ALTERNATIVE APPROACHES IN MACROECONOMICS

Essays in Honour of John McCombie

Edited by
Philip Arestis

Alternative Approaches in Macroeconomics

Philip Arestis
Editor

Alternative Approaches in Macroeconomics

Essays in Honour of John McCombie

palgrave
macmillan

Editor
Philip Arestis
Department of Land Economy
University of Cambridge
Cambridge, United Kingdom

ISBN 978-3-319-69675-1 ISBN 978-3-319-69676-8 (eBook)
https://doi.org/10.1007/978-3-319-69676-8

Library of Congress Control Number: 2018930525

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

It is a pleasure and privilege to write a short Foreword to this volume edited by long-standing mutual friend Philip Arestis in honour of my long-standing friend, John McCombie.

John McCombie is one of the finest scholars I have ever met. Most of all, I admire his integrity, independence and courage in taking on the entrenched mainstream use of aggregate production functions, culminating in his definitive volume with Jesus Felipe, *The Aggregate Production Function and the Measurement of Technical Change: 'Not Even Wrong'* (Edward Elgar 2013). He is also internationally respected for his writings on regional economics.

John McCombie is a shining example of what university and community citizenship should be, and so it is fitting that he should now be honoured in this splendid volume, with contributions from friends and colleagues.

GC Harcourt
Emeritus Reader in the History of Economic Theory, Cambridge (1998)
Emeritus Fellow Jesus College, Cambridge (1998)
Professor Emeritus, Adelaide (1988)
Honorary Professor, UNSW Sydney (2016–)

Contents

Notes on Contributors

Philip Arestis is Professor and Director of Research, Cambridge Centre for Economics and Public Policy, Department of Land Economy, University of Cambridge, UK, and Professor of Economics at the Department of Applied Economics V, University of the Basque Country, Spain. He is also Distinguished Adjunct Professor of Economics, University of Utah, USA, and Research Fellow, Levy Economics Institute, New York, USA. He served (2005–2013) as Chief Academic (External) Adviser to the UK Government Economic Service (GES) on Professional Development in Economics. He is holder of the 'Queen Victoria Eugenia' award of the British Hispanic Chair of Doctoral Studies and has been awarded 'homage' by the Brazilian Keynesian Association (AKB) for his contribution to Economics in Brazil. He is holder of the FAcSS (Fellow of the Academy of Social Sciences). He has published as sole author or editor, as well as co-author and co-editor, a number of books and papers in academic journals.

Michelle Baddeley is Professor at the Institute for Choice, UniSA Business School, University of South Australia, and was Professor in Economics and Finance of the Built Environment at UCL; previously she was Fellow and Director of Studies (Economics) at Gonville & Caius College, Cambridge. She has undergraduate degrees in Economics and Psychology from University of Queensland and an MPhil/PhD (Economics) from University of Cambridge. She has written books and articles and papers across a range of topics, including behavioural economics, neuroeconomics, cybersecurity, applied macroeconomics, regional economics and development economics. She is on the editorial boards of the *Journal of Cybersecurity*, the *American Review of Political Economy* and the *Journal of Behavioral Economics and Policy*, as well as the *Society for the Advancement of Behavioral Economics* (SABE)'s advisory board. She has an active interest in public policy and was a member of DEFRA's Hazardous Substances Advisory Committee 2013–2017, is an Associate Fellow—Cambridge Centre for Science and Policy—and was a member of the Blackett Review Expert Panel: FinTech Futures 2014–2015.

Gustavo Britto obtained a PhD from the University of Cambridge in 2008, where he worked at the Land Economy Department. He is currently a professor at the Department of Economics, Universidade Federal de Minas Gerais (UFMG), Brazil, where he lectures Development Economics to undergraduate and post-graduate students. He was the Head of the Department of Economics (UFMG) from 2014 to 2016. Currently, he is the vice-coordinator of the Postgraduate Programme in Economics at the Center for Development and Regional Planning (Cedeplar-UFMG). He has carried out research and supervisions on development economics and growth, balance-of-payments constrained and cumulative causation growth models, increasing returns to scale and regional productivity growth, regional development and planning, global innovation networks and university-industry interactions and on economic complexity and structural change. The results of these efforts can be seen in journals such as the *Cambridge Journal of Economics*, *Journal of Post Keynesian Economics, Metroeconomica and Scientometrics*.

Paul Davidson holds a bachelor's degree in Biochemistry from Brooklyn College, a MBA degree from the Baruch School of Business at City College of New York, and a PhD from the University of Pennsylvania. Davidson has the Emeritus Chair of Excellence in Political Economy at the University of Tennessee. He has held professorships at the Wharton School of Business, the University of Pennsylvania and at Rutgers University where he was also the Head of the Economics Discipline. He also held visiting appointments at the University of Bristol and Cambridge University in the UK, as well as Visiting Professorships at the University of Nice and the University of Strasbourg. He was the co-founder of the *Journal of Post-Keynesian Economics*. He is a member of the Advisory Board of the Institute for New Economic Thinking. Davidson is the author or co-author of 31 books and more than 200 professional articles.

Jesús Ferreiro is Professor of Economics at the Department of Applied Economics V, the University of the Basque Country UPV/EHU, in Bilbao, Spain; he is also an Associate Member of international research centres, like the Centre for Economic and Public Policy, University of Cambridge (UK), and the Núcleo de Investigação em Finanças Públicas e Política Monetária (NIFIP), at the University of Porto. His main research interests are in the areas of fiscal and monetary policies, labour market and wages and incomes policies, financialisation processes, Eurozone, European economies and international economy. He has published a high number of papers on these topics in edited books and also in international journals such as *American Journal of Economics and Sociology*, *Applied Economics*, *Economic and Industrial Democracy*, *European Planning Studies*, *International Labour Review*, *International Review of Applied Economics*, *Journal of Economic Issues*, *Journal of Economic Policy Reform*, *Journal of Post Keynesian Economics*, *Panoeconomicus* and *Transnational Corporations*, among others.

Giuseppe Fontana is Professor of Monetary Economics at the University of Leeds (UK), Associate Professor at the University of Sannio (Italy), Research Associate at the Levy Economics Institute (USA) and Associate Member at the Cambridge Centre for Economic and Public Policy (University of Cambridge, UK). He is a member of the co-ordinating team and work-package leader of the €8 million EU FP7 research project FESSUD (http://fessud.eu/). He is an expert of the Endogenous Money Theory and the New Consensus Macroeconomics Theory and of monetary and fiscal policies. He has authored and co-authored over 30 book chapters and over 50 international journal papers. He has also co-edited several books, including *Macroeconomic Theory and Macroeconomic Pedagogy* (Palgrave/Macmillan, 2010) with Mark Setterfield and *The Global Economic Crisis: New Perspectives on the Critique of Economic Theory and Policy* (Routledge, 2011) with Emiliano Brancaccio. He has also published the monograph *Money, Time, and Uncertainty* (Routledge, 2010).

Marc Lavoie holds a senior research chair from the University Sorbonne Paris Cité, located at the University of Paris 13. He has previously taught for 37 years at the Department of Economics at the University of Ottawa. He is also a research fellow at the Macroeconomic Research Institute of the Hans Böckler Foundation in Düsseldorf and a research associate at the Broadbent Institute in Toronto. Lavoie has published close to 200 articles or book chapters in a wide variety of fields, in particular macroeconomics and monetary economics. With Wynne Godley he has written *Monetary Economics: An Integrated Approach to Money, Income, Production and Wealth* (2007). He has recently edited *Wage-led*

Growth: An Equitable Strategy for Economic Recovery (2013, with E. Stockhammer) and also *In Defense of Post-Keynesian and Heterodox Economics* (2013, with F. Lee). His latest work is *Post-Keynesian Economics: New Foundations* (2014)—an exhaustive account of Post-Keynesian economic theory.

Selen Ozcelik received her BA degree in Economics from the Faculty of Political Science at Ankara University, Turkey, in 2013 and her MA degree in Economics from the New School for Social Research in 2017. She is currently a doctoral student and Prize Fellow in the Department of Economics at the New School for Social Research. Her research interests include Post-Keynesian theories of growth and distribution and classical political economy.

Marco Veronese Passarella is a lecturer of economics, based in the Economics Division of the University of Leeds (UK). He received a Laurea in economics at the University of Bologna in 2001 and a PhD in History of Economics thought in 2008 at the University of Florence. From 2009 to 2011, Marco worked as a teaching and research fellow in many Italian universities. He has been working at the University of Leeds since January 2012. His areas of expertise are theories of value and distribution, macroeconomics modelling, monetary economics and history of economic thought. He has authored a number of articles in economics journals, such as the *Cambridge Journal of Economics*, the *Journal of Economic Behavior & Organization*, the *Review of Political Economy*, and *Metroeconomica*. He has also authored several chapters in collective volumes, as well as reviews, popular articles and books. He is a member of the Reteaching Economics group.

Mark Roberts is a senior economist with the Social, Urban, Rural and Resilience Global Practice of the World Bank, where his work is primarily focused on the South Asia and Latin American and Caribbean regions. Prior to joining the Bank, Mark was a lecturer in Spatial Economics at the University of Cambridge and a fellow of Murray Edwards College, Cambridge. Mark has published widely in leading peer-reviewed journals on the subject of spatial economic development and is a former co-editor of the journal *Spatial Economic Analysis*. He is co-author of the Bank's South Asia Region Flagship Report, *Leveraging Urbanization in South Asia*, and was a main contributor to the 2009 World Development Report, *Reshaping Economic Geography*. He holds a PhD in Land Economy and an MA in Economics from the University of Cambridge, UK, as well as a MSc in Economics from Warwick University, UK.

João P. Romero obtained a PhD from the University of Cambridge in 2015, where he worked at the Land Economy Department. He is currently an associate professor at the Department of Economics, Universidade Federal de Minas Gerais (UFMG), Brazil, where he lectures International Economics, International Monetary Economics, and Development Economics to undergraduate and postgraduate students. In 2016, his PhD dissertation won the BRICS Economic Research Award, sponsored by Export-Import Bank of India. Since joining UFMG in 2016, his research has focused on demand-led growth models combining Kaldorian and Schumpeterian approaches with special attention to the investigation of the determinants productivity growth. He also has been carrying out applied work on the connections between long-term economic development, economic complexity and structural change. The results of these efforts can be seen in journals such as the *Review of Political Economy*, *Cambridge Journal of Economics*, *Metroeconomica* and *International Review of Applied Economics*.

Malcolm Sawyer is Emeritus Professor of Economics, University of Leeds, UK, and previously held a chair in economics at University of York. He is the lead co-ordinator for the EU-funded 8 million euro, 15 partners, a five-year project on Financialisation Economy Society and Sustainable Development (www.fessud.eu). Formerly Pro-Dean for Learning and Teaching, Leeds University Business School. He established and is managing editor of *International Review of Applied Economics*, and he is the editor of the book series *New Directions in Modern Economics* (Edward Elgar). He has authored 12 books (the most recent with Philip Arestis on Financial Liberalisation: Past, Present and Future with Palgrave Macmillan). He has also published over 100 papers in refereed journals on a wide range of topics including papers on financialisation and effects of financial sector on economic performance, the Eurozone and its failing, fiscal policies and alternatives to austerity, alternative monetary policies, public–private partnerships.

Mark Setterfield is Professor of Economics in the Department of Economics at the New School for Social Research in New York. He is also a member of the Centre d'Économie de l'Université Paris Nord (CEPN) at l'Université Paris XIII (France), an associate member of the Cambridge Centre for Economic and Public Policy at Cambridge University (UK) and a senior research associate at the International Economic Policy Institute at Laurentian University (Canada). He is the author of *Rapid Growth and Relative Decline: Modelling Macroeconomic*

Dynamics with Hysteresis (Macmillan, 1997), editor (or co-editor) of seven volumes of essays and has published in numerous peer-reviewed journals including the *Cambridge Journal of Economics*, *Journal of Post Keynesian Economics*, *European Economic Review*, *Review of Political Economy*, *Journal of Economic Issues* and *The Manchester School*. He and co-author Gilberto Lima were winners of the 2010 Haralambos Simeonidis Prize, awarded by the Brazilian Association of Graduate Programs in Economics (ANEPC).

Marta R.M. Spreafico is research fellow at the Faculty of Economics, Catholic University of the Sacred Heart, Milan, and an associate member of Cambridge Centre of Economic and Public Policy, Department of Land Economy, University of Cambridge, UK. She undertook her PhD ('*Doctor Europaeus*') in Economic Policy at the Catholic University of the Sacred Heart, Milan, Italy, and the thesis is entitled *Institutions and Growth: The Experience of the Former Soviet Union Economies.* Marta Spreafico's areas of research are the economics of inequality and poverty, economic growth and the role of institutions, macroeconomics and the subprime crisis, development and regional economics, panel data estimation and spatial econometrics. She has contributed to several funded research projects. Her more recent work includes the Icelandic banking collapse, the role of capital controls, productivity growth and spatial inequality in China, the 2008 Great Recession, the increased income share of the top one percent and financialisation.

Tony Thirlwall is Professor of Applied Economics at the University of Kent and has taught there (on and off) for 50 years. At the same time, he has held visiting positions at the University of Papua New Guinea, West Virginia University, Princeton University, University of Cambridge, Melbourne University, and La Trobe University. He has been consultant to several international development organisations including the African Development Bank and the Pacific Islands Development Programme in Hawaii. He is the author of many books in the field of growth and development, many published by Palgrave Macmillan, including his best-selling textbook, jointly authored with Penelope Pacheco- Lopez, *Economics of Development* (10th edition, May 2017). He has also written an intellectual biography of Nicholas Kaldor and is the General Editor of the Great Thinkers in Economics series published by Palgrave Macmillan. Tony Thirlwall has also published papers in a number of journals.

List of Figures

List of Tables

1

Introduction

Philip Arestis

This introductory chapter is comprised of two parts. The first is an appreciation of John McCombie in terms of his contributions to economics. The second part is an introduction to the chapters that follow.

1 John McCombie: An Appreciation

It is almost by pure chance that John McCombie ended up as an economist. He essentially began his academic career when he went up to Downing College at the University of Cambridge to read for the Geographical Tripos. He had an interest in economic geography, broadly defined. After spending two years studying for this Tripos, John considered that some training in Economics was essential for a full understanding of these topics. He therefore changed subject to read for Part II of the Economics Tripos and graduated in 1973.

In the face of fierce competition, he was then awarded a prestigious Commonwealth Scholarship. The Scholarship provides funding for

P. Arestis (✉)
Department of Land Economy, University of Cambridge, Cambridge, UK

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_1

postgraduate studies at leading Commonwealth universities. This provided John with the opportunity to return to geography and to study for an MA in Geography at McMaster University in Canada. At the time, the Department of Geography there had some notable quantitative and theoretical economic geographers under whom John wished to study. It was there that he wrote his first published paper, a comment on 'Utility Accessibility and Entropy in Spatial Modelling' (McCombie 1975),[1] which was the first of many influential papers to come. After the productive year spent in Canada, John won a SSRC grant (as it then was) to return to Cambridge University to undertake research for a PhD. He was faced with an embarrassment of riches as he had offers from both the Faculty of Economics and Politics and the Department of Geography; he chose the former.

John returned to his old College, Downing, as a postgraduate student and was subsequently elected to a Bye-Fellowship of the College. So began his love of teaching and working with students as he undertook supervisions for the College for the first time. His first academic post was at the University of Hull, where the Department of Economics was looking for a regional economist (he turned down a Fulbright scholarship, which provided funding for further research to be undertaken in a leading US university, to take up the lectureship). He took up this appointment in 1977 and in 1986 took leave of absence to spend three years at the Department of Economics at the University of Melbourne, Australia. He returned to the UK in 1989 to take up a post in the Department of Land Economy at the University of Cambridge and was immediately offered a Fellowship in Economics and Land Economy at his old College. The Land Economy Tripos is an interdisciplinary subject, drawing on the disciplines, inter alios, of law, economics and real estate management. This environment suited John's interdisciplinary background and his research. He continued to publish widely and to build on his international reputation.

His move to Cambridge brought with it a host of administrative duties including being Director of Studies in the three subjects Economics, Land Economy and Management Studies for Downing College and Director of Studies for Land Economy for Christ's and Girton Colleges. With the late Nigel Allington, he built up Downing economics with its

18 students, so that in terms of economics degree results, it became one of the top five Cambridge colleges. As Director of Studies and supervisor, John excelled at the small-group teaching, gaining great pleasure from imparting knowledge to some of the brightest students in the UK. He was a very popular supervisor. He also had many PhD students, enjoying the two-way interchange of ideas. Most of his students went on to distinguished careers in academia, including professorships, or to work for international institutions such as the World Bank.

John spent the rest of his academic career at Cambridge, becoming Professor in Regional and Applied Economics, and working in his large 200-year-old study, with its wonderful views over the Downing domus—there can be few nicer places to write academic papers. During his tenure at Land Economy, he achieved substantial private funding that enabled him and Philip Arestis, as Director of Research, to establish the Cambridge Centre for Economic and Policy. The Centre also funded a Senior Research Associate. It built up an international reputation in such diverse areas ranging from monetary economics (both theory and policy), regional economic theory and policy to post-Keynesian economics, in general. In collaboration with the Department of Applied Economics V of the University of the Basque Country, the Centre has regularly organized what has now become one of the major European economics conferences, which, at the time of writing, is into its 15th year. The conference regularly leads to high-quality publications.

During his academic career, John has been invited to hold a variety of important appointments outside the University. He was a co-editor of *Urban Analysis and Policy, Regional Studies* and a founding co-editor of *Spatial Economic Analysis.* He has been an Economic Consultant to both the World Bank and the Asian Development Bank, acting as editor of the *Asian Development Review* for a number of years. He was Specialist Advisor to the House of Lords European Union Sub-Committee on the Future of the EU Structural and Cohesion Funds (2008/2009), reflecting his high standing in regional economics. He was particularly delighted to be elected Fellow of the Academy of the Social Sciences (FAcSS) and Fellow of the Regional Sciences Association (FeRSA). He was a specialist for Town and Country Planning (reflecting his interdisciplinary background) in the Teaching Quality Assessment for *the Higher Education*

Funding Council of England and the *Scottish Higher Education Funding Council* in the late 1990s. Over the years, he has received numerous invitations from universities to come as a Visiting Professor. These included the University of Castilla-La Mancha, Spain; Keio University, Japan; Pomona College, USA; the Federal University of Minas Gerais, Brazil; and the University of Otago, New Zealand. He is also a Visiting Fellow at Centre for Globalization Research and Queen Mary University of London.

There is no doubt that a powerful influence on an economist's approach to his or her subject is the academic training, often for the doctorate. It is not for nothing that the Jesuits' motto is 'give me the child until he is seven and I will give you the man'. During the time John was pursuing his undergraduate and postgraduate studies in the early 1970s, there was what might be best termed a school of 'Cambridge Economics'. These included some intellectual heavyweights such as Nicky Kaldor, Joan Robinson, Luigi Pasinetti, Bob Rowthorn and Geoff Harcourt. It is probably fair to say that what gave the group its coherence was more its rejection of neoclassical economics than anything else. On the applied side, there was also Wynne Godley and the Cambridge Economic Policy Group.

John had come to Cambridge to do his PhD in the area of what was then called the 'New Urban Economics'. However, he became intrigued by the debate between Kaldor and Rowthorn at that time over what was then a deceptively simple relationship between the growth of productivity and output known as the Verdoorn law. This was interpreted as providing evidence of substantial dynamic and static increasing returns to scale.

This debate raised some intriguing issues and led first to what was going to be a temporary diversion of John's research. However, he eventually switched his PhD research to a full-time consideration of various aspects of the law and the Kaldorian approach to growth. This led to a large number of papers on the subject, including papers extending the Kaldor-Rowthorn debate (McCombie 1981a), new empirical evidence using US state data (McCombie and de Ridder 1983) and quantifying the importance of Kaldor's laws (McCombie 1980). Inevitably John's interests developed over time to other areas of what might be best termed post-Keynesian economics. But he returned to the Verdoorn law throughout his career. One thing that had bothered him was what he termed the 'Static-Dynamic Verdoorn law paradox'. Many studies using cross-regional data and conventional aggregate production functions estimated in log level found very small increasing or constant returns to scale. Paradoxically, using the same regional data set, the estimation of the Verdoorn law using growth rates gave estimates of significant increasing returns to scale. John's (McCombie 1982a) initial attempt to provide an explanation of this was not convincing, even to himself. It was not until several years later with Mark Roberts that he provided a satisfactory explanation (McCombie and Roberts 2007). It had to do with spatial aggregation bias.

Inevitably, recent developments in econometrics have led to a reconsideration of earlier studies and this is true of the extended Verdoorn law. In the latter case, it was the development of spatial autocorrelation estimation techniques. John's more recent work with his co-authors, using these more sophisticated techniques, has confirmed earlier results (Angeriz et al. 2008, 2009).

A further area of research where John has made a significant contribution has been in the area of balance-of-payments-constrained growth. This approach originated in a short note by Tony Thirlwall, published in 1979. John was at initially somewhat sceptical of this argument and this led him to publish a short comment on the analysis of the same journal (McCombie 1981b). The upshot was an invitation by Tony Thirlwall to visit him at his University to discuss their differences, which turned out to be small. There followed a fruitful collaboration on the subject for many years including a major book published in 1994. Since then there has been an explosion of the literature on balance-of-payments to which John has again returned to make influential contributions.

John has become a leading authority and developed an international reputation for the extension of a fundamental critique of the neoclassical aggregated production function. One of the outcomes of the Cambridge capital theory controversies of the late 1960s was that the results from an aggregate production function did not hold outside of a one-commodity world. Aggregation theorems also gave the same result. Over time, the Cambridge capital theory controversies have been relegated to the history of economic thought and largely forgotten. The reason seems to have been an implicit reliance on the not compelling Friedman's methodology of economics. According to Friedman, what matters is not the realism or otherwise of the assumptions of a model, but its predictive power. Douglas's statistical work in the 1930s with various colleagues using cross-industry or cross-state data found remarkably good statistical fits with the output elasticities very close to the factor shares. Time-series data usually, but not always, gave good fits. Hence, statistical estimations of aggregate production functions increased during the post-war period using ever more sophisticated statistical techniques. The good statistical fits were taken at both the textbook and advanced level to confirm that markets were competitive and factors of productions were paid their marginal products.

