoclassical economics and applied economics with its emphasis on competition, I was slowly convinced that she remained a closet follower of Rabindranath Tagore and his philosophy of love/cooperation and reconstruction. Her passion and constant endeavour for bringing up issues of societal and ecological concern into Neoclassical economics and research were driven in part by the contradictory pull towards the Tagorean philosophical angle. Till her retirement, particularly in the last phase, I found her discomforted with this uneasy alliance and trying to reconcile and balance the earthly academic work with her passion for humanist concern which came from the inside. Unlike many who bemoaned her exit from pure academics in which she undoubtedly still has much to contribute, I, for one, am not surprised to know that Sarmiladi has decided to pursue her deep-seated love and passion in the socially meaningful ways. This is what she is. How many of us would have the courage to take this leap into the unknown!