John found similar results and good statistical fits in the 1970s to the Cobb-Douglas aggregate production functions using the regional data he had collected and constructed for his PhD. This remained a puzzle until he came across, quite by chance, in the 1980s a small number of largely ignored papers that provided what he thought was a convincing explanation. Herbert Simon, for example, considered the critique of sufficient importance that he mentioned it explicitly in his Nobel Prize acceptance speech.[2]

If the arguments were logically correct, and John was convinced that they were, they had devastating implications for neoclassical macroeconomics. Theoretically, the aggregate production function should be expressed in *physical* terms, yet in empirical studies of the aggregate production function the measures of output and the capital stock are *constant price value terms*; not physical volumes. This is in spite of these measures misleadingly being called 'volumes', in much of the applied work on aggregate production functions. This is not an innocuous difference because it means that there is an underlying accounting identity, namely, that constant price value added equals the wage rate multiplied by the numbers employed plus the constant price value of capital multiplied by the rate of profit. It may be straightforwardly shown that if the accounting identity is differentiated, and then integrated, the result is a mathematical isomorphism that is identical to a Cobb-Douglas 'production function' with the estimated 'output elasticities' equalling the factor shares. The national accounting identity is, of course, compatible with any state of competition, whether there are increasing returns or not, and most importantly with the complete absence of an aggregate production function. The results are not dependent on factor shares being constant. If the identity is estimated, say, using data where the factor shares are changing, then a more flexible functional form (such as the CES 'production function') will give a better fit.

John first extended and revived this critique in a paper published in 1998 as nothing had been written on the topic for several years. It also led to a long and productive collaboration with Jesus Felipe in this area. In a series of papers, they examined some well-known papers that used the aggregate production function. Using only the identity and some 'stylised facts' that did not depend upon the putative aggregate production function, they were able to correctly predict the estimates before a single regression had been run. These papers include the following topics: estimating biased technical change in aggregate production functions (McCombie and Dixon 1991), estimating CES production functions (Felipe and McCombie 2001b), Hall's estimation of the mark-up in manufacturing (Felipe and McCombie 2002), Mankiw-Romer-Weil's test of the Solow growth model (Felipe and McCombie 2005b), the concept of total factor productivity (Felipe and McCombie 2007b) and estimations of labour demand functions (Felipe and McCombie 2009a). Moreover, they explained why Solow's 'startling' result that technical change explained by far the largest proportion of productivity growth was hardly surprising, when the accounting identity was taken into account. In fact it could hardly be otherwise.

What is surprising is that the argument is one of logic. There have been a few attempts to criticize the argument, none of them compelling. John and Jesus Felipe brought their papers together in a book published in 2014, *The Measurement of Technical Change and the aggregate Production Function. 'Not Even Wrong'.*

Consequently, neoclassical macroeconomists continue to behave analogously to the geocentrists who were confident that their predictions of the movement of the planets using ad hoc epicycles confirmed the view that the planets moved around the earth. The foundations of a Copernican revolution in economics are there, but it has yet to happen. This work on the aggregate production function had implications for the Verdoorn law in that Verdoorn in his original paper had derived the law from a Cobb-Douglas production function. This contradiction puzzled John until he eventually reconciled the two, in the process providing firmer foundations for Kaldor's 'technical progress function' (McCombie and Spreafico 2015).

John has also made a considerable contribution to a large number of areas within what may be best described as the post-Keynesian economics. Although he is now retired and is an Emeritus Professor of the University of Cambridge, and Emeritus Fellow in Economics at Downing College, he is still doing some teaching for the Department, attending international conferences and writing papers and books. He will no doubt continue to make an outstanding contribution to the debates in economics.

2 References

2.1 Articles

Allington, N. F. B., McCombie, J. S. L., & Pike, M. (2011). The failure of the new macroeconomic consensus: From non-ergodicity to the efficient markets hypothesis and back again. *International Journal of Public Policy*, *7*(1/2/3), 4–21.

Angeriz, A., McCombie, J. S. L., & Roberts, M. (2006). Productivity, efficiency and technical change in EU regional manufacturing: A DEA approach. *Manchester School*, *74*(4), 500–525.

Angeriz, A., McCombie, J. S. L., & Roberts, M. (2008). New estimates of returns to scale and spatial spillovers for EU regional manufacturing, 1986–2002. *International Regional Science Review*, *31*(1), 62–87.

Angeriz, A., McCombie, J. S. L., & Roberts, M. (2009). Increasing returns and growth of industries in the EU regions: Paradoxes and conundrums. *Spatial Economics Analysis*, *4*(2), 147–168.

Britto, G., & McCombie, J. S. L. (2015). Increasing returns to scale and regions: A multilevel model for Brazil. *Brazilian Keynesian Review*, *1*(2), 118–134.

Dixon, R., & McCombie, J. S. L. (1989). Inter-industry differences in the factor-augmenting bias of technological change. *Australian Economic Papers*, *28*(52), 103–111.

Felipe, J., & McCombie, J. S. L. (2001a). Biased technical change, growth accounting, and the conundrum of the East Asian miracle. *Journal of Comparative Economics*, *29*(3), 542–565.

Felipe, J., & McCombie, J. S. L. (2001b). The CES production function, the accounting identity and Occam's razor. *Applied Economics*, *33*(10), 1221–1232.

Felipe, J., & McCombie, J. S. L. (2002). A problem with some estimations and interpretations of the mark-up in manufacturing industry. *International Review of Applied Economics*, *16*(2), 187–215.

Felipe, J., & McCombie, J. S. L. (2003). Some methodological problems with the neoclassical analysis of the East Asian miracle. *Cambridge Journal of Economics*, *27*(5), 695–721.

Felipe, J., & McCombie, J. S. L. (2005a). How sound are the foundations of the aggregate production function? *Eastern Economic Journal*, *31*(3), 467–488. Reprinted in Harcourt, G. C., & Kriesler, P. (Eds.), *The oxford handbook of post-Keynesian economics.* Oxford: Oxford University Press.

Felipe, J., & McCombie, J. S. L. (2005b). Why are some countries richer than others? A sceptical view of Mankiw-Romer-Weil's test of the Solow growth model. *Metroeconomica*, *56*(3), 360–392.

Felipe, J., & McCombie, J. S. L. (2006). The tyranny of the identity. Growth accounting revisited. *International Review of Applied Economics, 20*(3), 283–299.

Felipe, J., & McCombie, J. S. L. (2007a). Is a theory of total factor productivity really needed. *Metroeconomica*, *58*(1), 195–229.

Felipe, J., & McCombie, J. S. L. (2007b). On the rental price of capital and the profit rate. The perils and pitfalls of total factor productivity growth. *Review of Political Economy*, *19*(3), 317–345.

Felipe, J., & McCombie, J. S. L. (2009a). Are estimates of labour demand functions mere statistical artefacts? *International Review of Applied Economics*, *23*(2), 147–168.

Felipe, J., & McCombie, J. S. L. (2009b). Why the data tell us nothing about returns to scale and externalities to capital in economic growth. *Economia e Sociedade*, *17*, 657–677.

Felipe, J., & McCombie, J. S. L. (2010). What is wrong with aggregate production functions. On temple's 'aggregate production functions and growth economics'. *International Review of Applied Economics*, *24*(6), 665–684.

Felipe, J., & McCombie, J. S. L. (2011/2012). On Herbert Simon's criticisms of the Cobb-Douglas and the CES production functions. *Journal of Post Keynesian Economics, 34*(2), 275–293.

Felipe, J., & McCombie, J. S. L. (2012). Problems with regional production functions and estimates of agglomeration economies: A caveat emptor for regional scientists. *Spatial Economic Analysis*, *4*, 461–484.

Felipe, J., & McCombie, J. S. L. (2014). The aggregate production function: 'Not even wrong'. *Review of Political Economy*, *26*(1), 60–84.

Felipe, J., Hasan, R., & McCombie, J. S. L. (2008). Correcting for biases when estimating production functions: An illusion of the laws of algebra? *Cambridge Journal of Economics*, *32*(3), 441–459.

Felipe, J., McCombie, J. S. L., & Naqvi, K. (2011). Is Pakistan's growth rate balance-of-payments constrained? Policies and implications for development and growth. *Oxford Review of Development Studies*, *38*, 477–496.

Fingleton, B., & McCombie, J. S. L. (1999). Increasing returns and economic growth. Evidence for manufacturing from the European Union regions. *Oxford Economic Papers*, *50*(1), 89–105.

Hein, E., & Lavoie, M. (2015). Interview with John McCombie: 'I think there's absolutely no way out for them: An aggregate production function does not make any sense at all!' *Intervention: The European Journal of Economics and Public Policy*, *12*, 1–6.

McCombie, J. S. L. (1975). Utility, accessibility and entropy in spatial modelling: A comment. *Swedish Journal of Economics*, *77*(4), 497–501.

McCombie, J. S. L. (1980). On the quantitative importance of Kaldor's laws. *Bulletin of Economic Research*, *32*(2), 102–112. Reprinted in King, J. E. (Ed.). (1994). *Economic growth in theory and practice: A Kaldorian perspective.* Cheltenham: Edward Elgar.

McCombie, J. S. L. (1981a). What still remains of Kaldor's laws? *Economic Journal*, *91*(361), 206–216. Reprinted in King, J. E. (Ed.). (1994). *Economic growth in theory and practice: A Kaldorian perspective.* Cheltenham: Edward Elgar.

McCombie, J. S. L. (1981b). Are international growth rates constrained by the balance of payments? A comment on Professor Thirlwall. *Banca Nazionale del Lavoro Quarterly Review*, *139*(December), 455–458.

McCombie, J. S. L. (1982a). Economic growth, Kaldor's laws and the static-dynamic Verdoorn law paradox. *Applied Economics*, *14*(3), 279–294.

McCombie, J. S. L. (1982b). How important is the spatial diffusion of innovations in explaining regional growth disparities? *Urban Studies*, *19*(4), 377–382.

McCombie, J. S. L. (1983). Kaldor's laws in retrospect. *Journal of Post Keynesian Economics*, *5*(3), 414–428.

McCombie, J. S. L. (1985). Increasing returns and manufacturing industries: Some empirical issues. *Manchester School*, *53*(1), 55–75.

McCombie, J. S. L. (1985–1986). Why cutting real wages will not necessarily reduce unemployment – Keynes and the "postulates of the classical economics". *Journal of Post Keynesian Economics*, *8*(2), 233–248. Reprinted in Blaug, M. (Ed.). (1991). *John Maynard Keynes (1883–1946)* (Vol. 2). Aldershot: Edward Elgar.

McCombie, J. S. L. (1986). On some interpretations of the relationship between productivity and output growth. *Applied Economics*, *18*(11), 1215–1225.

McCombie, J. S. L. (1987). Does the aggregate production function imply anything about the laws of production? A note on the Simon and Shaikh critiques. *Applied Economics*, *19*(8), 1121–1136.

McCombie, J. S. L. (1987–1988). Keynes and the nature of involuntary unemployment. *Journal of Post Keynesian Economics*, *10*(2), 202–215. Reprinted in Blaug, M. (Ed.). (1991). *John Maynard Keynes (1883–1946)* (Vol. 2). Aldershot: Edward Elgar.

McCombie, J. S. L. (1988a). A synoptic view of regional growth and unemployment: II – The post Keynesian theory. *Urban Studies*, *25*(5), 399–417. Reprinted in Haynes, K. E., et al. (Eds.). (1996). *Regional dynamics. Volume 1. Elgar reference collection. Modern classics in regional science.* Cheltenham: Edward Elgar.

McCombie, J. S. L. (1988b). A synoptic view of regional growth and unemployment: I – The neoclassical theory. *Urban Studies*, *25*, 267–281. Reprinted in Haynes, K. E., Button, K., Nijkamp, P., & Qiangshang,

L. (Eds.). (1996). *Regional dynamics. Volume 1. Elgar reference collection. Modern classics in regional science.* Cheltenham: Edward Elgar; and in Cheshire, P. C., & Evans, A. (Eds.). (1992). *Urban and regional economics, the international library of critical writings.* Cheltenham: Edward Elgar.

McCombie, J. S. L. (1991). The postwar productivity slowdown and the intersect oral reallocation of labour. *Australian Economic Papers*, *30*(56), 70–85.

McCombie, J. S. L. (1993). Economic growth, trade interlinkages and the balance-of-Payments constraint. *Journal of Post Keynesian Economics*, *5*(4), 471–505.

McCombie, J. S. L. (1997). The empirics of balance-of-payments constrained Growth. *Journal of Post Keynesian Economics*, *19*(3), 345–375.

McCombie, J. S. L. (1998). Are there laws of production? Some early criticisms of the Cobb-Douglas production function. *Review of Political Economy*, *10*(2), 141–173.

McCombie, J. S. L. (2000). Regional production functions and the accounting identity: A problem of interpretation. *Australasian Journal of Regional Studies*, *6*(2), 133–156.

McCombie, J. S. L. (2000–2001). The Solow residual, technical change, and aggregate production functions. *Journal of Post Keynesian Economics*, *23*(2), 267–297.

McCombie, J. S. L. (2001). What does the aggregate production function show? Further thoughts on Solow's 'second thoughts on growth theory'. *Journal of Post Keynesian Economics*, *23*(4), 589–615.

McCombie, J. S. L. (2012). Criticisms and defences of the balance-of-payments constrained growth model: Some old, some new. *PSL Quarterly Review* (formerly *Banca Nazionale Quarterly Review*), *64*(259), 353–392.

McCombie, J. S. L. comment on, Cohen, A., Harcourt, G. C., & Felipe, J. (2004). Whatever happened to the Cambridge capital theory controversies? *Journal of Economic Perspectives*, *1*, 299–301.

McCombie, J. S. L. (1992). 'Thirlwall's law' and balance-of-payments constrained growth: More on the debate. *Applied Economics*, *24*(5), 493–512.

McCombie, J. S. L., & de Ridder, J. R. (1983). Increasing returns, productivity and output growth: The case of the United States. *Journal of Post Keynesian Economics*, *5*(3), 373–387.

McCombie, J. S. L., & de Ridder, J. R. (1984). The Verdoorn controversy: Some new evidence using US state data. *Oxford Economics Papers*, *36*(2), 268–284.

McCombie, J. S. L., & Dixon, R. (1991). Estimating technical change in production functions: A critique. *International Review of Applied Economics*, *5*(1), 24–46.

McCombie, J. S. L., & Pike, M. (2013). No end to the consensus in macroeconomic theory? A methodological inquiry. *American Journal of Economics and Sociology*, *72*(2). Reprinted in Lee, F. S. (Ed.), *Markets, competition and the economy as a social system.* Oxford: Wiley.

McCombie, J. S. L., & Roberts, M. (2007). Returns to scale and regional growth: The static-dynamic Verdoorn law paradox revisited. *Journal of Regional Science*, *47*(2), 179–208.

McCombie, J. S. L., & Roberts, M. (2008). Effective demand constrained growth in a two-sector Kaldorian model. *Journal of Post Keynesian Economics*, *31*(1), 57–78.

McCombie, J. S. L., & Spreafico, M. R. M. (2015). Kaldor's 'technical progress function' and Verdoorn's law revisited. *Cambridge Journal of Economics*, *40*(4), 1117–1136.

McCombie, J. S. L., & Spreafico, M. R. M. (2016). Could the Icelandic banking collapse of 2008 have been prevented? The role of economists prior to the crisis. *European Journal of Economics and Economic Policies: Intervention*, *13*(3), 323–338.

McCombie, J. S. L., & Thirlwall, A. P. (1985). Economic growth, the Harrod foreign trade multiplier and the Hicks super-multiplier. *Applied Economics*, *17*(1), 55–72. Reprinted in Wood, J. C., & Woods, R. N. (Eds.). (1988). *Sir John Hicks: Critical assessments.* London: Routledge.

McCombie, J. S. L., & Thirlwall, A. P. (1997). The dynamic Harrod foreign trade multiplier and the demand-orientated approach to economic growth: An evaluation. *International Review of Applied Economics*, *11*(1), 5–26.

McCombie, J. S. L., Spreafico, M. R. M., & Xu, S. (2017). Productivity growth of the cities of the Jiangsu province, China: A Kaldorian approach. *International Review of Applied Economics*, 1–22. https://doi.org/10.1080/02692171.2017.1351529.

Ribeiro, R. S. M., Lima, G. T., & McCombie, J. S. L. (2016). Exchange rate, income distribution and technical change in a balance-of-payments constrained growth model. *Review of Political Economy*, *28*(4), 545–565.

Romero, J. P., & McCombie, J. S. L. (2016a). Differences in increasing returns between technological sectors: A panel data investigation using the EU KLEMS database. *Journal of Economic Studies*, *43*(5), 863–878.

Romero, J. P., & McCombie, J. S. L. (2016b). The multi-sectoral Thirlwall's law: Evidence from 14 developed European countries using product-level data. *International Review of Applied Economics*, *30*(3), 301–325.

Tharnpanich, N., & McCombie, J. S. L. (2013). Balance-of-payments constrained growth, structural change, and the Thai economy. *Journal of Post Keynesian Economics*, *35*(4), 569–598.

2.2 Books

Arestis, P., & McCombie, J. S. L. (Eds.). (2008). *Unemployment: Past and present*. Basingstoke: Palgrave Macmillan.

Arestis, P., & McCombie, J. S. L. (Eds.). (2009). *Missing links in the unemployment relationship*. Basingstoke: Palgrave Macmillan.

Arestis, P., Baddeley, M. C., & McCombie, J. S. L. (Eds.). (2001). *What global economic crisis?* Basingstoke: Palgrave.

Arestis, P., Baddeley, M. C., & McCombie, J. S. L. (Eds.). (2003). *Globalisation, regionalism and economic activity*. Basingstoke: Palgrave Macmillan.

Arestis, P., Baddeley, M. C., & McCombie, J. S. L. (Eds.). (2007). *Economic growth. New directions in theory and policy*. Cheltenham: Edward Elgar.

Arestis, P., McCombie, J. S. L., & Baddeley, M. C. (Eds.). (2006a). *The new monetary policy, implications and relevance*. Cheltenham: Edward Elgar.

Arestis, P., McCombie, J. S. L., & Vickerman, R. (Eds.). (2006b). *Development and economic growth: Essays in honour of A.P.Thirlwall*. Cheltenham: Edward Elgar

Felipe, J., & McCombie, J. S. L. (2013). *The aggregate production functions and the measurement of technical change: 'Not even wrong'.* Cheltenham: Edward Elgar (2013, hardback; 2014, paperback).

Fontana, G., McCombie, J. S. L., & Sawyer, M. C. (Eds.). (2010). *Macroeconomics, finance and money: Essays in honour of Philip Arestis.* Basingstoke: Palgrave Macmillan.

McCombie, J. S. L., & Gonzalez, C. R. (Eds.). (2007a). *Issues in finance and monetary policy*. Basingstoke: Palgrave Macmillan.

McCombie, J. S. L., & Gonzalez, C. R. (Eds.). (2007b). *The European Union: Current problems and prospects*. Basingstoke: Palgrave Macmillan.

McCombie, J. S. L., & Thirlwall, A. P. (1994). *Economic growth and the balance-of-payments constraint*. Basingstoke: Macmillan.

McCombie, J. S. L., Pugno, M., & Soro, B. (Eds.). (2002). *Productivity growth and economic performance: Essays on Verdoorn's law*. Basingstoke: Palgrave Macmillan.

Thirlwall, A. P., & McCombie, J. S. L. (2004). *Essays on balance of payments constrained growth*. London: Routledge.

3 Introductory Part

A number of John McCombie's colleagues who are familiar with his work have kindly offered to contribute to this book in honour of John. The rest of this introduction offers a brief summary of the chapters that follow.

Tony Thirlwall in Chap. 2, entitled, 'John McCombie's Contribution to the Applied Economics of Growth in a Closed and Open Economy', focuses on John McCombie's major contributions to our understanding of growth rate differences between countries. This chapter is divided into three parts. The first part deals with Kaldor's growth laws, and particularly John's work on Verdoorn's Law—its estimation—and resolving the static/dynamic paradox that increasing returns are found when the growth of labour productivity is regressed on the growth of manufacturing output but not when the log level of productivity is regressed on the level of manufacturing output. The second part outlines John's contribution to the theory of balance-of-payments-constrained growth, particularly showing that the dynamic Harrod trade multiplier result can be interpreted as the Hicks super multiplier. The third part shows that Kaldor's first law of growth, that manufacturing is the engine of growth, is also a reduced form of an export-led growth model, because the export growth of countries is closely related to the growth of manufactured exports.

Paul Davidson continues in Chap. 3, under the title, 'Why Neither Samuelson's Neoclassical Synthesis Keynesianism Nor New Keynesianism Theory Is Compatible with Keynes's General Theory'. This chapter demonstrates that the specified presumptions underlying Neoclassical Synthesis Keynesian Theory and New Keynesian Theory are in direct conflict with Keynes' statements regarding the foundation of his theory of involuntary unemployment. Stickiness of wages, and/or administered prices, is their fundamental cause of involuntary unemployment. Keynes's essential properties are that (1) all liquid assets have a zero elasticity of production, and therefore are nonproducible, and (2) the elasticity of substitution between all liquid assets and producible goods and services is zero. Thus, when people put their savings out of current income into the form of liquid assets, these savings find a resting place in nonproducibles even if all wages and prices are perfectly flexible. Consequently, every penny saved is a penny not earned by workers and enterprises that produce goods and services. In other words, in Keynes's general theory sticky money wages and/or administered prices are not the fundamental cause of unemployment.

Giuseppe Fontana and Marco Veronese Passarella in their Chap. 4, entitled, 'Aggregate Demand, Money and Finance in the New Consensus Macroeconomics: A Critical Appraisal', critically assess the 'New Consensus Macroeconomics' (NCM) theory and its recent developments. Building on the Wicksellian 'two-interest-rates model', the NCM highlights the role of interest rates in the transmission mechanism of monetary policy, whereas monetary aggregates are treated as residual variables. However, in contrast with Wicksell's theory, banks and financial institutions are usually neglected in the NCM theory. As a result, the financial instability and recurrent banking crises of modern economies have received little attention in modern macroeconomics. This chapter has three main goals. First, it aims to provide a critical analysis of the original NCM model and some recent developments. Second, it aims to show that few amendments to it are sufficient to account for the financial instability and banking crises of real-world economies. Third, it shows that some important policy-making conclusions logically follow once the role of banks, credit and finance is properly taken on board.

Malcolm Sawyer continues in Chap. 5 with considering the relationships between microeconomics, mesoeconomics and macroeconomics. Macroeconomic analysis has often been accused of lacking 'microeconomic foundations'. However, macroeconomic analysis always had microeconomic underpinnings though not ones based on life-time utility maximization. The use of the term 'microeconomic foundations' suggests causation runs from the micro level to the macro, whereas causation runs in both directions and involving the meso level. The notion that macroeconomic analysis can be based on the 'representative agent rational expectations' approach is strongly critiqued. The nature of macroeconomic relationships is discussed and the general proposition set out to suggest that macroeconomic conditions cannot be ready derived from microeconomic considerations. Economic analysis is only interesting when conducted above the individual level involving interactions between individuals and the resolution of their actions and decisions with consideration of whose decisions are decisive. Some of the problems surrounding the use of macro relationships which mimic micro relationships are set out.

Philip Arestis in the Chap. 6 entitled, 'A Coherent Approach to Macroeconomic Theory and Economic Policies', offers a new approach to macroeconomics, which focuses on the notion that there is often inadequacy of aggregate demand relative to what would be required for full employment of the factors of production. The level and distribution of productive capacity can often be inadequate to underpin full employment. Consequently, and under such circumstances, distributional effects are paramount and should be seriously taken on board in the analysis and policies; and such effects are actually considered in this contribution. Economic policies are thereby very relevant and important. We briefly summarize the theoretical framework that underpins the relevant economic policies before we turn our attention to the latter themselves. We suggest that in addition to the well-known economic policies, namely, fiscal and monetary policies, and of equal importance, co-ordination of them, two new, relevant and important policy dimensions emerge as paramount: distributional effects and financial stability. We also discuss briefly current ‘unorthodox’ monetary policies.

Marta Spreafico in Chap. 7, entitled ‘Is the Share of Income of the Top One Percent Due to the Marginal Product of Labour or Managerial Power?’, argues that the last 30 years have seen the rapid increase in the share of income of the top 1 per cent, especially in the USA. This has led to increasing concern in some quarters about the consequences of the increase in income inequality. However, for a long time, neoclassical economics has generally ignored the problem. This is largely because of its uncritical acceptance that all employees, including the highest paid, are paid their marginal products in competitive labour markets and receive their ‘just deserts’. The recent increase in overall inequality is also attributed to skill-biased technical change and the race between technology and education. These explanations are examined in light of empirical and theoretical arguments that question the existence of the aggregate production function and the marginal productivity theory of distribution. It is concluded that the explanation for the increase in income of the top 1 per cent must lie elsewhere such as an increase in managerial power.

Jesús Ferreiro, in the Chap. 8, entitled ‘Macroeconomic Lessons from the Financialisation Process’, has the objective in contribution to outline the main macroeconomic lessons resulting from the financialization process. This chapter is structured into four main sections. The first section will focus on the definition of the financialization process. The second section will focus on the consequences of the financialization process on economic activity in general and on the activity carried out by particular sectors and agents. The third section will deal with the Great Recession as far as there is an extended consensus on the key role played by the excessive growth of finances on the burst of the crisis. This study will pay attention to the different impacts of the economic and financial crises in European countries and on the consequences generated by the management of macroeconomic policies, mainly in developed and European countries. A final section will be devoted to the consequences of financialization on the European integration process.

Michelle Baddeley in Chap. 9 entitled, ‘Financial Instability and Speculative Bubbles: Behavioural Insights and Policy Implications’, draws on themes from Baddeley and McCombie's (2001) exploration of speculative bubbles, which applied different models of speculation to analysing famous historical speculative episodes, specifically Tulipmania and the South Sea Bubble. This chapter re-assesses these insights in the light of all that has happened during the US sub-prime mortgage crisis and subsequent global financial crises of 2007/2008. It also extends the analysis to include new insights from behavioural finance about the nature and causes of speculative bubbles, blending insights from behavioural finance and post-Keynesian economics. Speculative bubbles throughout history have a number of common, predictable features so why have we not learnt more from these past experiences? In answering these questions, this chapter concludes with an analysis of policy implications—including fiscal and monetary measures that could be implemented to minimize the destabilizing real-side impacts from speculative bubbles and the financial shocks and crises which often follow, with significant de-stabilizing impacts for real economies.

João P. Romero and Gustavo Britto in Chap. 10, entitled 'Sophistication, Productivity and Trade: A Sectoral Investigation', argue that in balance-of-payments-constrained growth models, income elasticities of exports and imports are the crucial parameters determining the long-term growth rate. Consequently, it is critical to understand what determines the level of these elasticities. The chapter investigates whether measures of productive sophistication developed by Hausmann et al. (2007) and Hidalgo and Hausmann (2009) can explain not only productivity growth but also the size of income elasticities of trade in different technological sectors. It does so by testing the impact of initial industry sophistication on subsequent productivity growth for low and high-tech industries as well as by assessing if changes in industry sophistication are associated with higher exports and imports in these sectors. The empirical investigation uses product-level trade data from UN Comtrade, combined with price data from Feenstra and Romalis (2014) and with productivity data from EU KLEMS for 13 industries from seven countries, over the period 1984–2007.

Mark Roberts in the Chap. 11, entitled 'Patterns of Urban Growth in South Asia: A View from Outer Space', examines the South Asia's case where since the turn of the century, the area has added an estimated

130 million people to its towns and cities, a number equivalent to the entire population of Japan. In the process, its urban share of the population has grown from 27.3 per cent in 2000 to 30.9 per cent in 2011, implying a pace of urbanization that has been on a par with that in sub-Saharan Africa. However, although South Asia's relatively rapid pace of urbanization over the last decade and a half is well-known, less understanding exists of patterns of physical urban expansion and economic growth across the region's cities. This chapter, therefore, makes use of data on night-time lights for the period 1999–2010 which has been remotely collected by satellites orbiting the earth to provide new insights into these patterns. The chapter furthermore explores the empirical links between urbanization and rates of extreme poverty across subnational regions within South Asia.

Marc Lavoie in Chap. 12, entitled 'Production Functions, the Kaldor-Verdoorn Law and Methodology', focuses on John McCombie contributions and argues that he has been an unrelenting critic of the neoclassical production function for over 30 years. With his co-author Jesus Felipe, along with Anwar Shaikh, he has provided a number of proofs demonstrating that the apparent empirical successes of neoclassical production functions could be attributed to the fact that these production functions were reproducing the identities of the national accounts. Kaldor's technical progress function and the Kaldor-Verdoorn equation, however, do share some similarities with these identities, and thus one may wonder if they are subjected to the same critique. It is shown that the Kaldor-Verdoorn equation is impervious to the critique. Some of the methodological considerations are advanced by John McCombie, notably those concerning the instrumentalist approach of mainstream economics and its DSGE model. The chapter concludes with a pledge in favour of meta-regression analysis, recalling that a recent such analysis has shown that the Kaldor-Verdoorn effect is genuine.

Finally, Mark Setterfield and Selen Ozcelik in Chap. 13, entitled 'Is the Balance-of-Payments-Constrained Growth Rate Time-Varying? Exchange Rate Overvaluation, Deindustrialization, and Long-Run Growth', examine the long-held view among macroeconomists in the UK and USA that sustained currency over valuation, often the result of financial-sector dominance, weakens domestic macroeconomic performance and results in premature deindustrialization. Similar concerns have been expressed about persistent, policy-induced recessions. According to balance-of payments-constrained growth (BPCG) theory, meanwhile, the BPCG rate in a multi-sector economy varies directly with the share of manufacturing in total output. This chapter develops a simple model that combines these observations to show how a temporary but persistent shock to the nominal exchange rate and/or domestic demand can both affect the actual rate of growth in the short run (by moving it away from the long-run equilibrium BPCG rate) and alter the BPCG rate itself (by lowering the income elasticity of demand for exports as a result of induced premature deindustrialization). The result is a time-varying balance-of payments constrained growth (TV-BPCG) rate. Because actual growth and the TV-BPCG rate vary directly, the latter is also characterized as quasi path dependent.

Notes

1. The numbers reported in the text refer to the ones in 'selected bibliography' as below.
2. Simon, H.A. (1979), 'Rational Decision-making in Business Organizations', *American Economic Review,* 69, pp. 493–513.

2

John McCombie's Contribution to the Applied Economics of Growth in a Closed and Open Economy

A. P. Thirlwall

1 Introduction[1]

It is a great pleasure to write this essay in honour of John McCombie who retired in 2017 from his Chair in the Department of Land Economy in Cambridge University, which he first joined in 1990 when Gordon Cameron was the Director. I have known John since 1980, first by correspondence and then in person. Over the years, we have conducted a lot of research and writing together, with John invariably being the major author. Our first contact was when John questioned my 1979 paper 'The Balance of Payments Constraint as an Explanation of International Growth Rate Differences' (Thirlwall 1979). He tried to argue that the simple rule, $y = x/\pi$, where y is the growth of output (GDP), x is the growth of real exports and π is the income elasticity of demand for imports, is a tautology because if the income elasticity of demand for imports is defined as $\pi = m/y$, where m is the growth of imports, the simple rule amounts to saying $x = m$. What he had failed to realise was

A. P. Thirlwall (✉)
School of Economics, Keynes College, University of Kent, Canterbury, UK

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_2

that the π I used was estimated econometrically, controlling for changes in the relative prices of foreign and domestic goods, so what the rule really says is that it is not relative price changes that equilibrate the balance of payments of countries, working through the price elasticity of imports (and exports), but output growth. John generously conceded the point (McCombie 1980b), and so began a long and fruitful friendship and collaboration that culminated in our book *Economic Growth and the Balance of Payments Constraint* (McCombie and Thirlwall 1994). Later we collected together a series of empirical studies on balance of payments-constrained growth entitled *Essays on Balance of Payments Constrained Growth* (McCombie and Thirlwall 2004).

Over the years, John and I have met in several places. He spent a sabbatical term with me at the University of Kent in 1984, and we overlapped at the University of Melbourne when John was a lecturer there between 1985 and 1988, and I was a visitor in 1988. We have also participated in several memorable conferences together in different parts of the world including the Post Keynesian conferences in Knoxville, Tennessee, organised by Paul Davidson; a fiftieth anniversary conference celebrating Verdoorn's 1949 paper 'Fattori che Regolano lo Sviluppo della Produttivita del Lavoro' (Verdoorn 1949) held at the University of Genoa in 1999; Keynesian conferences in Pula, Croatia, organised by Soumitra Sharma; and a conference in 2011 on balance of payments-constrained growth held in Coimbra, Portugal, organised by Elias Soukiazis and Pedro Cerqueira, out of which was published a book *Models of Balance of Payments Constrained Growth: History, Theory and Empirical Evidence* (Soukiazis and Cerqueira 2012).

I also have an anecdote to tell. I was the first person to take John to the continent of Europe when he was already in his 40s. When I asked him why he had not visited before, he replied 'because they don't speak English'! John is quintessentially English, brought up as an only child, educated at Dulwich College and Cambridge University where he read geography and had a grandfather who was Postmaster General, and reserved in character—but very clever and a little iconoclastic. 'Irony' and 'putative' are two of his favourite words.

John's PhD at Cambridge, supervised by Robert Rowthorn, and examined by Roger Tarling and Keiran Kennedy in 1982, was entitled 'Post-War Output and Productivity Growth in the Advanced Countries', and this was the starting point for his interest in Kaldor's growth laws, and particularly Verdoorn's Law relating the rate of growth of labour productivity in manufacturing to the growth of output in manufacturing via static and dynamic increasing returns. He continued to mix his interest in Kaldor's growth laws and balance of payments-constrained growth models well into the new millennium. Then his research interest started to focus on a critique of the neoclassical production function and its application for understanding the sources of growth, which culminated in his book with Jesus Felipe, *The Aggregate Production Function and the Measurement of Technical Change: 'Not Even Wrong'* (Felipe and McCombie 2013).

This essay will be organised under three main heads. The first will be on Kaldor's growth laws, and Verdoorn's Law, and John's contribution to our understanding of them. The second will be balance of payments-constrained growth, and John's innovative contributions to the literature. Kaldor's growth laws refer to a closed economy, while balance of payments-constrained growth models deal with an open economy. Kaldor also gave great importance to the role of exports in economic growth (Kaldor 1970), which is missing from his writing on manufacturing industry as the engine of growth. I shall end the essay, therefore, by marrying together Kaldor's first law of growth that manufacturing is the engine of growth in the closed economy with his export-led growth model for the open economy, and show that the former can be regarded as a reduced form of the latter. I shall give some empirical results which show this across a sample of 89 developing countries.

2 Kaldor's Growth Laws

In his Cambridge Inaugural Lecture in 1966 (Kaldor 1966) and in his Frank Pierce Memorial Lectures at Cornell University in the same year (Kaldor 1967), Kaldor enunciated a series of growth laws, and subsidiary propositions, which he believed explained differences in the growth performance of countries at different stages of development. John spent much of the early part of his career in the 1980s critically examining and testing Kaldor's growth laws. The basic thrust of the Kaldorian vision consists of the following propositions: (i) manufacturing industry is the engine of growth—sometimes referred to as Kaldor's first law. The faster the rate of growth of the manufacturing sector, the faster will be the rate of growth of GDP, not simply in a definitional sense in that manufacturing is a component of GDP, but for fundamental economic reasons connected with induced productivity growth inside and outside the manufacturing sector. This is essentially a *structural* explanation of why growth rates differ between countries, as opposed to the one-good model of orthodox neoclassical growth theory in which structure (and demand) plays no part; (ii) productivity growth in the manufacturing sector is induced by the growth of manufacturing output because of static and dynamic returns to scale, otherwise known as Verdoorn's Law—or Kaldor's second law. Static returns relate to economies of scale, while dynamic returns relate to induced capital accumulation and embodied technical progress, plus learning by doing. There is also the phenomenon to consider of macro-economies of scale in the Allyn Young (1928) sense arising from the interaction between manufacturing industries in the presence of micro-economies of scale within industries and a price elasticity of demand for products greater than unity which sets up a cumulative interactive process leading to fast output and productivity growth. We will consider later John's attempt to understand what lies behind Verdoorn's Law; (iii) productivity growth outside manufacturing is induced by manufacturing output growth because the faster manufacturing grows, the faster the rate of transference of labour from other sectors of the economy where there are diminishing returns or no relationship exists between employment growth and output growth—sometimes called Kaldor's third law. A reduction in the amount of labour in these sectors will raise the average product of labour and therefore will raise productivity growth in those sectors. As the scope for absorbing labour from diminishing returns activities dries up, or as output comes to depend on employment in all sectors of the economy, the degree of overall productivity growth induced by manufacturing output growth is likely to diminish, with the overall growth of GDP correspondingly reduced. It is in this sense that Kaldor believed that countries at a high level of development, with little or no surplus labour in agriculture or other non-manufacturing activities, suffer from a shortage of labour and will experience deceleration of growth, not in the sense that the manufacturing industry is constrained by a shortage of labour which he suggested in his Inaugural Lecture as the UK's problem which he soon retracted (Kaldor 1968). This is an important point because it makes a difference to the choice of independent variables to use in testing the Verdoorn relationship between productivity growth in industry and output growth, whether output growth should be the independent variable or employment growth as argued by Cripps and Tarling (1973) and Rowthorn (1975) (see later); (iv) the growth of manufacturing output is not constrained by labour supply but is fundamentally determined by demand from agriculture in the early stages of development and export growth in the later stages. These are the two fundamental sources of autonomous demand to offset leakages from the industrial sector in the form of payments for food from agriculture and imported inputs from other countries; (v) a fast rate of growth of exports and output will set up a virtuous circle of growth through the link between output growth and productivity growth. Fast export growth leads to fast output growth; fast output growth leads to fast productivity growth; fast productivity growth makes exports more competitive; and greater competitiveness leads to fast export growth. The virtuous circle is complete. The export-led growth model originally came from Kaldor's address to the Scottish Economics Society on 'The Case for Regional Policies' (Kaldor 1970) and was formalised by Dixon and Thirlwall (1975) (see also Thirlwall 2014).

Kaldor (1966, 1967) tested the first three propositions for a cross-section of 12 advanced economies over the period 1954–1964 and found the 'laws' were supported. A strong relation existed between manufacturing output growth and GDP growth, but not between the growth of other sectors and GDP growth, and there is a strong inverse relation between the growth of employment outside manufacturing and overall productivity growth.

John's first paper in this field (McCombie 1980a) attempts to quantify the extent to which the reallocation of labour between sectors of an economy explains overall labour productivity growth. Twelve advanced countries are taken over the two periods 1950–1965 and 1965–1973, with the overall level of productivity disaggregated between industry, agriculture and the rest of the economy, and using different assumptions about increasing returns and surplus labour. It transpires that sectoral differences in the *levels* of productivity *by themselves* explain only a small proportion of the growth of overall productivity, but once the transfer of labour is combined with increasing returns in industry, over 30 per cent of total productivity growth can be explained in at least five of the 12 countries. John reveals here for the first time (at least in print) his Keynesian credentials because he concludes the article by saying that Kaldor is correct in his emphasis on the importance of the transfer of labour from agriculture *but* 'this is by no means an indispensable element in his explanation of why growth rates differ'. Since productivity growth is the difference between output growth and employment growth 'the key to the understanding of differences in productivity growth lies in explaining large differences between countries in the *growth of demand for output* (emphasis added). This stands in marked contrast to the neoclassical approach with its emphasis on the supply side'.

Kaldor regarded his third law of the relationship between industrial output growth and productivity growth outside of manufacturing as important for two basic reasons because first of all it provides an explanation of differences in the growth of *total* productivity in an economy and second, by confirming the existence of surplus labour outside industry, it justifies using output not employment as the independent variable (or regressor) in the testing of Verdoorn's Law (see later). The normal test of Kaldor's third law is to run a regression across countries of the form:

$$P_{\mathrm{T}} = a + b\left(g_{\mathrm{I}}\right) - c\left(e_{\mathrm{NI}}\right) \tag{2.1}$$

where P_{T} is the growth of total productivity, g_{I} is the growth of industry output and e_{NI} is the growth of employment outside of industry. The coefficient on e_{NI} is supposed to provide an estimate of the negative effect of non-industrial growth on total productivity growth, but John argues (McCombie 1981) that because P_{T} is definitionally related to g_{I} and e_{NI}, the estimated coefficients simply reflect the share of industrial output in total output and the share of non-industrial employment in total employment and therefore cannot be given any behavioural interpretation. To see this, total productivity growth can be disaggregated as follows:

$$P_{\mathrm{T}} = a\left[g_{\mathrm{I}}\right] - b\left[e_{\mathrm{I}}\right] + \left(1-a\right)\left[g_{\mathrm{NI}}\right] - \left(1-b\right)\left[e_{\mathrm{NI}}\right] \tag{2.2}$$

where a is the share of industry output in total output and b is the share of industry employment in total employment. Comparing Eqs. (2.1) and (2.2), it can be seen that the coefficient on g_{I} in Eq. (2.1) is picking up the share of industry output in total output and the coefficient on e_{NI} is picking up the share of non-industrial employment in total employment. By excluding g_{NI} and e_{I}, the estimates will be biased. Studies that have estimated Eq. (2.1), such as Hansen and Zhang (1996) across 28 provinces of China, and Wells and Thirlwall (2003) across 45 countries of Africa, need to be treated, therefore, with some caution unless it can be shown that the omitted variables are orthogonal to the regressors.

John's first attempt to test Kaldor's first law (McCombie and de Ridder 1983) takes state data for the USA where it is hard to argue that state growth is supply determined because capital and labour are freely mobile across states. Forty-nine states are taken and a subset of 20 largest states. State GDP growth is taken as the regressand and also non-manufacturing output growth to avoid any spurious correlation between manufacturing output growth and total output growth. The results are very similar to the cross-country results originally found by Kaldor. When GDP growth is regressed on manufacturing growth for the 49 and 20 states, the coefficients are 0.632 and 0.622, respectively. When non-manufacturing growth is regressed on manufacturing growth, the coefficients are 0.444 and 0.466, respectively. Manufacturing industry as the engine of growth is supported.

2.1 Verdoorn's Law

Verdoorn's Law, or Kaldor's second law, derives from P.J. Verdoorn's paper 'Fattori che Regalano lo Sviluppo della Produttivita del Lavoro' published in Italian in 1949 in the obscure Italian journal *L'Industria,* where

Verdoorn examines the relationship between labour productivity growth in industry and manufacturing output growth across a variety of countries and industries, and finds a regression coefficient of approximately 0.5. Verdoorn was one of Kaldor's staff in the Research and Planning Division of the Economic Commission for Europe in Geneva where Kaldor was Director between 1947 and 1949. Kaldor never used or quoted Verdoorn's work until his 1966 Inaugural Lecture, but somehow it resurfaced in Kaldor's mind when it became convenient to do so in explaining the UK's poor economic growth record compared to other European countries.[2] Kaldor gives two specifications for testing the Verdoorn relation. One is:

$$p_{\mathrm{m}} = a + b\left(g_{\mathrm{m}}\right) \tag{2.3}$$

where p_{m} is manufacturing productivity growth and g_{m} is manufacturing output growth. The second is:

$$e_{\mathrm{m}} = -a + \left(1-b\right)g_{\mathrm{m}} \tag{2.4}$$

where e_{m} is employment growth in manufacturing. The two equations are two ways of looking at the same relationship because $g_{\mathrm{m}} = p_{\mathrm{m}} + e_{\mathrm{m}}$. Kaldor estimated both ways, deriving a Verdoorn coefficient (b) of 0.484 and an R^2 greater than 0.8.

In fact, from Eqs. (2.3) and (2.4), there are two other ways of specifying the Verdoorn relation. One is:

$$g_{\mathrm{m}} = \frac{a}{1-b} + \frac{1}{1-b}e_{\mathrm{m}} \tag{2.5}$$

The other is:

$$p_{\mathrm{m}} = \frac{a}{1-b} + \frac{b}{1-b}e_{\mathrm{m}} \tag{2.6}$$

Only if the equations are exact will the estimates be the same. From an economic and econometric point of view, the specification is not a matter of indifference. Cripps and Tarling (1973) estimate the Verdoorn relationship with employment growth as the independent variable because Kaldor had argued in 1966 that in the UK at least employment was the constraint on manufacturing output growth, even though Kaldor (1968) had retracted his view about the UK economy in reply to some niggling points of criticism made by Wolfe (1968). Cripps and Tarling show that their version of the Verdoorn Law held from 1951 to 1965, but seemed to break down in the period 1965–1970. Rowthorn (1975), with no reference to Kaldor's reply to Wolfe, also continued to interpret Kaldor as believing that manufacturing output growth is endogenous and employment growth is exogenous and used the same formulation as Cripps and Tarling. Rowthorn claimed to show that Kaldor's results, as well as those of Cripps and Tarling, are heavily dependent on the inclusion of Japan in the sample which, because of its deviant position on the scatter diagram, must be regarded as a special case. Rowthorn criticises Kaldor for estimating a Verdoorn coefficient 'indirectly' (using Eq. 2.4) rather than what he considers 'directly' (using Eq. 2.6). He argues that had Kaldor done so, his estimate of the Verdoorn coefficient would have been much lower than 0.48. But if output growth is exogenous and employment growth is endogenous, the Cripps-Tarling-Rowthorn specification of the Verdoorn relation is incorrect for well-known econometric reasons. Moreover, Kaldor's original results using the correct specification of the Verdoorn relation do *not* depend on the existence of Japan in the sample. The R^2 between p_m and g_m excluding Japan is 0.536 and between e_m and g_m is 0.685.

2.2 Measuring Increasing Returns

A Verdoorn coefficient less than unity ($b < 1$) implies increasing returns, but to measure the degree of increasing returns, the role of capital accumulation in the determination of productivity growth needs to be recognised. The Verdoorn relation, including the contribution of capital, is:

$$p_{\mathrm{m}} = a + b\left(g_{\mathrm{m}}\right) + \varphi\left(k\right) \tag{2.7}$$

where k is the rate of growth of capital. Kaldor was aware of this issue, and in Kaldor (1978b), he introduces the gross investment/output ratio in the Verdoorn equation, but the equation was never tested omitting Japan. When John does this (McCombie 1983), substantial economies of scale are found. If the underlying relationship is a Cobb-Douglas production function, then:

$$g = a_1 + \alpha\left(l\right) + \beta\left(k\right) \tag{2.8}$$

where l is the growth of the labour force. Now $p = g - l$, so:

$$p = g - l = a_1 + \alpha\left(l\right) + \beta\left(k\right) - l \tag{2.9}$$

Therefore:

$$p = a_1 + l\left(\alpha - 1\right) + \beta\left(k\right) \tag{2.10}$$

but $l = g - p$. Therefore:

$$p = a_1 + \left(g - p\right)\left(\alpha - 1\right) + \beta\left(k\right) \tag{2.11}$$

Therefore:

$$p\left(1 + \alpha - 1\right) = a_1 + \left(\alpha - 1\right)g + \beta\left(k\right) \tag{2.12}$$

Therefore:

$$p = \frac{a_1}{\alpha} + \left(\frac{\alpha - 1}{\alpha}\right)g + \frac{\beta}{\alpha}\left(k\right) \tag{2.13}$$

and the Verdoorn coefficient is: $(\alpha - 1)/\alpha$. Now let us suppose that the capital-output ratio is constant, so $g = k$. Therefore, from Eq. (2.13):

$$P = \frac{a_1}{\alpha} + \left[\left(\frac{\alpha - 1}{\alpha}\right) + \frac{\beta}{\alpha}\right]g = \frac{\alpha_1}{\alpha}\left(\frac{\alpha + \beta - 1}{\alpha}\right)g \tag{2.14}$$

If the Verdoorn coefficient is 0.5, then $(\alpha + \beta - 1)/\alpha = 0.5$. If $\alpha = \beta$, then $(2\alpha - 1)/\alpha = 0.5$. Therefore, $\alpha = \beta = 0.66$, and the returns to scale are 1.32.

Kaldor is clear that the Verdoorn Law is a dynamic relationship reflecting static and dynamic returns to scale. To quote him directly 'it is a dynamic rather than a static relationship—between the rates of change of productivity and of output rather than between the *level* of productivity and the *scale* of output—primarily because technical progress enters into it, and is not just a reflection of economies of large scale production' (Kaldor 1966 p. 10). But John argues in several papers (e.g. McCombie 1981, 1982, 1984) that the Verdoorn Law may also be derived from:

$$E_t = A(\exp at)Q_t^{\,b} \tag{2.15}$$

where E and Q are the *levels* of employment and output. Interestingly, Verdoorn (1949) himself derived the law from a static Cobb-Douglas production function, but that does not necessarily imply that integration of the growth equation will lead to the level equation. This will depend on the assumption made about the constant of integration. But this leads to what John has called the static/dynamic paradox because invariably when the law is tested using *levels* of productivity and output, the coefficient does not differ significantly from unity (constant returns), while when the law is tested using growth rates, increasing returns are found.[3]

To give some examples of John's findings in this field: McCombie (1982) takes a sample of OECD countries over the time period 1950–1973 and finds that taking *levels* of productivity and output, it is not possible to reject the hypothesis of constant returns. McCombie and de Ridder (1983) use US state data from 1963 to 1973 and estimate a significant Verdoorn coefficient using growth rates of productivity and output, giving returns to scale of 1.45. But using level data, there is no evidence of increasing returns. McCombie (1986) looks at the manufacturing sector of nine OECD countries over the period 1955–1979, including capital accumulation in the dynamic Verdoorn equation, and estimates a Verdoorn coefficient of 0.35 (increasing returns). McCombie and Fingleton (1998) use 178 regions across 13 EU countries over the period 1978–1989 and estimate a Verdoorn coefficient of 0.575 (allowing for country dummies). When the static law is estimated, the Verdoorn coefficient falls to 0.057. A novel feature of this study is that they allow for the fact that some of the productivity growth may be due to catch-up. The log of the initial level of productivity in the base year is therefore included in the equation and turns out to be highly significant. The Verdoorn coefficient falls to 0.275. Angeriz, McCombie and Roberts (2008) take 54 regions of the EU over the period 1986–2002 using a variety of spatial econometric techniques, testing both the static and dynamic versions of Verdoorn's Law. The static version gives constant returns, while the dynamic version gives a Verdoorn coefficient of between 0.50 and 0.67, depending on the method of estimation. Finally, McCombie, Angeriz and Roberts (2009) estimate Verdoorn's Law in a spatial econometric framework for six individual manufacturing industries using EU regional data over the period 1991–2002. In this study total factor productivity growth is taken as the dependent variable, and as in the other studies above, the static/dynamic Verdoorn Law paradox is apparent.

It is not entirely clear what lies behind the paradox. Simultaneous equation bias in the dynamic specification is sometimes mentioned, but it is not clear which way the bias goes (McCombie 1982). There may be some bias in the dynamic estimation due to omitted variables, such as capital, but this is not a problem if capital is included, or if the capital-output ratio is constant. Errors in variables (McCombie 1981, 1982) may be another explanation, but it is not clear why measurement errors should be more or less between data in levels and data in growth rates. Spatial aggregation bias is a possibility when regional data are used. McCombie and Roberts (2007) attempt to show this using a simulation exercise showing that spatial aggregation bias biases the estimates of returns to scale downwards using the static specification in log levels.[4] McCombie, Angeriz and Roberts (2009) are more categoric when they say that 'spatial aggregation bias resolves the issue'. But equally they argue that 'the dynamic formulation is the correct specification'. The argument goes back to John's earlier and original conclusion (McCombie 1982) when he says: 'the argument suggests that the dynamic Verdoorn coefficient may provide the unbiased estimate and the result of constant returns to scale provided by the static law may be due to the misspecification of the function. An implication is that the conventional static production function may understate the degree of returns to scale by their abstraction from the dynamic components that Kaldor argues are so important'. In other words, the paradox lies in the second-order identification problem that differentiating the level equation yields the growth equation, but integrating the growth equation will not necessarily yield the level equation because this depends on the constant of integration (as argued earlier).

In fact, John had come to the view much earlier (McCombie 1986) that Kaldor's interpretation of Verdoorn's Law as reflecting various types of dynamic increasing returns is the most satisfactory and accords very closely with Kaldor's linear technical progress function (Kaldor 1961) where the rate of growth of output per man is a function of the rate of growth of capital per man and the rate of productivity growth depends on autonomous productivity growth on the one hand and the extent to which technical progress is embodied in capital accumulation on the other. Dixon and I (Dixon and Thirlwall 1975) first showed how the Verdoorn coefficient can be derived from Kaldor's technical progress function:

$$\text{Let } p_m = d + \pi(k) \tag{2.16}$$

where p_m is the growth of output per man, and k is the growth of capital per man.

$$\text{Now let}: d = \alpha_1 + \beta_1(g_m) \tag{2.17}$$

where a_1 is 'pure' disembodied technical progress and β_1 reflects learning by doing.

$$\text{Now let}: k = \alpha_2 + \beta_2 \left(g_m \right) \tag{2.18}$$

where β_2 reflects induced capital accumulation (the accelerator principle).

Substituting (2.17) and (2.18) into (2.16) gives:

$$p_m = \left(\alpha_1 + \pi\alpha_2 \right) + \left(\beta_1 + \pi\beta_2 \right) g_m \tag{2.19}$$

So, the Verdoorn coefficient depends on learning by doing (β_1); induced capital accumulation (β_2); and the extent to which technical progress is embodied in capital (π). If this interpretation is accepted, it means that the conventional Cobb-Douglas production function is not the correct underlying structure of the Verdoorn Law, and this may be another reason why the estimates of the returns to scale from the static law are biased downwards.

3 Balance of Payments-Constrained Growth

John has made major contributions to the development of the balance of payments-constrained growth model that I first outlined in Thirlwall (1979). He has shown that the simple result I derived, that a country's long-run growth rate can be approximated by the ratio of the growth of exports (x) to the income elasticity of demand for imports (π), is a reduced form of the Hicks' super multiplier (McCombie 1985b). He defends very well the attack on the model by McGregor and Swales (1985, 1986, 1991) and Palley (2002) and discusses extensively the role of non-price competition in the model reflected in the income elasticities of demand for exports and import (McCombie 1989, 1992). He also devised a simple parametric test of the model for individual countries. He showed with myself

(McCombie and Thirlwall 1997a) that if there is a limit to the current account deficit given by the debt to GDP ratio, capital inflows make little difference to the predicted growth rate from the simple model ($y = x/\pi$). On top of this, he has conducted several case studies of the model by himself and with colleagues (McCombie 1997; McCombie and Britto 2009; McCombie et al. 2010; McCombie and Tharnpanich 2013).

3.1 John's Initial Attack on the Model

John was initially hostile to the model (McCombie 1980b). He accused me of circular reasoning. He argued that if we follow Thirlwall and use an estimate of the income elasticity of demand for imports (π) by regressing import growth (m) on income growth (y), it is not surprising that the balance of payments equilibrium growth rate (y_B) closely approximates to the actual growth rate (y) because the analysis borders on circular reasoning.[5] There is a problem in determining the direction of causality; whether growth is demand constrained or supply constrained. My response (Thirlwall 1981) was to say in the context of the UK that if growth was constrained *before* the balance of payments constraint became important, why didn't the UK experience growing balance of payments surpluses like Japan? I continued 'while the simple model itself may not be able to discriminate easily between the demand and supply-led growth hypotheses, I think the *results* of applying the model, combined with judgment, can'.

John wrote to me on 7 November 1980 saying that his balance of payments paper was 'written to a certain extent in the spirit of Devil's Advocate'. He went on: 'while from the point of view of the demand – oriented explanation of growth, I would have been worried if the law did not hold, and it is a remarkable empirical generalisation, I am not so convinced that the law necessarily confirms the hypothesis of export-led growth'. I replied to him on 13 November 1980 that 'I am still a little puzzled why you cannot bring yourself to believe that if the rate of growth of exports were higher, the rate of growth of output could also be higher and that the rate of growth of output is not constrained by a shortage of factor supplies'. John didn't agree with this. He wrote (19 November) 'the reason why the UK ran a deficit is that the government pursued policies trying to increase the growth rate above that permitted by the growth of factor supplies. This, in turn, induced a greater growth of imports and hence a deficit'. Clearly, at that time, John did not believe in the endogeneity of factor supplies! But then, finally, in the same letter, he seems to concede: 'certainly, at the moment, I find the demand-oriented approach more plausible than the supply-constrained assumption, and the model, together with your formulation of the Verdoorn growth model (Thirlwall 1980), is very attractive'. So began a long and fruitful collaboration which still continues, but culminated in our book *Economic Growth and the Balance of Payments Constraint* published in 1994.

3.2 Balance of Payments-Constrained Growth and the Hicks' Super Multiplier

The simple rule $y_B = x/\pi$ turns out to be the dynamic version of the static Harrod foreign trade multiplier of $Y = X/m$, where Y is the *level* of income, X is the *level* of exports and m is the marginal propensity to import (Harrod 1933; Thirlwall 1982). The two 'multipliers' are derived on the same assumptions of no change in the real terms of trade and the existence (necessity) of long-run balance of payments equilibrium.

John (McCombie 1985b) showed that the dynamic Harrod trade multiplier, $y_B = x/\pi$, can be thought of as reflecting a reduced form of the Hicks' super multiplier where all components of demand adapt to the exogenous rate of growth of exports which provides the foreign exchange to pay for the import content of consumption, investment, government expenditure and exports themselves. John shows that the rule $y_B = x/\pi$ can be decomposed into two parts according to the formula:

$$y_B = \frac{1}{k}\left(w_x x + w_a a_B\right) = \frac{x}{\pi} \tag{2.20}$$

where k is the Keynesian multiplier for an open economy; w_x is the share of exports in GDP; a_B is the growth of other components of autonomous expenditure necessary, for a given growth of export, to maintain the growth of income at the balance of payments equilibrium rate; and w_a is the share of autonomous expenditure (excluding exports) in total income. The balance of payments equilibrium growth rate is thus determined jointly by the growth of exports, via the multiplier (w_x/k), and the growth of 'induced' autonomous expenditure working through the associated domestic multiplier, w_a/k. This is identical to the effect of the growth of exports working through the super multiplier, $1/\pi$. The importance of this result is that it shows clearly that not only does export growth have a direct effect on output growth but also an indirect effect by allowing other components of demand to grow faster because export growth pays for the import content of consumption, investment and so on. Exports are a unique component of demand in this respect. Kaldor (1975) was responsible for reviving the doctrine of the Harrod trade multiplier and already in 1970 had presented an export-led growth model applicable to regions and countries alike with cumulative features, but lacking a balance of payments constraint (see Thirlwall 2014).

3.3 Defence of the Model Against McGregor and Swales and Palley

In a series of papers, McGregor and Swales (1985, 1986,1991) attack the balance of payments-constrained growth model as 'incoherent' and lacking empirical support. They make three basic criticisms of the model. Firstly, that if relative prices remain constant because of the 'law of one price', the model is not distinguishable from a neoclassical model in which a country can sell any amount of its goods at a given price, so that exports and output growth are supply constrained not demand constrained. Secondly, the model doesn't capture satisfactorily non-price competition. Thirdly, there is no relation empirically across countries between actual growth (y) and the estimates of the balance of payments equilibrium growth rate (y_B).

John (McCombie 1989, 1992) had no difficulty in refuting each of these criticisms. It is true that if the 'law of one price' holds, there can be no balance of payments constraint because exports would adjust to imports with no need for domestic income adjustment. If true, however, it would mean that the price elasticity of demand for exports is infinitely elastic and that the income elasticity of demand for exports would be insignificant—neither of which are the case empirically. The world income variable is always highly significant in export growth equations which is not consistent with the small open economy assumption of the 'law of one price'. And it needs to be borne in mind, of course, that the 'law of one price' is not the only explanation of why relative prices measured in a common currency, or the real exchange rate, may be 'sticky'. More likely explanations are relative price changes mirroring nominal exchange rate changes, and oligopolistic market structures at least in the production of industrial goods.

On the question of non-price competition, McGregor and Swales are simply wrong. Non-price competition is captured by the income elasticities of demand for exports and imports. McGregor and Swales refute this because they argue that income elasticities will determine the growth of exports and imports but not changes in the share of markets which the country's exports (and imports) take. John points out that there is plenty of empirical evidence for non-price competition, particularly for changes in export shares which cannot be explained by relative price movements—the so-called Kaldor Paradox (Kaldor 1978a).

Not much is known about the determinants of the income elasticities of demand for exports and imports (reflecting non-price competitiveness). Some recent work explores the connection between aggregate income elasticities and the sectoral composition of trade. Gouvea and Lima (2010) and Romero et al. (2011) have estimated export and import demand functions for different technological sectors and find that high-tech sectors have higher income elasticities. Gouvea and Lima (2013) find that capital goods have higher income elasticities than consumption and intermediate goods. McCombie and Romero (2016a) take five technological sectors in 14 developed countries and find higher income elasticities for medium- and high-tech manufactures. McCombie and Tharnpanich (2013) find in Thailand that manufactures have a higher income elasticity of demand than primary commodities.

McCombie and Romero (2016b) modify export and import demand functions by introducing the direct effect of productivity growth on export and import growth via improvements in non-price competitiveness, and Ribeiro, McCombie and Lima (2016) endogenise the income elasticities of demand for exports and imports via changes in the technological gap and income distribution. For example, if a poor country can narrow its technological gap with a rich country, it will raise the ratio of its income elasticity of demand for exports to imports. More unequal countries will tend to import more luxury goods and export necessities, reducing the ratio. This is a research field still ripe for enquiry.

As far as the predictive power of the model is concerned, McGregor and Swales purport to show that y_B is not a good predictor of y across countries. They test by using the linear regression $y = a + b(y_B)$ with the null hypothesis that $a \neq 0$ and $b \neq 1$. Using Thirlwall's (1979) sample of countries, they do not reject the null hypothesis. John criticises the test on two grounds. Firstly, the estimates of y_B depend on the estimated coefficient π which has a standard error. Inverse least squares should therefore be used. Secondly, the cross-section test has outliers which misleadingly rejects the rule that y_B can predict y for individual countries. John shows (McCombie 1992) that if Japan and the USA are excluded from the sample, McGregor and Swales are wrong.

John develops a much more suitable parametric test for individual countries—now called the McCombie test. First calculate the income elasticity of demand for imports π^* that equates the ratio of the rate of growth of exports to the actual growth of output y, that is, $\pi^* = x/y$, and then compare π^* with the statistical estimate of π (π^) from an import growth equation including as a regressor the rate of change of relative prices. If there is no significant difference between π^* and π^, then y_B will be a good predictor of y.

Palley (2002) also attacks the balance of payments-constrained growth model on the grounds that there is no mechanism in the model for reconciling the growth of supply and demand. He argues that if y_B is less than potential output growth (y_N), the income elasticity of demand for imports will fall to equate y_B and y_N, so, in effect, no country is balance of payments constrained in the long run: 'the steady state growth rate [is] uniquely determined by supply-side factors' (Palley 2002, p. 15). There are a number of problems with this argument as John points out (McCombie 2011). Palley claims that $y_B < y_N$ is not observed in practice so there must be some adjustment mechanism, but the adjustment could equally be on the supply side, as Setterfield (2006) has argued. Weak demand growth through a balance of payments constraint can affect adversely both determinants of the rate of growth of productive potential, that is, the rate of growth of the labour force and the growth of labour productivity (by reducing the Verdoorn coefficient). Moreover, if $y_B < y_N$ and governments expand demand to get to y_N, the income elasticity of demand for imports could rise rather than fall which would worsen the situation. Lanzafame (2014) has shown for a panel of 22 OECD countries over the period 1960–2010 that the direction of causation runs from the balance of payments-constrained growth rate (y_B) to the actual growth rate (y) to the potential growth rate (y_N). As Setterfield says 'the demand-side thus rules the roost in what can be identified as a model of *fully demand-determined growth*' (p. 55).

3.4 Capital Flows

Thirlwall and Hussain (1982) were the first to include capital flows into the balance of payments-constrained growth model, which potentially relaxes a balance of payments constraint on growth, but no limit was imposed on the current account or debt to GDP ratio that capital inflows might be associated with.

In 1996, John and I were asked separately by Philip Arestis whether we would write an essay in honour of Geoffrey Harcourt. We decided to join forces and to address the question, which hadn't been asked before, of what difference do capital flows make to the sustainable growth rate, assuming there is a limit to the current account or debt to GDP ratio. It was mainly John who worked on the model and came up with the interesting, but not obvious, conclusion that even if the current account deficit as a proportion of GDP is allowed to be as high as 10 per cent, it makes a relatively small quantitative difference to the growth rate determined by the basic dynamic Harrod trade multiplier result of $y_B = x/\pi$ (McCombie and Thirlwall 1997a).[6] Moreno-Brid (1998, 2003) subsequently derived the same result as us in a simpler (more elegant) way. We both include interest payments on past debt in the full model, but first, for clarity, let us model without interest payments. The fundamental balance of payments identity is:

$$P_{\mathrm{d}}X + FP_{\mathrm{d}} = P_{\mathrm{f}}ME \tag{2.21}$$

where X is the volume of exports, P_{d} is the domestic price of exports, M is imports, P_{f} is the foreign price of imports, E is the exchange rate to convert the value of imports in foreign currency into domestic currency, F is the current account deficit in *real* terms and FP_{d} is nominal capital inflows (C) in domestic currency to finance the deficit. Taking logs of Eq. (2.21) and differentiating with respect to time gives:

$$\theta\left(p_{\mathrm{d}} + x\right) + \left(1 - \theta\right)\left(f + p_{\mathrm{d}}\right) = m + p_{\mathrm{f}} + e \tag{2.22}$$

where θ is the proportion of imports financed by exports and $(1 - \theta)$ is the proportion of imports financed by capital flows. Now the growth of exports can be written as:

$$x = \eta\left(p_{\mathrm{d}} - p_{\mathrm{f}} - e\right) + \varepsilon\left(z\right) \tag{2.23}$$

and the growth of imports as:

$$m = \psi\left(p_{\mathrm{d}} - p_{\mathrm{f}} + e\right) + \pi\left(y\right) \tag{2.24}$$

where $(p_{\mathrm{d}} - p_{\mathrm{f}})$ is the difference in the rate of change of domestic and foreign prices; e is the rate of change of the exchange rate and y and z are the growth of domestic and foreign income, respectively; $\eta(<0)$ and $\psi(>0)$ are the price elasticities of exports and imports, respectively; and π and ε are the income elasticities of imports and exports. Substituting Eqs. (2.23) and (2.24) into (2.22) and setting $f = y$, so that the ratio of the current account deficit to GDP is constant, gives:

$$y_{\mathrm{D}} = \frac{\theta\varepsilon z + \left(\theta\eta + \psi + 1\right)\left(p_{\mathrm{d}} - p_{\mathrm{f}} - e\right)}{\pi - \left(1 - \theta\right)} \tag{2.25}$$

If the real terms of trade remain unchanged, the constrained growth rate consistent with a fixed deficit/GDP ratio is:

$$y_{\mathrm{D}} = \frac{\theta x}{\pi - (1-\theta)} \tag{2.26}$$

With no deficit, $\theta = 1$, and the simple rule holds—$y_{\mathrm{D}} = x/\pi$. Now suppose that the deficit to GDP ratio is allowed to be as high as 10 per cent of GDP, so $\theta = 0.9$, and $x = 10$ per cent, and $\pi = 2$, the simple rule gives a balance of payments equilibrium growth rate of 5 per cent, and the modified model gives a prediction of 4.73 per cent—hardly any difference.

If the current account deficits are financed by debt-creating flows, the model needs further modification for interest rate payments. McCombie and Thirlwall (1997a) included this and so too have Elliot and Rhodd (1999), Ferreira and Canuto (2003), Vera (2006) and Alleyne and Francis (2008). Following Moreno-Brid (2003), we can modify Eq. (2.22) by taking interest payments out of capital flows to get:

$$\theta(p_{\mathrm{d}} + x) - \theta_1(p_{\mathrm{d}} + i) + (1 - \theta - \theta_1)(p_{\mathrm{d}} + f) = m + p_{\mathrm{f}} + e \tag{2.27}$$

where i is the rate of growth of real net interest payments abroad (the negative sign implies the country is a net debtor) and θ_1 is the share of foreign exchange devoted to interest payments. Again, setting $f = y$, and substituting for x and m, gives:

$$Y_{\mathrm{I}} = \frac{\theta \varepsilon z - \theta_1 i + (\theta\eta + \psi + 1)(p_{\mathrm{d}} - p_{\mathrm{f}} - e)}{\pi - (1 - \theta - \theta_1)} \tag{2.28}$$

And if the real terms of trade are constant:

$$Y_{\mathrm{I}*} = \frac{\theta x - \theta_1 \mathrm{i}}{\pi - (1 - \theta - \theta_1)} \tag{2.29}$$

If there are no interest payments on debt, Eq. (2.29) reduces to Eq. (2.26). But now interest rate payments have the potential to reduce the sustainable growth rate depending on the growth of interest payments and the share of foreign exchange (θ_1) devoted to interest payments. For example, if $i = 4$ per cent per annum and $\theta_1 = 0.2$, the sustainable growth rate will be 4.09 per cent compared with 5 per cent from the simple model.

3.5 Case Studies of Balance of Payments-Constrained Growth

John has used his considerable applied econometric skills to test the balance of payments-constrained growth model for several different countries. His first study (McCombie 1997) was for the USA, Japan and the UK. This was followed by detailed case studies for Brazil (McCombie and Britto 2009), Pakistan (McCombie et al. 2010) and Thailand (McCombie and Tharnpanich 2013).

Crucial to the estimation of the model is a well-determined estimate of the income elasticity of demand for imports. This requires the absence of unit roots in the data and making allowances for any structural breaks. In the case of the study for the USA, Japan and the UK, both log levels of data are used, and first differences of the logs and the import elasticity results are roughly similar. The technique of rolling regressions is also used covering 15-year sub-periods. For the USA, the model predicts well for 1970–1984, but for much of the 1980s, the USA was growing faster than its balance of payments equilibrium growth rate—and then reverted. So over the long period 1974–1993, the growth of the US economy did not differ significantly from its balance of payments equilibrium growth rate. For the UK, the model fits very well over the period 1952–1993. For Japan, growth was always below its balance of payments equilibrium growth rate, as Thirlwall (1979) originally found for the 1950s and 1960s, with Japan running huge balance of payments surpluses.

In the study of Brazil, the model is tested for the period 1951–2006, with the import demand function estimated using Vector Auto Regression (VAR). The estimated income elasticity of demand for imports is 1.7, but using the McCombie test, the hypothetical income elasticity to equate the actual growth rate with the balance of payments equilibrium growth rate is 1.15. The basic model, therefore, turns out not to be a good predictor of growth performance. But when the extended model with capital flows and interest payments on debt is used, the hypothetical import elasticity lies between 1.46 and 1.73, so the extended model is a good predictor of actual growth performance. McCombie and Britto conclude that 'Brazil's growth fluctuates around its long term trend determined by the extended version of Thirlwall's Law'.

In the study for Pakistan, the model is estimated over the period 1980–2007 using co-integration techniques. The estimated income elasticity of demand for imports is 0.91 which is very close to the hypothetical elasticity of 0.88 which would make the actual growth rate and balance of payments-constrained growth rate equal. The maximum annual growth rate consistent with balance of payments equilibrium is 5 per cent compared with Pakistan's target rate of 7–8 per cent. Pakistan has frequent balance of payments crises.

In the study of Thailand, the model is estimated over the period 1962–2009, and the results show that the economy grew at, or very near to, the rate constrained by the balance of payments, but there is a marked deceleration of growth post-1999. This seems to have been due to a fall in the income elasticity of demand for exports as a result of structural changes in the economy and a slowdown of manufacturing output growth. This explains the slowdown of growth from over 9 per cent per annum up to 1998 to only 4 per cent from 1999 to 2009.

4 Manufacturing Output-Led Growth Versus Export-Led Growth

Kaldor's growth laws give primacy to the growth of the manufacturing sector, while the balance of payments-constrained growth model, and Kaldor's (1970) model of export-led growth, gives primacy to the growth of exports. It might be said, therefore, that there is an uneasy connection between the closed economy model of growth rate differences between countries based on the structure of production, and an open economy model in which export growth is the driving force. There *is* an uneasy connection, but it is easy to see that manufacturing as the engine of growth is also a reduced form of export-led growth in which GDP growth is a function of export growth, but export growth is a function of manufacturing output growth. In other words:

$$g_{\text{gdp}} = a_1 + b_1(x) \tag{2.30}$$

$$x = a_2 + b_2(g_{\text{m}}) \tag{2.31}$$

and substituting (2.31) into (2.30) gives:

$$g_{\text{gdp}} = (a_1 + b_1 a_2) + (b_1 b_2) g_{\text{m}} \tag{2.32}$$

Kaldor's first law of growth is a reduced form of two structural equations and depends on the elasticity of GDP growth with respect to export growth (b_1), and the elasticity of export growth with respect to manufacturing output growth (b_2). A colleague and I have tested these relationships across a sample of 89 developing countries over the period 1990–2011 (Pacheco-López and Thirlwall 2014). Figure 2.1 shows the relationship between GDP growth and manufacturing output growth (Kaldor's first law).

The estimated equation is (t-values in brackets):

$$g_{\text{gdp}} = \underset{(9.07)}{2.16} + \underset{(9.43)}{0.43}\, g_{\text{m}} \quad r^2 = 0.50$$

Figure 2.2 shows the relation between manufacturing output growth and export growth.

The estimated equation is:

$$x = \underset{(5.7)}{3.59} + \underset{(6.19)}{0.75}\, g_{\text{m}} : \quad r^2 = 0.30$$

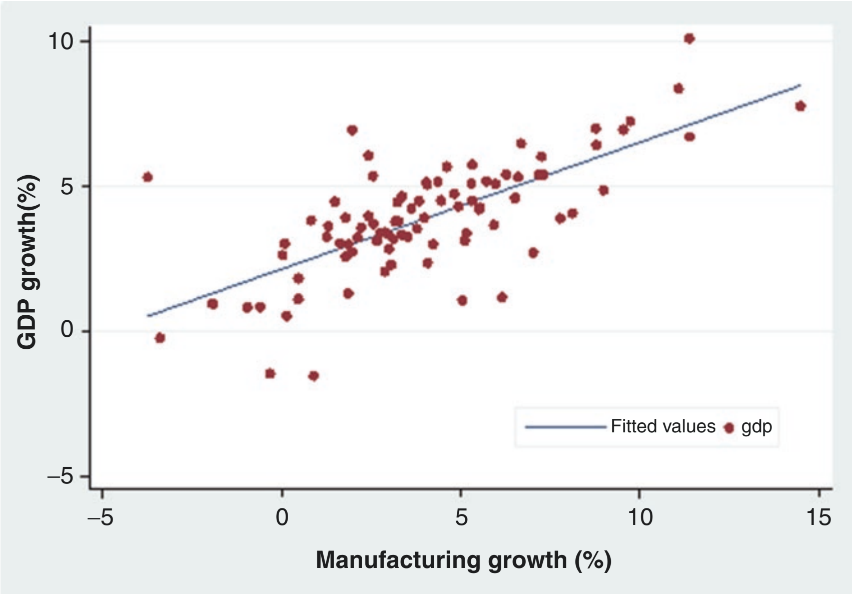


Fig. 2.1 Association between GDP growth and manufacturing growth, 1990–2011

The strong positive relation should occasion no surprise. For any given growth of world income, the growth of exports will depend on the structure of production and the income elasticity of demand for different products. Export growth is endogenous in this sense and is likely to be related to the growth of manufacturing output since all manufactures are potentially tradable. Primary products are also potentially tradable, but they do not have the same production and demand characteristics. Their demand growth in international trade is low (Engel's Law). Some services are tradable, but many are not, and their income elasticity in world markets is not likely to be as high as for medium- and high-technology manufactured goods.

Figure 2.3 shows the link between export growth and GDP growth. The estimated equation by two-stage least squares is:

$$g_{\text{gdp}} = \underset{(0.21)}{0.09} + \underset{(9.43)}{0.57}\ x : r^2 = 0.50$$

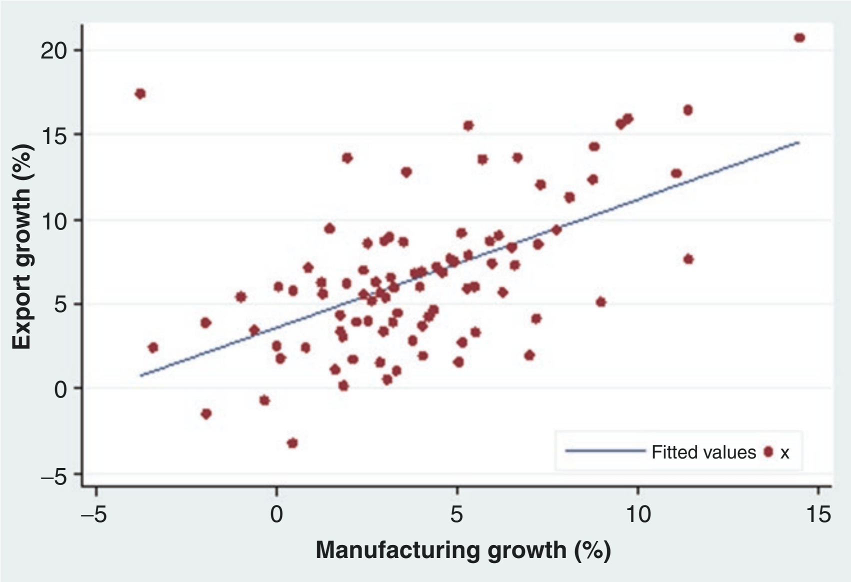


Fig. 2.2 Association between export growth and manufacturing growth, 1990–2011

There are three major reasons for expecting a priori a close link between export growth and GDP growth. Firstly, there is the neoclassical supply-side argument which focuses on the static and dynamic gains from trade and the externalities that the export sector can confer on the non-export sector and the rest of the economy (Feder 1983). Exports also allow the import of inputs and investment goods that may be more productive than domestic resources, thus increasing the supply capacity of the economy. Secondly, if domestic demand is constrained by a shortage of foreign exchange, faster export growth will help relax that constraint. All components of demand have an import content which need to be paid for, and only exports can do so. Exports are a unique component of demand in that respect (McCombie 1985b). Thirdly, export growth may set off a virtuous circle of growth, as outlined earlier (Kaldor 1970).

The results of this research across a wide sample of developing countries support the work of Hausmann, Hwang and Rodrik (2007) on 'What You Export Matters' which shows a close association between what they call EXPY and growth rate differences across countries. EXPY is a weighted

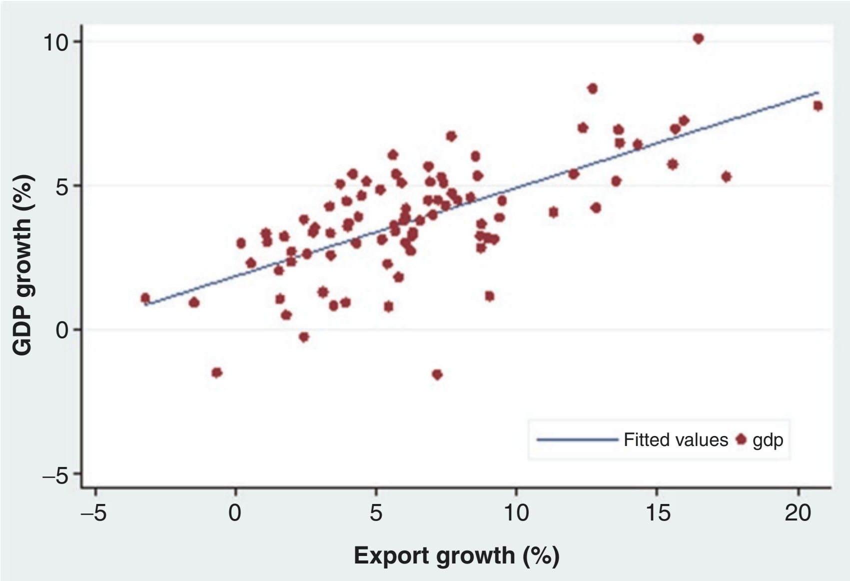


Fig. 2.3 Association between GDP growth and export growth, 1990–2011

average of what they call PRODY which measures the income level that each good produced is associated with. Countries grow fast if they have an export structure geared to the production and income levels of rich countries where the demand for high value-added goods is strong. Countries producing manufactured goods with a high income elasticity of demand in world markets will have a higher growth of exports and a higher growth of GDP. Hausmann et al. show a close correlation across countries between PRODY, EXPY (a weighted average of the PRODYs) and GDP growth. As they remark 'types of goods in which a country specialises have important implications for subsequent economic performance'.

5 Conclusion

John has led the life of a scholar ensconced in Downing College, Cambridge. It has seemed his natural home in a spacious study overlooking the beautiful green of the College quad where he could think, research and write to his heart's content. He did not only live in an ivory tower, however. He took on major advisory jobs, particularly for international development organisations such as the Asian Development Bank where Jesus Felipe was the senior research economist.

His contribution to the understanding of the dynamics of growth in a closed and open economy has been immense. He must surely be the world's leading expert on Verdoorn's Law, confirming that there is something special about the production characteristics of manufacturing industry as opposed to other sectors of the economy. Likewise, he has been the foremost researcher confirming that many countries' growth can be approximated by the simple dynamic Harrod trade multiplier rule—and this cannot be an accident.

I owe a huge debt of gratitude to John for the inspiration he has given me and for all the work we have done together over the years.

Notes

1. The author is grateful to Dr. Penelope Pacheco-Lopez for helpful comments on an early draft of the paper.
2. There are only three references to Verdoorn's 1949 paper between 1949 and 1966: two by Colin Clark (1957, 1962) and one by Kenneth Arrow (1962) (see McCombie et al. 2002).
3. Interestingly, the static/dynamic paradox does not seem to exist using time series data or with panel estimation using two-way fixed effects. The latter is illustrated in Leon-Ledesma (2000) for Spanish regions and, also, Angeriz et al. (2008) across 54 European regions 1986–2002.
4. The authors show that it arises through adding up the output and inputs of so-called Functional Economic Areas within a region to estimate the static law, whereas taking the dynamic specification, the *growth rates* of outputs and inputs are dimensionless.
5. He had forgotten that the income elasticities used from Houthakker and Magee (1969) were estimated controlling for relative price changes in the equation.
6. Allowance for interest rate payments on past debt makes a bigger difference (see later).

References

Angeriz, A., McCombie, J. S. L., & Roberts, M. (2008). New estimates of returns to scale and spatial spillovers for EU regional manufacturing 1986–2002. *International Regional Science Review, 31*(1), 62–87.

Alleyne, D., & Francis, A. A. (2008). Balance of payments constrained growth in developing countries: A theoretical perspective. *Metroeconomica, 59*(2), 189–202.

Arrow, K. (1962). The economic implications of learning by doing. *Review of Economic Studies, 29*, 155–173.

Clark, C. (1957). *The conditions of economic progress* (3rd ed.). London: Macmillan.

Clark, C. (1962). *British trade in the common market*. London: Stevens and Sons Ltd.

Cripps, T. F., & Tarling, R. (1973). *Growth in advanced capitalist economies 1950–70*. Cambridge: Cambridge University Press.

Dixon, R. J., & Thirlwall, A. P. (1975). A model of regional growth rate differences on Kaldorian lines. *Oxford Economic Papers, 27*(2), 201–224.

Elliot, D., & Rhodd, R. (1999). Explaining growth rate differences in the highly indebted countries: An extension to Thirlwall and Hussain. *Applied Economics, 31*(9), 1145–1148.

Feder, G. (1983). On exports and economic growth. *Journal of Development Economics, 12*(1/2), 59–74.

Felipe, J., & McCombie, J. S. L. (2013). *The aggregate production function and the measurement of technical change: 'Not even wrong'*. Chletenham: Edward Elgar.

Ferreira, A., & Canuto, O. (2003). Thirlwall's law and foreign capital in Brazil. *Momento Economico, 125*, 18–29.

Gouvea, R., & Lima, G. (2010). Structural change, balance of payments constraint and economic growth: Evidence from the multi-sectoral Thirlwall's law. *Journal of Post Keynesian Economics, 33*(1), 169–204.

Gouvea, R., & Lima, G. (2013). Balance of payments constrained growth in a multi-sectoral framework: A panel data investigation. *Journal of Economic Studies, 40*(2), 240–254.

Hansen, J., & Zhang, J. (1996). A Kaldorian approach to regional economic growth in China. *Applied Economics, 28*(6), 679–685.

Harrod, R. (1933). *International economics*. London: Macmillan.

Hausmann, R., Hwang, R., & Rodrik, D. (2007). What you export matters. *Journal of Economic Growth, 12*(1), 1–25.

Houthakker, H., & Magee, S. (1969). Income and price elasticities in world trade. *Review of Economics and Statistics, 51*(2), 111–125.

Kaldor, N. (1961). Capital accumulation and economic growth. In F. Lutz (Ed.), *The theory of capital*. London: Macmillan.

Kaldor, N. (1966). *Causes of the slow rate of growth of the United Kingdom*. Cambridge: Cambridge University Press.

Kaldor, N. (1967). *Strategic factors in economic development*. New York: Cornell University Press.

Kaldor, N. (1968). Productivity and growth in manufacturing industry: A reply. *Economica, 35*(140), 385–391.

Kaldor, N. (1970). The case for regional policies. *Scottish Journal of Political Economy, 17*(3), 337–348.

Kaldor, N. (1975). What is wrong with economic theory? *Quarterly Journal of Economics, 89*(3), 347–357.

Kaldor, N. (1978a). The effects of devaluation on trade in manufactures. *Further essays in applied economics*. London: Duckworth.

Kaldor, N. (1978b). *Further essays in economic theory*. London: Duckworth.

Lanzafame, M. (2014). The balance of payments constrained growth rate and the natural rate of growth: New empirical evidence. *Cambridge Journal of Economics, 38*(4), 817–838.

León-Ledesma, M. (2000). Economic growth and Verdoorn's law in the Spanish regions 1962–91. *International Review of Applied Economics, 14*(1), 59–69.

McCombie, J. S. L. (1980a). On the quantitative importance of Kaldor's laws. *Bulletin of Economic Research, 32*(2), 102–112.

McCombie, J. S. L. (1980b). Are international growth rates constrained by the balance of payments? *Banca Nazionale del Lavoro Quaterly Review, 139*, 455–458.

McCombie, J. S. L. (1981). What remains of Kaldor's laws. *Economic Journal, 91*(361), 206–216.

McCombie, J. S. L. (1982). Economic growth, Kaldor's laws and the static-dynamic Verdoorn law paradox. *Applied Economics, 14*(3), 279–294.

McCombie, J. S. L. (1983). Kaldor's laws in retrospect. *Journal of Post Keynesian Economics, 5*(3), 414–429.

McCombie, J. S. L. (1984). The Verdoorn law controversy: Some new empirical evidence using US state data. *Oxford Economic Papers, 36*(2), 268–284.

McCombie, J. S. L. (1985a). Increasing returns and manufacturing industries: Some empirical issues. *Manchester School, 53*(1), 55–75.

McCombie, J. S. L. (1985b). Economic growth, the Harrod trade multiplier and the Hicks super-multiplier. *Applied Economics, 17*(1), 52–72.

McCombie, J. S. L. (1986). On some interpretations of the relationship between productivity and output growth. *Applied Economics, 18*(11), 1215–1225.
McCombie, J. S. L. (1989). Thirlwall's law and balance of payments constrained growth: A comment on the debate. *Applied Economics, 21*(5), 611–629.
McCombie, J. S. L. (1992). Thirlwall's law and balance of payments constrained growth: More on the debate. *Applied Economics, 24*(5), 493–512.
McCombie, J. S. L. (1997). The empirics of balance of payments constrained growth. *Journal of Post Keynesian Economics, 19*(3), 345–375.
McCombie, J. S. L. (2011). Criticisms and defences of the balance of payments constrained growth model: Some old, some new. *PSL Quarterly Review, 64*(259), 353–392.
McCombie, J. S. L., & Britto, G. (2009). Thirlwall's law and the long run equilibrium growth rate: An application to Brazil. *Journal of Post Keynesian Economics, 32*(1), 115–136.
McCombie, J. S. L., & de Ridder, J. (1983). Increasing returns, productivity and output growth: The case of the United States. *Journal of Post Keynesian Economics, 5*(3), 414–429.
McCombie, J. S. L., & Fingleton, B. (1998). Increasing returns and economic growth: Evidence from the European Union Regions. *Oxford Economic Papers, 50*(1), 89–105.
McCombie, J. S. L., & Roberts, M. (2007). Returns to scale and regional growth: The static-dynamic Verdoorn law paradox revisited. *Journal of Regional Science, 47*(2), 179–200.
McCombie, J. S. L., & Romero, J. P. (2016a). The multi-sectoral Thirlwall's law: Evidence from 14 developed European countries using product-level data. *International Review of Applied Economics, 30*(3), 301–325.
McCombie, J. S. L., & Romero, J. P. (2016b). *Thirlwall's law and the specification of export and import demand functions: An investigation of the impact of relative productivity growth on trade performance*. Cambridge Centre for Economic and Public Policy Working Paper Series (CCEPP WP 02-2016).
McCombie, J. S. L., & Tharnpanich, N. (2013). Balance of payments constrained growth, structural change and the Thai economy. *Journal of Post Keynesian Economics, 35*(4), 569–598.
McCombie, J. S. L., & Thirlwall, A. P. (1994). *Economic growth and the balance of payments constraint*. London: Macmillan.
McCombie, J. S. L., & Thirlwall, A. P. (1997a). Economic growth and the balance of payments constraint revisited. In P. Arestis, G. Palma, & M. Sawyer (Eds.), *Markets, unemployment and economic policy: Essays in honour of G. Harcourt* (Vol. 2). London: Edward Elgar.

McCombie, J. S. L., & Thirlwall, A. P. (1997b). The dynamic Harrod foreign trade multiplier and the demand-oriented approach to economic growth: An evaluation. *International Review of Applied Economics, 11*(1), 5–26.

McCombie, J. S. L., & Thirlwall, A. P. (2004). *Essays on balance of payments constrained growth*. London: Routledge.

McCombie, J. S. L., Pugno, M., & Soro, B. (2002). *Productivity growth and economic performance: Essays on Verdoorn's law*. London: Macmillan.

McCombie, J. S. L., Angeriz, A., & Roberts, M. (2008). New estimates of returns to scale and spatial spillovers for EU regional manufacturing 1986–2002. *International Regional Science Review, 31*(1), 62–87.

McCombie, J. S. L., Angeriz, A., & Roberts, M. (2009). Increasing returns and growth in industries in the EU regions: Paradoxes and conundrums. *Spatial Economic Analysis, 4*(2), 147–168.

McCombie, J. S. L., Felipe, J., & Naqvi, K. (2010). Is Pakistan's growth rate balance of payments constrained? Policies and implications for development and growth. *Oxford Review of Development Studies, 38*(4), 477–496.

McGregor, P. G., & Swales, J. K. (1985). Professor Thirlwall and balance of payments constrained growth. *Applied Economics, 17*(1), 17–32.

McGregor, P. G., & Swales, J. K. (1986). Balance of payments constrained growth: A rejoinder to Professor Thirlwall. *Applied Economics, 18*(12), 1265–1274.

McGregor, P. G., & Swales, J. K. (1991). Thirlwall's law and balance of payments constrained growth: Further comment on the debate. *Applied Economics, 23*(1), 9–20.

Moreno-Brid, J. (1998). On capital flows and the balance of payments constrained growth model. *Journal of Post Keynesian Economics, 21*(2), 413–433.

Moreno-Brid, J. (2003). Capital flows, interest rate payments and the balance of payments constrained growth model: A theoretical and empirical analysis. *Metroeconomica, 54*(2/3), 346–365.

Pacheco-López, P., & Thirlwall, A. P. (2014). A new interpretation of Kaldor's first law of growth in an open developing economy. *Review of Keynesian Economics, 2*(3), 384–398.

Palley, T. (2002). Pitfalls in the theory of growth: An application to the balance of payments constrained growth model. In M. Setterfield (Ed.), *The economics of demand-led growth*. Cheltenham: Edward Elgar.

Ribeiro, R., McCombie, J. S. L., & Lima, G. (2016). Exchange rate, income distribution and technical change in a balance of payments constrianed growth model. *Review of Political Economy, 28*(4), 545–565.

Romero, J. P., Silveira, F., & Jayme, F. G. (2011). Brazil: Structural change and balance of payments constrained growth. *CEPAL Review, 105*, 185–208.
Rowthorn, R. (1975). What remains of Kaldor's laws. *Economic Journal, 85*(337), 10–19.
Setterfield, M. (2006). Thirlwall's law and Palley's pitfalls: A reconsideration. In P. Arestis, J. S. L. McCombie, & R. Vickerman (Eds.), *Growth and economic development: Essays in honour of A.P. Thirlwall*. Cheltenham: Edward Elgar.
Soukiazis, E., & Cerqueira, P. (Eds.). (2012). *Models of balance of payments constrained growth: History, theory and empirical evidence*. London: Palgrave Macmillan.
Thirlwall, A. P. (1979). The balance of payments constraint as an explanation of international growth rate differences. *Banca Nazionale del Lavoro Quarterly Review, 32*, 45–53.
Thirlwall, A. P. (1980). Rowthorn's interpretation of Verdoorn's law. *Economic Journal, 90*(358), 386–388.
Thirlwall, A. P. (1981). Balance of payments constrained growth: A reply to Mr. McCombie. Banca Nazionale del Lavoro. *Quarterly Review, 139*, 458–459.
Thirlwall, A. P. (1982). The Harrod trade multiplier and the importance of export-led growth. *Pakistan Journal of Applied Economics, 1*(1), 1–21.
Thirlwall, A. P. (2014). Kaldor's 1970 regional growth model revisited. *Scottish Journal of Political Economy, 61*(4), 341–347.
Thirlwall, A. P., & Hussain, M. N. (1982). The balance of payments constraint, capital flows and growth rate differences between developing countries. *Oxford Economic Papers, 34*(3), 498–510.
Vera, L. A. (2006). The balance of payments constrained growth model: A north-south approach. *Journal of Post Keynesian Economics, 29*(1), 66–92.
Verdoorn, P. J. (1949). Fattori che Regalano lo Sviluppo della Produttivita del Lavoro. *L'Industria, 1*, 3–10.
Wells, H., & Thirlwall, A. P. (2003). Testing Kaldor's growth laws across the countries of Africa. *African Development Review, 15*(2/3), 89–105.
Wolfe, T. N. (1968). Productivity and growth in manufacturing industry: Some reflections on Professor Kaldor's inaugural lecture. *Economica, 35(138):* 117–126.
Young, A. A. (1928). Increasing returns and economic progress. *Economic Journal, 38*(152), 527–542.

3

Why Neither Samuelson's Neoclassical Synthesis Keynesianism Nor New Keynesianism Theory Is Compatible with Keynes's General Theory Explanation of the Cause of Unemployment

Paul Davidson

1 Introduction

I define a classic in economics as a book everyone cites but practically no one reads. Thus, Keynes's *General Theory of Employment Interest and Money* is truly an economic classic. I should add that even among those few who claim to have read Keynes's book, it is obvious that most of these readers have not comprehended Keynes's message.

Keynes (1936) stated that "The outstanding faults of the economic society in which we live are its failure to provide for full employment and its arbitrary and inequitable distribution of income and wealth" (p. 372). Some 70 years after Keynes wrote this, the Great Recession, beginning in August 2007, has demonstrated that these 'faults' are still plaguing the

P. Davidson (✉)
Emeritus Chair of Excellence in Political Economy, University of Tennessee, Knoxville, TN, USA

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_3

developed nations around the globe. This chapter will demonstrate that the persistence of these economic faults of the economic system in which we live are due in large part to the fact that what passes for Keynes's theory in mainstream professional journals, textbooks, and so on is not the theory that Keynes specified in his *General Theory* (1936).

Modern macroeconomic theory has flourished professionally in its pursuit of the secrets of long-run economic growth, but has completely neglected the short-run economic problem of unemployment. Modern macroeconomics assumes that in the long run, prices are flexible and the growth of the economy is determined primarily by the growth in the ability to supply goods and services. Thus, increases in technology which produces increases in labor productivity are often cited as the basis for economic growth and prosperity.

But in the short run, it is assumed that we live in an economy where prices and/or money wages are not perfectly flexible. Consequently, growth and full-employment prosperity can be held back because prices and/or money wages are too high, and as a result, market demand is too low to absorb all that can be produced at full employment. Samuelson's mainstream Neoclassical Synthesis Keynesian theory and the next generation of New Keynesian theorists, as well as most widely sold *Principles of Economics* textbooks, claim that Keynes's general theory associates any short-run unemployment problem as being caused by the stickiness or even rigidity of money wages and/or administered prices.

This claim that Keynes's analysis required short-run wage and/or price rigidity, such as a sticky, money wage rate and administered product prices, is, however, exactly what classical theory specified to explain the causes of unemployment. This rigidity of money wages presumption is also the fundamental basis of Samuelson's neoclassical synthesis interpretation of Keynes's *General Theory*. But this presumption of rigidity is in direct conflict with Keynes's own words about what he believed was the basis of his *General Theory of Employment, Interest and Money.*

As Keynes (1936) put it:

> For the classical theory has been so accustomed to rest the supposedly self-adjusting character of the economic system on the assumed fluidity of money wages; and, when there is rigidity, to lay on this rigidity the blame

of maladjustment..... My difference from this theory is primarily a difference of analysis. (p. 257)

To understand the difference between Keynes's analysis of the cause of involuntary unemployment and the rigidity of the money wage rate presumption of Samuelson and modern Neoclassical Synthesis, Keynesian theorists and New Keynesian theorists, and in terms of the causes of unemployment, requires an explanation of how Keynes's macroeconomic analysis was perverted by Samuelson's neoclassical synthesis interpretation of Keynes.[1] We will then be able to provide the explanation of how Keynes's liquidity serious monetary theory and its specification of the essential properties of money and all other liquid assets is, as Keynes noted (1936, p. 257), 'primarily a difference of analysis' of the cause of involuntary unemployment from the classical theory that presumes wage and price rigidity or stickiness is the basic cause of unemployment in the short run—or even in the long run. Once the reader is made aware of what Keynes specified as 'the essential properties' of money and all other liquid assets, then the reader will be able to understand the principles of why international payments imbalances and the financial crisis of 2007–8 has created the worst global economic recessionary period since the Great Depression.

2 Samuelson's Abortion of Keynes's General Theory

In 1941, Samuelson's PhD dissertation won the Wells prize for the best PhD dissertation in economics at Harvard. This dissertation was polished by Samuelson and finally published as *Foundations of Economic Analysis* (1947). Neoclassical Synthesis Keynesianism and New Keynesianism theories are both based on what Samuelson's *Foundations of Economic Analysis* asserts as the necessary Walrasian classical microfoundation of all valid economic theories. If the microfoundations of any macroeconomic analysis are not Walrasian, then, according to Samuelson, this non-Walrasian micro-based macroeconomics is neither a valid theory of economics nor what Keynes meant in his *General Theory*.

Nevertheless, Keynes (1936, pp. 176–177) stated that Walras "is strictly in the classical tradition"—a tradition that Keynes's *General Theory* was attempting to replace. Moreover, as already noted on p. 257 of the *General Theory*, Keynes explicitly denied that his theory of unemployment required the classical theory presumption of rigid money wages.

After reading Keynes's *General Theory of Employment, Interest and Money* in 1936, Samuelson stated that he found Keynes's *General Theory* analysis 'unpalatable' and not comprehensible (Colander and Landreth 1996, p. 159). Samuelson then explicitly stated that "The way I finally convinced myself was to just stop worrying about it (about understanding Keynes's analysis). I asked myself: why do I refuse a paradigm that enables me to understand the Roosevelt upturn from 1933 till 1937? … *I was content to assume that there was enough rigidity in relative prices and wages to make the Keynesian alternative to Walras operative*" (Colander and Landreth 1996, pp. 159–160, emphasis added).

In the Preface to the German language edition of *The General Theory*, however, the following sentences appear: "One of the reasons for justifying my calling my theory a *General* theory. Since it is based on fewer restrictive assumptions ("weniger enge Vorasusstezungen Stutz") than the orthodox (classical) theory. It is also more easily adopted to a large area of different circumstances".[2] In other words, Keynes's general theory is a classical theory from which some restrictive axioms fundamental to classical theory have been removed.

On the other hand, Samuelson has claimed that the Walrasian general equilibrium analysis is the fundamental foundation of any general theory of economics and Keynes's analysis was a special case where an additional restrictive assumption of rigidity of money wages and/or product prices was added to the Walrasian microfoundation to develop an analysis where the gross substitution effect (which require freely flexible prices and wages) cannot work to assure full employment in the short run.

Samuelson (as in Colander and Landreth 1996, p. 163) explicitly stated that in his view Keynes's analysis is merely a "very slowly adjusting disequilibrium …. [where] the full Walrasian equilibrium was not realized" in the short run because prices and money wages do not adjust rapidly enough to an exogenous shock. Nevertheless, the economic system would, if left alone, achieve full employment in the long run as all prices and wages are variable. It therefore follows that Samuelson as a 'Keynesian' has urged additional government spending to achieve full employment before there was at the long run Walrasian adjustment to flexible wages and prices, merely because he is too impatient to wait to leave it to the market to permit the slowly adjusting Walrasian system to achieve this longer-run goal of full employment.

It is worth noting that Samuelson's slowly adjusting Walrasian micro-foundation interpretation of Keynes apparently convinced all mainstream 'Keynesian' economists after the Second World War to treat Keynes's theory as a 'special case' of the *general* classical (Walrasian) theory where full employment is inevitable if all money wages and money prices are freely flexible. In the Samuelson and mainstream Keynesian economists' post-Second World War view, the classical Walrasian general theory involved freely flexible wages and prices, while Keynes's theory was merely a special case of the Walrasian system, where this special case presumed an additional restrictive assumption of sticky or rigid money wages and/or prices is added to the general Walrasian theory. Following Samuelson's insistence that the foundation of all economic theories explaining the economic system in which we live is a Walrasian slowly adjusting process, mainstream Keynesian theorists after the Second World War always taught their students that Keynes's explanation of involuntary unemployment required the fact that money wages and prices were not freely flexible.

Since Samuelson claimed he read *The General Theory of Employment, Interest and Money* and found it 'unpalatable', most economists after the war were convinced by Samuelson's use of mathematics in economics that reading exactly what Keynes wrote would not be very productive. Instead they merely adopted Samuelson's interpretation of Keynes's book as a classic in economics and they apparently never even tried to read Keynes's book. Thus, all mainstream Keynesian theories ignored Chap. 19 in the *General Theory*, entitled 'Changes in Money Wages'.

In this chapter, Keynes explicitly denies the validity of this rigidity of wages and prices assertion as a basis for his theory of involuntary unemployment. Using the Marshallian micro theory—instead of the Walrasian micro theory—as the microfoundation of his general theory, Keynes is able to demonstrate in Chap. 19 that even if the economy possessed freely flexible money wages and prices, involuntary unemployment equilibrium can still occur and persist.

Samuelson never tried to comprehend Keynes's use of Marshallian micro analytical foundation and framework for his *General Theory*. For in 1986 Samuelson was still claiming that "we [Keynesians] always assumed that the Keynesian underemployment equilibrium floated on a substructure of administered prices and imperfect competition" (Colander and Landreth 1996, p. 160). When pushed by Colander and Landreth as to whether this requirement of rigidity was ever formalized in his work, Samuelson's response was: "There was no need to" (Colander and Landreth 1996, p. 161). Clearly had Samuelson and any of his Keynesian followers read and understood Chap. 19 of Keynes's classic in economic theory, Samuelson would have to have noted that there was at least one 'Keynesian' named John Maynard Keynes, who had not assumed that the "Keynesian underemployment equilibrium floated on a substructure of administered prices and imperfect competition".

Specifically in Chap. 19 of the *General Theory*, and even more directly in Keynes's published response to Dunlop (1938) criticism of Keynes's analysis, Keynes (1939) had already responded in the negative to this question of whether his analysis of less than full-employment equilibrium required imperfect competition, administered prices, and/or rigid wages. Dunlop (1938) had argued that the purely competitive model was not empirically justified; therefore, it was monopolistic price and wage fixities that must be the realistic basis of Keynes's involuntary unemployment equilibrium analysis. Keynes's (1939) reply to Dunlop was simply: "I complain a little that I in particular should be criticised for conceding a little to the other view" (p. 411).

In Chaps. 17–19 of his *General Theory*, Keynes explicitly demonstrated that even if a purely competitive economy with perfectly flexible money wages and prices existed ('conceding a little to the other side'), there was no automatic competitive market mechanism that could restore the full-employment level of effective demand in his theory if some involuntary unemployment existed. In other words, Keynes's general theory, using Marshallian microfoundations, could show that, as a matter of logic, less than full-employment equilibrium could exist in a purely competitive economy with freely flexible wages and freely competitive product prices.

Obviously Samuelson, who became the premier American 'Keynesian' of his time, had either not read or not comprehended (1) Keynes's response to Dunlop or even (2) Chap. 19 of the *General Theory*.

Keynes (1936) indicated that to assume that wage rigidity was *the* sole cause of the existence of an unemployment equilibrium implied accepting the argument that the Marshallian micro-demand functions "can only be constructed on some fixed assumption as to the nature of the demand and supply schedules of other industries and fixity as to the amount of aggregate effective demand. It is invalid, therefore to transfer the argument to industry as a whole unless we also transfer the argument that the aggregate effective demand is fixed. Yet this assumption reduces the argument to an *ignoratio elenchi*" (p. 259).

An ignoratio elenchi is a fallacy in logic of offering a proof irrelevant to the proposition in question. Unfortunately Samuelson invoked the same classical ignoratio elenchi when he argued that Keynes's general theory was simply a slowly adjusting Walrasian general equilibrium system where if there is insufficient aggregate effective demand to produce full employment, then rigid wages and prices created a temporary disequilibrium that prevented full-employment equilibrium from being restored in the short run.

As Keynes (1936) went on to explain,

> whilst no one would wish to deny the proposition that a reduction in money wages *accompanied by the same aggregate effective demand as before* will be associated with an increase in employment, the precise question at issue is whether the reduction in money wages will or will not be accompanied by the same aggregate effective demand as before measured in term of money, or, at any rate, by an aggregate effective demand which is not reduced in full proportion to the reduction in money-wages. (pp. 259–260)

Keynes (1936) then spent the rest of Chap. 19 analyzing the question that if the economy was not initially at full employment, how any reduction in the money wage rate would affect both the aggregate supply function and the aggregate demand function. A reduction in money wage rate reduced money costs of production at every level of employment and therefore would shift downward the aggregate supply curve function 'in terms of money'. The money wage reduction implies a reduction in aggregate money wage income at every possible level of employment. This reduction in wage earners' money income would reduce their aggregate money spending on consumption by wage earners at each level of employment. The result is to also shift downward the aggregate demand function 'in terms of money'.

Does the downward shift in both the aggregate demand curve and the aggregate supply curve in money terms result in their intersection at the same initial level of unemployment or will a new point of intersection of these aggregate supply and aggregate demand curves after their shifts be at full employment, or at least a higher level of employment? Keynes's Marshallian microfoundation answer was there was no reason to believe that the two downward-shifting curves' intersection would produce a higher equilibrium level of employment.

This question of where the point of effective demand would be if all prices and wages were flexible, by assumption, is not relevant to a Walrasian system or Samuelson's Neoclassical Synthesis Keynesianism which assumes a slowly adjusting Walrasian system. A Walrasian system is built on the assumption that with flexible wages and prices, there will always be a sufficient market demand to purchase all that is produced by a fully employed economy at profitable prices for the entrepreneurs. There can never be a lack of aggregate effective demand if all wages and prices are flexible.

At the same time that Samuelson was developing his Neoclassical Synthesis Keynesianism, he was working on cleaning up his masterful *Foundations of Economic Analysis* (1947) in which Samuelson 'demonstrates' that the Walrasian system is *the* foundation for all economic theory. In this 1947 book, Samuelson asserted certain specific classical axioms are necessary for the foundation of all economic analysis. For example, Samuelson (1947) noted that "in a purely competitive world it would be foolish to hold money as a store of value as long as other assets had a positive yield" (pp. 122–124). This statement implies that (1) producible capital goods (plant and equipment) that provide a positive yield of output are preferable to money (or any other liquid asset) as a substitute form in which to hold one's store of value savings and therefore (2) money is neutral in the sense that changes in the quantity of money *per se* cannot affect the level of employment and output.

Keynes (1935), however, rejected the neutral money axiom when he wrote:

> the theory which I desiderate would deal…with an economy in which money plays a part of its own and affects motives and decisions, and is, in short, one of the operative factors in the situation, so that the course of events cannot be predicted either the long period or in the short, without a knowledge of the behavior of money between the first state and the last. And it is this which we mean when we speak of a monetary economy…. Booms and depressions are peculiar to an economy in which …*money is not neutral*. (pp. 408–409)

Furthermore, in Chap. 17 of the *General Theory*, Keynes (1936, p. 231) explicitly stated that in his liquidity theory, real producible capital goods are not gross substitutes for money or any liquid asset as a form for holding one's savings. Accordingly, Keynes explicitly rejected the ubiquitous use of the gross substitution *and* the neutral money presumption of classical economic theory as a foundation of his macroeconomic theory. Consequently, by rejecting the ubiquitous use of the gross substitution and neutral money axioms (which is a specific requisite of the Walrasian classical theory), Keynes is providing a more general theory (because it is based on fewer assumptions) than the classical Walrasian theory.

Furthermore, Samuelson (1969) argued that the "ergodic hypothesis (axiom)" (p. 184) is a necessary foundation if economics is to be a hard science. But as we have also explained in detail elsewhere (Davidson 1982–1983, 2007, 2011, 2015), Keynes's concept of an uncertain future requires the rejection of this ergodic axiom.

Since a general theory is one that has fewer restrictive axioms than another theory, Keynes's theory, which rejects three restrictive classical presumptions, namely, the neutral money axiom, the ubiquity of the gross substitution axiom, and the ergodic axiom, is, by definition, a more general theory than the foundations of the Neoclassical Synthesis Keynesian macroeconomic theory developed from the Walrasian classical equilibrium theory by Samuelson. These three classical restrictive axioms are "the postulates of the classical theory …applicable to a special case only and not to the general case.…Moreover the characteristics of the special case assumed by classical [Walrasian] theory happen not to be those of the economic society in which we actually live, with the result its teaching is misleading and disastrous if we attempt to apply it to the fact of experience" (Keynes 1936, p. 3).

3 Keynes's Theory Is Primarily a Difference of Analysis

As Keynes's developed his theory of liquidity preference, he recognized that his explanation of the existence of involuntary unemployment required specifying 'The Essential Properties of Interest and Money' (1936, Chap. 17) that differentiated his analytical results from classical theory. These 'essential properties' assured that money and all other liquid assets are never neutral. Keynes (1936, pp. 230–231) specified these 'essential properties' as:

1. The elasticity of production of all liquid assets including money is zero or negligible; and
2. The elasticity of substitution between liquid assets (including money) and reproducible goods is zero or negligible.

"The attribute of 'liquidity' is by no means independent of the presence of these two characteristics" (Keynes 1936, p. 241n.1). In other words, all liquid assets have these two *essential* elasticity characteristics.

A zero elasticity of production means that money does not grow on trees and consequently workers cannot be hired to harvest money trees to provide people with more money when the demand for money increases. Or as Keynes put it: "money…cannot be readily reproduced; labour cannot be turned on at will by entrepreneurs to produce money in increasing quantities as its price rises" (Keynes 1936, p. 230).

Accordingly, when people save some portion of their current income instead of spending their entire current income on buying producible goods, then, ceteris paribus, the demand for producibles, is reduced, while the resulting savings out of income is used to demand money (and/or other liquid assets with the same essential elasticity properties). Since the market demand for producibles is reduced by any increase in savings out of current income, then employers will hire fewer workers as they face a decline in market demand, and there will be unemployment and less production of goods and services.

Since the production elasticity of money and liquid assets is zero, private sector entrepreneurs cannot hire the unemployed labor to produce more money (or other liquid assets) to meet this increase in demand by savers for liquid nonreproducible (by the private sector) assets. In Keynes's theory, liquidity is a concept where if one has sufficient liquidity, one has the ability to meet all money contractual obligations as they come due.

Friedman, in his permanent income theory, avoids this problem where we have specified that any additional positive aggregate savings will go to the demand for liquid assets and therefore, ceteris paribus, create less demand for producible goods and services, thereby reducing the demand for entrepreneurs to employ workers. For Keynes and most normal persons, savings is typically defined as being that part of current income that is not spent on producible goods and services. Savers store their savings in the form of liquid assets.

Friedman, however, defines savings differently. For Friedman savings is the utility stored in a producible durable goods that is not consumed in the current accounting period. Since the consumption of any good, in Friedman's theory, produces utility for the consumer, the purchase price of a newly produced durable good at the moment of purchase is the value of all the utility that is stored in this produced durable good—the utility will be provided to the consumer over the useful life of this produced durable. Thus, at the moment of purchase of a newly produced durable, the total utility that the durable will produce is being *saved* for the future periods when the durable will be slowly consumed over its useful life.

According to Friedman's definition of savings out of current income, savings are not stored in money or other liquid assets such as bonds, stocks, bank savings accounts, and so on. Instead savers are using current savings out of income to purchase newly produced durables; and therefore, spending on savings out of current income creates jobs in the producible durable goods industries just as much as consumption spending out of current income on nondurables produced goods and services creates jobs.

In classical Walrasian theory, however, money is a reproducible commodity. In many neoclassical textbook models of a Walrasian system, peanuts or some other easily reproducible product of industry is the money commodity or numeraire. Peanuts may not grow on trees, but they do grow on the roots of bushes. The supply of peanuts can easily be augmented by the hiring of additional workers by private sector entrepreneurs when the demand for peanut money (or any readily producible asset money) increases.

Keynes insisted that one essential property of money and all liquid assets is the elasticity of production is approximately zero. Accordingly, if the savings out of current income take the form of increasing the demand for nonproducibles money (or other liquid assets), then the price of this demand for liquid assets rises relative to the price of producible durables. The zero elasticity of substitution assures that portion of income that is not spent on by the products of industry for consumption purposes, that is, savings, will, in Hahn's (1977, p. 31) terminology, find 'resting places' in the demand for nonproducibles, that is, liquid assets. Producible real capital goods are not a gross substitute for liquid assets as places where savers will store their savings. Some 40 years after Keynes, Hahn (op, cit.) rediscovered Keynes's point that a stable involuntary unemployment equilibrium could exist *even in a purely competitive system with flexible wages and prices* whenever there are "resting places for savings in other than reproducible assets" (p. 31).

Hahn (1977) rigorously demonstrated what may have been logically intuitive to Keynes. Hahn (op. cit.) showed that the view that with "flexible money wages there would be no unemployment has no convincing argument to recommend it .... Even in a pure tatonnement in traditional models convergence to [a general] equilibrium cannot be generally proved" (p. 37) if savings were held in the form of nonproducibles. Hahn (1977) also argued that "any non-reproducible asset allows for a choice between employment inducing and non-employment inducing demand" (p. 39).

Accordingly, given the 'essential properties' of money and other liquid assets specified by Keynes, the existence of a demand for any liquid non-reproducible assets (when the products of the capital goods-producing industries are not gross substitutes) as a store of 'savings' creates the potential for involuntary unemployment. Any saved income is not, in the short or long run, necessarily spent on the products of industry. Households that save (i.e., they do not spend a portion of their income on the products of industry) store that portion of their income that they do not consume in liquid assets are choosing, in Hahn's (1977) words, "a non-employment inducing demand" (p. 39) for their savings.

However, if the gross substitution axiom was universally applicable, any new savings that would increase the demand for nonproducibles would therefore increase the price of nonproducibles (whose production supply curve is, by definition, perfectly inelastic). The resulting relative price rise in nonproducibles vis-à-vis the price of producibles would, if gross substitution was universally applicable, induce savers to increase their demand for reproducible durables as a substitute for nonproducibles for storing their savings holdings. Consequently, nonproducibles could not be the ultimate 'resting places' for savings for when the price of nonreproducible liquid assets rose, savers will substitute producibles and therefore their savings will spill over into a demand for producible goods.

Samuelson's assumption of a Walrasian system where all demand curves are based on a ubiquitous gross substitution axiom implies that everything is a substitute for everything else. In Samuelson's foundation for economic analysis, therefore, producibles must be good (or better) gross substitutes for any existing nonproducible liquid assets (including money) when the latter are used as stores of savings. Accordingly, Samuelson's *Foundations of Economic Analysis* explicitly denies the logical possibility of involuntary unemployment as long as all prices are perfectly flexible. Samuelson's brand of Keynesianism is merely a form of the classical special case analysis that is "misleading and disastrous" (Keynes 1936, p. 3) if applied to the real world. In Keyes's general theory, any increase in the demand for 'savings' in liquid form over time raises the relative price of nonproducibles, but will not spill over into a demand for producible goods.

Accordingly, if at a full-employment level of income decision makers want to save a portion of their income in liquid nonproducibles, then they will have made a choice for 'nonemployment-inducing demand'. To offset this nonemployment-induced demand and maintain full employment, other decision makers must spend more than their full-employment income in the marketplace on producibles. To spend in excess of their full-employment income, these decision makers must either spend a portion of their previous savings on producibles or borrow new money from the banking system to spend on producible goods and services.

In an international context, if any nation runs a persistent surplus in its balance of payments, then it is saving its excess of income earned from exports over its payment for imports, to obtain liquid foreign reserves, namely, an international noninducing employment demand. Thus, Keynes (1941) put forth as a principle that if any nation persistently runs international payments surpluses, it creates a significant shortage of international effective demand for its trading partners. Consequently, to remove the international sector from creating employment problems for any nation, Keynes (1941) required any persistent surplus creditor nation to spend down its saved accumulated liquid international reserves. Keynes (op. cit., p. 176) argued that the onus should be on the creditor saver nations to solve this accumulating noninducing employment demands for international liquid reserves and therefore to encourage the creditor nations to provide international expansionary economic forces. After all, this onus is not costly to the creditor nation for it has the international liquidity wherewithal to engage in such an activity. Thus, Keynes saw the necessity of creating an international institution where all trading nations agree to a 'rule of the road' that requires persistent creditor nations to spend down their excessive foreign reserves. This would solve any international payment imbalance problem by placing an expansionist pressure on world trade.

If, instead of relying on an international institution's rule of the road and creating an expansionist pressure on global trade, the onus of reducing a nation's deficit in its international payments was placed on the debtor nation to somehow obtain a devaluation in its exchange rate to improve its balance of payments position by making its industries 'more competitive', then the effect would not only reduce the standard of living for the residents of the debtor nation, but it would put 'in place a contractionist pressure on world trade' (Keynes 1941, p. 176).

Finally, Keynes (1941) argued that only in a money-using entrepreneur economy where the future is uncertain (and therefore could not be reliably predicted) would money (and all other liquid assets) always be nonneutral as they are used as a store of value savings. In essence Keynes (op. cit.) viewed the economic system as moving through calendar time from an irrevocable past to an uncertain, not statistically predictable, 'real' future. This required Keynes to reject the ergodic axiom.

Samuelson's slowly adjusting Walrasian system view of Keynes's theory resulted in aborting Keynes's revolutionary analysis from altering the foundation of mainstream macroeconomics from classical microeconomic theory. Consequently, what passes as conventional macroeconomic wisdom of mainstream economists, such as Krugman, Akerloff, and Temin, at the beginning of the twenty-first century, is nothing more than a high-tech and more mathematical version of the nineteenth-century classical theory.

Current economic policies, such as the need for 'austerity' and the fear of government deficit spending increasing the national debt, policies adopted in the United States, the United Kingdom, and the Euro zone, demonstrate the real-world economic damage that Samuelson's proclamation that his 'reconstructed' Keynesianism provided the correct analytical foundations for understanding the economic world in which we live. Instead, had the correct explanation of Keynes's *General Theory* been taken up by mainstream economists and politicians, the world we live in would have been a more prosperous and civilized economic society.[3]

4 A Serious Monetary Theory

Arrow and Hahn in their book *General Competitive Equilibrium* (1971; emphasis added) explicitly state that even in a general equilibrium system "The terms in which contracts are made matter. In particular, if money is the goods in terms of which contracts are made, then the prices of goods in terms of money are of special significance. This is not the case if we consider an economy without a past or future…. If a *serious monetary* *theory* comes to be written, the fact that contracts are made in terms of money will be of considerable importance" (256–257).

Keynes wrote a 'serious monetary theory' since his theory of liquidity recognized that (1) the economy has a past and an uncertain future and (2) all market transactions are organized by the use of legal money-denominated contracts that specify money as the means of contractual settlement for all spot and forward contractual obligations. All market transactions are organized by the use of spot and/or forward money-denominated contracts. *The essence of a capitalist economic system involves a legal money-denominated contract system*. Hence, the need for liquidity to meet one's market money contractual obligations must have an important impact on decision makers choosing what to buy and what to sell and when to take on these contractual obligations. This legal money contract analysis is absent from the works of Samuelson, and other mainstream 'Keynesians' and therefore their theory is not even a 'serious monetary theory'.

This use of money to settle all market transaction contracts, including international transaction contracts, is ignored by mainstream economists when they discuss changing the exchange rate, where the latter, in fact, must alter the sum of the specified money that must be obtained to settle an international contractual obligation for the buyer or seller or both. Indeed, whenever any international forward contract spans the moment in calendar time when an exchange rate change occurs, this can create a liquidity problem in obtaining sufficient funds in terms of the money specified in the international contract—a problem which apparently is not important to today's mainstream international macroeconomists! But it is important to Keynes in his international analysis and even important to entrepreneurs engaged in international contractual transactions in the world we live in.

Since Greece and the other southern Mediterranean Euro nations have the same currency as their Euro trading partners, it is not possible for these deficit nations to look to a monetary exchange rate devaluation to achieve a better trade balance. Accordingly 'austerity' is a policy designed to induce a more favorable trade balance for Greece since it is presumed that austerity depresses the income of Greek workers and Greek enterprises sufficiently so that a reduced Euro price of Greek products induces a gross substitution effect to reduce Greek imports from Germany (and the rest of the Euro-zone nations) and significantly increase Greek exports to the Euro-zone nations. The result will be socially distressing and divisive politically for all Greek residents but would supposedly end up with slower outflow of Euros, in which case the inflow of Euros on trade contractual obligations emerges. Would not an institutional rule such as the Keynes Plan of Bretton Woods, which would put the onus on the nations running a trade surplus to spend more Euros on the products of the balance of trade deficit nations be better? The answer is positive in that the result would be that the deficit balance of the trade nation could earn sufficient Euros to service its debt obligations to the surplus nations, while employment and income of the deficit nation's population would rise and the nation would become more prosperous.

5 Concluding Remarks: The Role of Monetary Policy

Finally, it should be pointed out that the classical presumption of neutral money means that any increase in the money supply will immediately be spent to purchase newly produced goods and services, while none of the increase is used merely to store savings. Accordingly, the neutral money axiom provides a necessary condition for the quantity theory of money. This quantity theory states that any increases in the supply of money greater than any increase in total production by a fully employed labor force will directly increase the rate of price inflation. Nobel Laureate Milton Friedman (1970) stated his belief that money is neutral when he wrote: "We have accepted the quantity theory presumption…that changes in the quantity of money as such *in the long run* have a negligible effect on real income, so that nonmonetary forces are 'all that matter' for changes in real income [total production or GDP] over the decades and 'money does not matter'. On the other hand, we have regarded the quantity of money …as all that matter for …the price level" (p. 27).

Oliver Blanchard (1990) characterized all the mainstream econometric models used by government agencies, central banks, and in academia, as follows: “All the models we have seen impose the neutrality of money as a maintained assumption. This is very much a matter of faith, based on theoretical considerations rather than on empirical evidence” (p. 828).

It follows that most mainstream economists and almost all politicians believe that the primary role of the central bank is to control the rate of expansion of the money supply in order to directly affect the rate of inflation. Since in Keynes’s theory, money is not neutral and therefore the role of monetary policy is not to directly affect the rate of price inflation but rather to assure orderliness in the operation of public financial markets by providing sufficient liquidity in order to avoid any financial crisis.

For the most part, central bankers such as Alan Greenspan or Janet Yellen tend to indicate to the public that easing or tightening of monetary policy depends on the rate of inflation that the economy is experiencing. In recent years, the Board of Governors of the Federal Reserve has indicated a 2 percent rate of price inflation is a target rate for monetary policy. Thus, when the rate of inflation is less than 2 percent, the public can expect an easy monetary policy, while if the rate of inflation exceeds 2 percent, one can expect that the Federal Reserve will engage in a tightening of monetary policy.

When the global financial crisis of 2007–8 occurred, however, the Federal Reserve engaged in a policy that was labeled quantitative easing or QE. In essence, a policy of quantitative easing was adopted because the Federal Reserve recognized that the disorderly price collapse of many derivative financial markets involved a rush by market participants to make a fast exit from holding these derivative securities in their portfolios. This set off fears of participants in many other financial markets of a possible spreading of price disorderliness. Such fears induced financial market participants to make a fast exist in order to substitute some global liquid money holdings as the safe financial asset harbor to protect the liquidity of their portfolio holdings.

Consequently, the Federal Reserve apparently saw the need for it to become the market maker in US government bonds and many derivative securities by entering these markets to directly purchase over $4 billion of US government bonds and derivative securities. By doing so, the Federal Reserve created significant additional liquidity and therefore the feeling of more security for participants in financial markets.

It is evidently clear that the quantitative easing policy of the Federal Reserve prevented the Great Recession which began in 2008 (and was a result of the global financial crisis of 2007–8) from developing into a Great Depression similar to the global economic collapse of the 1930s.

Notes

1. A more detailed analysis of how mainstream macroeconomics is not Keynes's economics is presented in Davidson (2015, chap. 5).
2. These sentences do not appear in *The Collected Writings of John Maynard Keynes*, vol. 17. B. Schefold (1980, pp. 175–176) has called attention to the fact these sentences appear in the German language edition but not in *The Collected Writings.*
3. For example, see Davidson (2017).

References

Arrow, K. J., & Hahn, F. H. (1971). *General competitive equilibrium*. San Francisco: Holden-Day.

Blanchard, O. (1990). Why does money affect output? In B. M. Friedman & F. H. Hahn (Eds.), *Handbook of monetary economics* (Vol. 2). New York: North Holland Press.

Colander, D. C., & Landreth, H. (1996). *The coming of Keynesianism to America*. Cheltenham: Edward Elgar.

Davidson, P. (1982–1983). Rational expectations: A fallacious foundation for studying crucial decision making processes. *Journal of Post Keynesian Economics, 5*(2), 182–198.

Davidson, P. (2007). *John Maynard Keynes*. New York: Palgrave Macmillan.

Davidson, P. (2011). *Post Keynesian macroeconomic theory* (2nd ed.). Cheltenham: Edward Elgar.

Davidson, P. (2015). *Post Keynesian theory and policy: A realistic analysis of the market oriented capitalist economy*. Cheltenham: Edward Elgar.

Davidson, P. (2017). *Who's afraid of John Maynard Keynes?* London: Palgrave Macmillan.

Dunlop, J. G. (1938). The movement of real and money wages. *The Economic Journal, 58*(191), 413–434.

Friedman, M. (1970). A theoretical framework for monetary analysis. In R. J. Gordon (Ed.), *Milton Friedman's monetary framework: A debate with his critics*. Chicago: University of Chicago Press.

Hahn, F. A. (1977). Keynesian economics and general equilibrium theory. In G. C. Harcourt (Ed.), *The microfoundations of macroeconomics*. London: Macmillan.

Keynes, J. M. (1935). A monetary theory of production. In D. Moggridge (Ed.), *The collected writings of John Maynard Keynes* (Vol. XIII). London: Macmillan. 1973. All references are to reprint.

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. New York: Harcourt, Brace.

Keynes, J. M. (1939). Relative movements of real wages and output. *The Economic Journal, 49*. Reprinted in Moggridge, D. (Ed.). (1973). *The collected writings of John Maynard Keynes* (Vol. XIV). London: Macmillan. All references are to reprint.

Keynes, J. M. (1941). Proposal for an international currency union. Reprinted in Moggridge, D. (1980). *The collected writings of John Maynard Keynes* (Vol. 25). Cambridge. All references are to reprint.

Samuelson, P. A. (1947). *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Samuelson, P. A. (1969). Classical and neoclassical theory. In R. W. Clower (Ed.), *Monetary theory*. London: Penguin.

4

The Role of Commercial Banks and Financial Intermediaries in the New Consensus Macroeconomics (NCM): A Preliminary and Critical Appraisal of Old and New Models

Giuseppe Fontana and Marco Veronese Passarella

1 Introduction[1]

The early 1990s were marked by a convergence of views in mainstream macroeconomics. That convergence gave rise to the so-called New Consensus in Macroeconomics (NCM hereafter), which conquered the academic word, central banks and other major policy-making institutions around the world (Arestis 2007; Tovar 2009; Woodford 2009). The NCM was regarded as a 'new neoclassical synthesis' incorporating important elements of both New Keynesian economics and Real Business Cycle economics (Goodfriend and King 1997; Goodfriend 2004; Dixon 2008;

G. Fontana (✉)
University of Leeds, Leeds, UK

University of Sannio, Benevento, Italy

M. V. Passarella
University of Leeds, Leeds, UK

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_4

Fontana 2009b; McCombie and Pike 2013). In 2000 John Taylor, a leading contributor to the development of this new macroeconomic paradigm, listed the most original features of the NCM:

> First, the long-run real GDP trend, or potential GDP, can be understood using the growth model that was first developed by Robert Solow and that has now been extended to make 'technology' explicitly endogenous. Second, there is no long run trade-off between inflation and unemployment, so that monetary policy affects inflation, but is otherwise neutral with respect to real variables in the long run. Third, there is a short run trade-off between inflation and unemployment with significant implications for economic fluctuations around the trend of potential GDP; the trade-off is due largely to temporarily sticky prices and wages. Fourth, expectations of inflation and of future policy decisions are endogenous and quantitatively significant. Fifth, monetary policy decisions are best thought of as rules, or reaction functions, in which the short-term nominal interest rate (the instrument of policy) is adjusted in reaction to economic events. (Taylor 2000, p. 90)

Much has happened between 2000 and today, including a worldwide dramatic financial and banking crisis, a consequent devastating recession and a prolonged stagnation period, which continues today. There is now a unanimous consensus in the economic profession that commercial banks (banks for short) and financial intermediaries are at the heart of these remarkable economic events. Therefore, it may seem odd that Taylor (2000) does not mention banks and financial intermediaries among the most original features of the NCM. It was not a mistake or oversight of the paper. Banks and financial institutions were rarely mentioned, let alone modelled, in the original NCM model (e.g. Woodford 2003; see, also, for a critical analysis, Goodhart 2010). This is really extraordinary vis-à-vis the fact that the NCM model was enthusiastically adopted by most central banks and treasuries around the world (see, e.g. Adolfson et al. 2007; Smets and Wouters 2003, 2007; Tovar 2009). Yet, some interesting attempts to account for banks and financial institutions in mainstream macroeconomic modelling were made in the early 1980s. The so-called financial accelerator mechanism (FAM hereafter) literature pioneered by Ben Bernanke (Bernanke 1981, 1983; Bernanke and Gertler

1989) analysed the role of banking and financial frictions as triggers or amplifiers of the business cycle. The FAM literature has been rediscovered after the 2007–2008 financial crisis. It has been used to amend the NCM model to account for the nature and role of banks and financial intermediaries in modern economies (see, among others, Christiano et al. 2013; Del Negro et al. 2014).

The purpose of this chapter is to offer a preliminary and critical review of the progress made in mainstream macroeconomics in the last two to three decades. Did really the NCM model ignore banks and financial intermediaries? Are the core theoretical propositions of the NCM unable to explain some important features of real-world economies, and especially the remarkable economic events of the last decade? And how the insights and results of the FAM literature have been encompassed into the NCM model? What are the prospects, if any, for the new NCM cum FAM model to explain the nature and role of banks and financial intermediaries in modern economies? These are some of the main questions that this chapter tries to answer.

The chapter is organised as follows. Section 2 presents the three-equation model describing the macroeconomic core of the NCM in a closed economy, the so-called benchmark NCM model. It highlights the nature and role of the 'rational expectations hypothesis' (REH hereafter), and the concept of the 'natural equilibrium' in the benchmark NCM model. It also proposes an amended version of the latter, which takes into account criticisms raised against the use of the REH and natural equilibrium in the NCM. The so-called augmented NCM model allows for the possibility of interdependence between aggregate demand and aggregate supply, a hallmark of real-word economies captured by demand-led Post Keynesian economic models (Setterfield 2002, 2010; Fontana 2009a, 2010; Palacio-Vera 2009), via some hysteresis effects. Section 3 reviews the core features and main results of the FAM literature and presents a simple set of equations describing the 'benchmark FAM model'. It also draws attention to the nature and role of banks and financial intermediaries in the latter model. Furthermore, it proposes an augmented NCM cum FAM model ('augmented FAM model', for short), which allows for hysteresis effects. A table summarises all models discussed in the chapter, namely, the benchmark NCM model, the augmented NCM model, the benchmark

FAM model, and the augmented FAM model. Section 4 assesses the current state of mainstream macroeconomics. It reviews many attempts that have been made to amend the NCM model in order to fix its failure to explain the remarkable economic events of the last decade. It highlights how the absence of banks and financial intermediaries, and systematic errors in inflation forecasting, has been accounted for in the new augmented FAM model. It also discusses the inability of the latter to allow for the possibility that financial instability is an endogenous by-product of the normal functioning of modern economies. Furthermore, it considers the monetary policy implications of the augmented FAM model, including the possibility of replacing the dominant price stability goal with an alternative financial stability goal, which aims to stabilise the market value of private and public financial assets. Finally, Sect. 5 offers some final remarks.

2 A Critical Analysis of the Benchmark NCM Model

The macroeconomic core of the NCM can be described through three reduced-form (or aggregate) equations, namely, an aggregate demand equation, an inflation equation and an interest rate rule. Each macroeconomic equation is in turn strictly microeconomics-founded, that is, every relationship among aggregate magnitudes is derived from the constrained inter-temporal optimisation of an individual utility function. This function underpins the behaviour of a single, sovereign, completely rational representative agent with perfect foresight, who maximises its utility over an infinite horizon by combining labour supply/leisure time and consumption/saving in each period. McCombie and Pike (2013) label these features the 'paradigmatic heuristics' or 'pseudo-assumptions' of the NCM model (see also McCombie and Negru 2014, who explore the general question of paradigm-dependent economic theories).

In simple algebraic terms, the reduced-form benchmark model can be represented as follows (Clarida et al. 1999; De Grauwe 2010; see, for a critical assessment of it, Arestis 2007, 2009; Arestis and Sawyer 2004, 2006, 2008):

$$Y_t^g = a_0 + a_1 Y_{t-1}^g + a_2 E\left(Y_{t+1}^g\right) - a_3\left[r_t - E\left(\pi_{t+1}\right)\right] + \varepsilon_1 \tag{4.1}$$

$$\pi_t = b_1 Y_t^g + b_2 \pi_{t-1} + b_3 E\left(\pi_{t+1}\right) + \varepsilon_2 \tag{4.2}$$

$$r_t = \left(1 - c_3\right)\left[RR_t^* + E\left(\pi_{t+1}\right) + c_1 Y_{t-1}^g + c_2\left(\pi_{t-1} - \pi^T\right)\right] + c_3 r_{t-1} + \varepsilon_3 \tag{4.3}$$

where Y_t^g is the current output gap, π_t is the current inflation rate, π^T is the target inflation rate, r_t is the current nominal interest rate, RR_t^* is the natural or optimal real interest rate, $E(\cdot)$ defines future expected values, a_0, a_1, a_2, a_3, b_1, c_1, $c_2 > 0$, $(b_2 + b_3) = 1$ and $0 < c_3 < 1$.[2]

Equation (4.1) corresponds to the old *IS* curve and is grounded on the separation between aggregate demand and aggregate supply, with the (growth of) natural output being supply-determined and independent of the level, and rate of change, of aggregate demand (Fontana 2010). It shows that the output gap—that is, the difference between the (logarithm of) actual output and its 'natural' or potential or long-run level—depends negatively on the expected real interest rate. The output gap depends also positively on the past and expected future output gaps. Equation (4.2) corresponds to the 'accelerationist' (or expectations-augmented, New Keynesian) Phillips curve, acting as the aggregate supply function. It shows that the inflation rate depends positively on the output gap (and also on the past inflation and the expected future inflation), signalling demand pressures. For this reason, it is sometimes called the 'inflation-adjustment (IA) line' (e.g. Romer 2000; Taylor 2000). Equation (4.2) can be considered as the equivalent of the NAIRU principle (e.g. Lavoie 2006, p. 169): the inflation rate accelerates whenever the actual (growth rate of) demand and output exceeds the natural (rate of growth of) output. Equation (4.3) is the monetary policy rule or the reaction function of the central bank. It incorporates the well-known 'Taylor rule' (e.g. Taylor 1993, 1999), according to which changes in the nominal interest rate set by the central bank are positive function of the 'natural' real interest rate, the expected future inflation rate, the past output gap and the past inflation gap (i.e. the deviation of the actual inflation in previous period from its target value). In formal terms, it is usually derived from the minimisation of the 'loss function' of the central bank, where the losses for each period are a weighted average in quadratic terms of the deviation of inflation from its target rate, and of current output relative to its potential level (Woodford 2003, p. 381). Since prices are supposed to be sticky in the short run, and changes in expected inflation are taken into account, when steering the nominal rate, the central bank is effectively setting the real interest rate (Romer 2000, p. 155).

Two points are worthy to mention here. First, the interest rate policy rule replaces the traditional *LM* curve in the *IS-LM-AS* model, along with its assumption that the central bank targets the money supply. In the NCM the central bank is able to influence the short-run real interest rate, and money is a residual (Meyer 2001). Second, the short-run stickiness of prices also explains the limited effectiveness of monetary policy. In the long run, prices are flexible, and hence the central bank is unable to influence the real interest rate. Therefore, monetary policy affects real variables and inflation in the short run, but is neutral in the long run. Finally, notice that combining Eq. (4.1) with Eq. (4.3) gives a negatively sloped relationship between inflation and output gap, which represents the aggregate demand function of the model (see, among others, Romer 2000; Taylor 2000; Fontana and Setterfield 2009).

The closure of the model (4.1)–(4.3) requires the specification of the nature of expectations, that is, of the form of the set of functions $E(\cdot)$. In this regard, NCM authors admit that expected values of inflation and output may deviate from actual values in the short run. This discrepancy, in turn, may temporarily push the economic system out of its natural equilibrium state (or natural growth path). Consequently, there is some room for public intervention in the short run, though mainly through the steering of the target interest rate, in order to anchor inflation expectations. By contrast, forecasts could not be systematically wrong over time. The rational expectations hypothesis, that is, the assumption that agents know the right economic model and can use all information efficiently, remains the first theoretical pillar of the NCM. Exogenous non-systematic shocks may affect the equilibrium in the long run: in Eqs. (4.1), (4.2), and (4.3), this random component is 'captured' by ε_i (with

i = 1, 2, 3). But, apart from this, every systematic economic policy is doomed to leave real magnitudes, notably output and employment rate, unchanged. For instance, the only long-run effect of a long-lasting expansive fiscal stimulus would be an increase in inflation and (both nominal and real) interest rates (Fontana 2009b, c). This result is the NCM equivalent of the well-known neoclassical principle of the long-run neutrality of demand-led macroeconomic policies (Fontana 2011). In addition to rational expectations, the other theoretical pillar of the NCM is the notion of a natural (or long-run or trend) equilibrium, namely, the state towards which a fully competitive economy would tend in the long run, when the inflation expectations of agents are utterly fulfilled. In the natural equilibrium state, output and employment levels are determined by three factors: (i) the quantity of labour-force and capital (i.e. the stock of resources), (ii) the system of preferences of individual agents (i.e. the utility function of consumers or households), and (iii) the available technology (i.e. the production function of firms).

The mechanics of the NCM model follows from the theoretical pillars discussed above. A departure of output from its natural level (or natural growth rate) causes inflation to change, which in turn leads the central bank to move the short-run nominal interest rate, and given the stickiness of price, the short-run real interest rate, such that to bring current output back to its normal level.[3] This is the so-called nominal-anchor function of monetary policy (Allsopp and Vines 2000, p. 11). The institutional structure of the economy, including prevailing conditions on the labour market, is sometimes considered, but the natural or potential level of output is always independent of aggregate demand changes, including fiscal and monetary policy led changes.

However, the two theoretical pillars of the NCM modern, namely, the rational expectations hypothesis and the notion of a natural (or long-run or trend) equilibrium, are problematic (Hargreaves-Heap 1980). Real-world economies are essentially non-ergodic and path-dependent systems (Davidson 1978; Hanngsen 2006). Economic variables do not progress steadily towards an exogenously given unique and stable equilibrium. They can reach several (suboptimal) equilibria, and each of the equilibria achieved depends on past values. On the whole, it is not clear how the natural equilibrium would be reached in the long run. The achievement of such an optimal position is simply postulated. In order to clarify this point, a simplified version of the previous three-equation NCM model is presented below:

$$Y_t = \alpha_0 - \alpha_1 \left(r_{t-1} - \pi_{t-1} \right) + \varepsilon_1 \quad (4.4)$$

$$\pi_t = \pi_{t-1} + \beta_1 \left(Y_{t-1} - Y^n_{[t-1]} \right) + \varepsilon_2 \quad (4.5)$$

$$r_t = \pi_t + RR^*_t + \gamma_1 \left(\pi_{t-1} - \pi^T \right) + \gamma_2 \left[Y_t - E\left(Y^n_{[t+1]} \right) \right] + \varepsilon_3 \quad (4.6)$$

where α_0, α_1, β_1, γ_1, $\gamma_2 > 0$. The main difference with the previous model is that Eq. (4.4) now determines the current value (or growth rate) of output, Y_t, instead of its gap with the natural level (or growth rate), $Y^n_{[t]}$ (where square brackets show that, in principle, natural output is independent of current conditions). In addition, for the sake of simplicity, Eqs. (4.4) and (4.5) are assumed not to be forward-looking. The variable RR^* in Eq. (4.6) is the real rate of interest assuring the ex ante matching of savings and investment at the natural level of output. It corresponds to the Wicksellian natural rate of interest (Fontana 2007) and can be derived by substituting Eq. (4.4) in Eq. (4.6). Then, by imposing that the actual inflation rate equals the target rate, and that the output gap is nil, it follows:

$$RR^*_t = \frac{\left(\alpha_0 - Y^n_t \right)}{\alpha_1} \quad (4.7)$$

If the central bank sets the value of RR^*_t in accordance with Eq. (4.7), then the economy reaches its natural equilibrium, and the system (4.4)-(4.5)-(4.6)-(4.7) behaves like the system (4.1)-(4.2)-(4.3). Yet, the assumption that the level (or growth rate) of potential output is an exogenous variable has been criticised by several authors. Labour productivity (e.g. the impact of learning by doing of workers, technological innovations and investment in fixed capital) and the availability of labour force (e.g. migration flows) are strictly linked to the current level of demand

and output (Setterfield 2002; León-Ledesma and Thirlwall 2002; Lavoie 2006; Fontana and Palacio Vera 2007; McCombie and Pike 2013; Sawyer 2013). All these factors affect the future potential output of the economy. Following Lavoie (2006, p. 182), the reduced-form NCM model (4.4)-(4.5)-(4.6)-(4.7) should, therefore, be amended by introducing an additional equation:

$$Y_t^n = Y_{t-1}^n + \phi\left(Y_{t-1} - Y_{t-1}^n\right) + \varepsilon_4 \quad (4.8)$$

where $0 < \phi \le 1$.

Equation (4.8) means that the short-run level of output affects the long-run potential or natural level of output (Lavoie 2006, p. 181; see also Flaschel 2000; Fontana 2010). In other words, Eq. (4.8) allows for hysteresis effects to be introduced into the benchmark NCM model, in this way allowing for the interdependence between the aggregate demand for and the aggregate supply of goods and services. For this reason, the reduced-form NCM model (4.4)-(4.5)-(4.6)-(4.7)-(4.8) is labelled the augmented NCM model in the rest of this chapter.

3 Adding Banks and Financial Intermediaries to the NCM Model

In the aftermath of the 2007–2008 financial crisis, several scholars argued that the NCM model is not fit for modern economies. It does not capture fundamental aspects of the working of financially sophisticated capitalist economies, including the possibility of financial turmoil, financial and banking crises and related prolonged recessions (e.g. Foley and Farmer 2009; Krugman 2009; Buiter 2009; Spaventa 2009). Lucas (2009) seems to agree with this view. He maintains that the 2007–2008 financial crisis was not predicted because such events cannot be predicted by NCM model (and related DSGE models alike): simulations based on the NCM model are not an 'assurance that no crisis would occur, but […] a forecast of what could be expected conditional on a crisis not occurring' (Lucas 2009). In this regard, one of the main theoretical issues, with significant practical consequences (see, e.g. Allington et al. 2012), is that the benchmark NCM model relies on both the 'efficient market hypothesis' (EMH hereafter) and the 'Modigliani-Miller theorem' (MMT hereafter), in the medium to long run at least (Veronese Passarella 2014). According to the EMH, prices of traded assets always reflect all available information, while the MMT maintains that, under a number of restrictive assumptions, the value of a firm is unaffected by how it is financed. As a result, given enough long time, money and finance would not affect output and employment, but only inflation and nominal interest rates. This again is not surprising: if an autonomous investment function of firms is ruled out of the model, then conditions of financing of investment (and current production) cannot, by definition, influence real variables.

The explicit analysis of the possible interaction between the real economy and the prevailing conditions in the banking and financial sectors is the core feature of the 'financial accelerator mechanism' (FAM) literature, originally developed by Bernanke, Gertler and Gilchrist during the early 1980s (e.g. Bernanke 1981, 1983; Bernanke and Gertler 1989; see also Bernanke et al. 1996, 1999). The FAM literature recognises that firms need external finance in order to realise their investment projects. Furthermore, it brings to light the informational asymmetries between lenders (i.e. banks and financial intermediaries) and borrowers (i.e. firms). On this basis, then it analyses the process by which negative shocks to the real sector of an economy are amplified by the workings of the banking and financial sectors.

The FAM literature introduces several innovative aspects into the mainstream macroeconomic debate. First, the informational asymmetries between lenders and borrowers make both the EMH and the MMT so cherished by NCM authors inapplicable. Second, these informational asymmetries mean that lenders have little information about the reliability of borrowers. Lenders face conventional agency costs, including monitoring costs and potential bankruptcy risks, which in turn translate into a premium for firms of the cost of external finance vis-à-vis internal finance. Third, in the face of informational asymmetries, banks and financial intermediaries assess the ability of repaying loans by using the market value of the net worth of firms, that is, the collateralised assets of firms.

Two important implications follow from these theoretical innovations. First, the net worth, and hence the ability to borrow of firms moves procyclically (e.g. Bernanke and Gertler 1989). An increase in asset prices and cash flows raises the net worth of firms, and reduces the premium of external finance on internal finance. This in turn boosts investment, aggregate demand and economic activity, which have then positive feedbacks on the net worth of firms and so on. Similarly, a fall in assets prices triggers a vicious self-reinforcing cycle. A reduction in the net worth of firms leads banks and financial intermediaries to tighten financing conditions. This reduces the ability of firms to borrow and finance investment. Economic activity falls, which then further reduces assets prices and the net worth of firms, and so on. This is the core of the FAM. An initial shock to the economy, however small it is, is likely to be amplified by changes in the balance sheets of firms and, more generally, by conditions in the banking and financial sectors. Second, the dynamics of the FAM is intrinsically nonlinear, since it depends on both the current level of internal finance of firms, and the general conditions of the economy. For instance, the more an economy is in a deep recession, the less likely would be the availability of external and internal finance, and hence the stronger will be the autoregressive movement in demand (e.g. Bernanke and Gertler 1989, pp. 14–15; Bernanke et al. 1996, pp. 3–4). This, in turn, will produce dramatic effects for firms. They will be accumulating excess inventories, while reducing the employment level and/or real wages bargained with workers (e.g. Greenwald and Stiglitz 1993, p. 109).

It is worthy to note that references to an exogenously given natural level or rate of growth of output are rare in the FAM literature. On the one hand, it is clearly stated that the methodological starting point of the FAM model is the benchmark NCM model. On the other hand, FAM scholars ignore long-run financial relationships in their works (e.g. Bernanke and Gertler 1989, p. 15). In other words, price flexibility is no longer regarded as the natural or long-run condition of the system, but just as the limiting case—as Bernanke et al. (1999, p. 6) call it. The long run is regarded as an ideal path, rather than as the historical tendency of capitalist economies. But then, if the relationship between price stickiness and price flexibility is reversed, with the latter being the exception rather than the norm, short-run sub-optimal equilibria become the rule, and so it does public intervention. This controversial interpretation of the FAM literature is supported by the repeated reference to the debt-deflation theory of Fisher (1933), and also by mention of the work of Minsky and Kalecki (e.g. Bernanke et al. 1999; Bernanke 1983, which quotes Minsky 1977). In fact, the discussion by FAM scholars of agency costs resonate the Minskian 'objectivation' of the lender risk into interest rates, fees and commissions that firms have to pay in order to access external financing (e.g. Minsky 1986). In this regard, another interesting feature of FAM models is the assumed heterogeneity of agents. As Bernanke et al. (1996) explain these models 'step outside the convenient representative-agent paradigm … [since] the distribution of wealth affects the dynamics of the economy in a nontrivial way' (pp. 3–4).

According to FAM scholars, during economic recessions the reallocation of bank lending from firms whose net worth is decreasing to more solvent firms triggers a 'flight-to-quality' (or 'flight-to-safety') process. This, in turn, increases the financial fragility of a country. Against this background, it is argued that large corporations are likely to be less hit by the greater cost (or difficulty) in obtaining credit in downturns compared to small firms. FAM scholars then conclude that 'recessions that follow a tightening of monetary policy are perhaps most likely to involve a flight to quality, because of the adverse effect of increased interest rates on balance sheets and because of monetary tightening may reduce flows of credit through the banking system' (Bernanke et al. 1996, p. 6; see also Bernanke and Blinder 1988). To put it differently, FAM scholars seem to argue that monetary policy affects output and other real magnitudes not so much because prices are sticky, as it is assumed in the benchmark NCM model, but rather because it affects the price and access to external finance, which has a crucial impact on the investment and production plans of firms.

As far as the formal modelling is concerned, the benchmark FAM model is usually obtained through a process of microeconomic foundation of the macroeconomic dynamics. This is done by considering a production (or investment) technology that involves asymmetric information between firms, who have direct access to the technology, and banks and financial intermediaries, who have not. In addition, it is assumed that banks and financial intermediaries incur agency costs in order to observe the investment returns of firms. These costs are in turn assumed to be a decreasing function of the soundness of the balance-sheet of borrowers, that is, the net wealth of firms. Finally, since the latter is likely to move pro-cyclically, agency costs will behave counter-cyclically, therefore improving lending conditions in booms and deteriorating them in recessions. In this way, the accelerator (macroeconomic) effect of income on investment is brought back to a simple (microeconomic) principal-agent problem (Bernanke et al. 1996, p. 27).

The simplest way to include the FAM mechanism within the benchmark NCM model discussed in the previous section is to replace Eq. (4.1) with the following:

$$Y_t^g = a_0 + a_1 Y_{t-1}^g + a_2 E\left(Y_{t+1}^g\right) - a_3\left[r_t - E\left(\pi_{t+1}\right)\right] + a_4 H_{t-1} + \varepsilon_1 \tag{4.9}$$

with:

$$H_t = H_{t-1} + \omega Y_t^g + \varepsilon_4 \tag{4.10}$$

where H_t is the net worth of investing firms, $0 < \omega < 1$ is the share of aggregate (retained) profits and capital gains in total output (gap) and $\alpha_4 > 0$ is the sensitivity of total output gap to changes in the creditworthiness of firms, through changes in the finance available for investment. The basic idea underpinning Eqs. (4.9) and (4.10) is that investment, and hence current output, is crucially affected by the financial soundness of the consolidated balance-sheet of firms. More precisely, the lower (higher) the amount of internal funds accumulated by firms over the previous periods, the lower (higher) will be current investment and output. It is worthy to note that changes in internal funds can affect production decisions both through the self-financing of investment (direct channel) and through the degree of creditworthiness of firms used by banks and financial intermediaries (indirect channel). Whatever the prevalent channel, the result is a strengthening and extension of the short-run effects of aggregate demand on output and employment levels.

Table 4.1 Four different mainstream macroeconomic models

	Without finance	With finance (accelerator)
Temporary effect of demand	(I) Benchmark NCM	(III) Benchmark FAM
Permanent effect of demand (hysteresis)	(II) Augmented NCM	(IV) Augmented FAM

Source: Authors' construction

Table 4.1 presents the four different mainstream macroeconomic models discussed in this chapter, notably the benchmark NCM model (I), the augmented NCM model (II), the benchmark FAM model (III) and the augmented FAM model (IV). Models (I)–III) have been examined above, while model (IV) is a modified version of model (II). It takes into account the cumulative effects on investment of changes in the market value of the net worth of firms, as it occurs in model (III). Yet, unlike model (III), model (IV) does not involve any exogenously given natural level of output towards which the economy is assumed to move. In algebraic terms, it is derived by replacing Eq. (4.4) of model (II) with Eq. (4.11):

$$Y_t = \alpha_0 - \alpha_1 \left(r_{t-1} - \pi_{t-1} \right) + \alpha_2 H_{t-1} + \varepsilon_1 \tag{4.11}$$

where $\alpha_2 > 0$, while Eq. (4.10) can be rewritten as:

$$H_t = H_{t-1} + \omega \left(Y_t - Y_t^n \right) + \varepsilon_4 \tag{4.12}$$

Consequently, the interest rate rule defined by Eq. (4.7) must be replaced by Eq. (4.13):

$$RR_t^* = \left(\alpha_0 - Y_t^n + \alpha_2 H_{t-2} \right) / \alpha_1 \tag{4.13}$$

The model determined by the system of equations (4.11)-(4.5)-(4.6)-(4.13)-(4.8)-(4.12) is a synthesis of models (II) and (III): like in model (III) changing conditions in the banking and financial sectors amplify real shocks and can trigger booms and recessions. In addition, like in model (II), long-run levels of output and employment are affected via hysteresis effects by the current level of demand. This second feature is what distinguishes it from the benchmark FAM model.

Interestingly, FAM scholars acknowledge that the financial accelerator introduces a long-lasting (though not ever-lasting) hysteresis effect of aggregate demand into the benchmark NCM model. In the absence of information asymmetries—it is argued—investment demand can be safely assumed to be fixed over time, in the first approximation at least. By contrast, 'when information asymmetries are present, investment demand will vary and be history-dependent' (Bernanke and Gertler 1989, p. 20). This effect has important policy implications, namely, that one of main goals of central banks should be to strengthen the balance-sheets of economic agents, through the stabilisation of financial asset (viz. collateral) markets. This policy implication of the FAM is explored in great details in the next section.

4 The Current State of Macroeconomics

The repeatedly wrong predictions, and especially the failure in providing a satisfactory explanation of the 2007–2008 US crisis and the subsequent global financial crisis and economic recession, have represented a serious blow for the reputation of the NCM. There have been two main reactions to this in the economic discipline. Some scholars have argued that the proclaimed consensus around the benchmark NCM model was short-lived and finally unsuccessful (e.g. Buiter 2009). Other scholars have accepted the shortcomings of their original macroeconomic analyses and tried to amend the NCM model. As argued by McCombie and Pike (2013), the analytical core of the NCM model is in fact still 'seen by many to be relatively unscathed (but with the imperative to build in assumptions that allow for debt default and bankruptcy)' (p. 521). To be fair, attempts to make the benchmark NCM model more realistic were made before the onset of the 2007–2008 financial crisis. The most popular way was to modify the benchmark NCM model to allow for the possibility that a fraction of households or consumers cannot access financial markets. As these non-Ricardian consumers cannot borrow or save to smooth consumption, they follow a simple 'rule of thumb', namely, they always spend all current labour income on current consumption. Galì et al. (2004) showed that 'if the weight of such rule-of-thumb consumers is large enough, a Taylor-type rule must imply a (permanent) change in the nominal interest rate in response to a (permanent) change in inflation that is significantly above unity, in order to guarantee the uniqueness of equilibrium. Hence, the Taylor principle becomes too weak a criterion for stability when the share of rule-of-thumb consumers is large' (p. 740). Furthermore, the presence of non-Ricardian consumers is proved to affect significantly the reaction of an economy to fiscal policy shocks. For instance, an increase (decrease) in government spending entails now a remarkable increase (decrease) in output, in the short to medium run at least. This conclusion has been further strengthened by recent work indicating that the actual size of the multiplier of government spending is larger than one, either when the zero-lower bound on the nominal interest rate binds or the nominal interest rate is constant (e.g. Christiano et al. 2009).

In the aftermath of the financial crisis, attempts to improve or update the benchmark NCM model have multiplied. There have been two main targets of the original benchmark NCM model. First, scholars have focused their efforts on systematic errors in inflation forecasting. Second, and related to the previous point, scholars have tried to model financial markets and financial frictions. Starting with the former, the overestimation of deflationary effects of the financial crisis in the benchmark NCM model has been usually regarded as the consequence of the underestimation of price stickiness, which is captured by the so-called Calvo parameter in the accelerationist Phillips curve (Calvo 1983), namely, Eq. (4.2). The underestimation of the degree of price rigidity has, in turn, been explained by the lack of financial frictions in the benchmark NCM model. Once these frictions are introduced, it is argued that the NCM model accurately predicts the behaviour of the US economy since 2008, including the weak drop in inflation rate. Intuitively, the rationale is that financial frictions make the Phillips curve 'flatter', that is, they reduce the parameter b_1 in Eq. (4.2), or the parameter β_1 in Eq. (4.5), presented above. The

US crisis could therefore be interpreted and modelled as the result of aggregate demand shocks in the presence of a flat aggregate supply curve (e.g. Del Negro et al. 2014, pp. 19–21).

As far as the explicit modelling of financial markets and financial frictions is concerned, some NCM scholars have explored the effects of volatile risk premia, by assuming that fluctuations in these premia are the most important shocks driving the business cycle. This insight closely follows the work of Bernanke and Gertler (1989) and Bernanke et al. (1999). It represents an attempt of developing the benchmark FAM model, where the major difference between old and new models is mainly the accuracy of theoretical modelling and econometric techniques. In this regard, a fundamental contribution has been provided by Christiano et al. (2013), who assume that firms combine internal funds with external funds, namely, bank loans, to acquire raw (physical) capital, and that the interest rate on loans includes a 'premium' covering the costs of default of firms. The production of goods and services is then likened to a process in which firms convert raw capital into effective capital under 'idiosyncratic uncertainty' or 'risk'. Christiano et al. (2013) show that increases in risk premia raise the premium charged by banks, and reduce the supply of loans. In this way, they argue that increases in risk premia could account for some key features of the 2007–2008 financial crisis and related economic recession:

> With fewer financial resources, entrepreneurs acquire less physical capital. Because investment is a key input in the production of capital, it follows that investment falls. With this decline in the purchase of goods, output, consumption and employment fall. For the reasons stressed in [Bernanke et al. 1999], the net worth of entrepreneurs – an object that we identify with the stock market – falls too. This occurs because the rental income of entrepreneurs falls with the decline in economic activity and because they suffer capital losses as the price of capital drops. Finally, the overall decline in economic activity results in a decline in the marginal cost of production and thus a decline in inflation. So, according to the model the risk shock implies a countercyclical credit spread and procyclical investment, consumption, employment, inflation, stock market and credit. These implications of the model correspond well to the analogous features of US business cycle data. (Christiano et al. 2013, p. 2)

In short, fluctuations in risk premia over the risk-free interest rate should be regarded as the main trigger (or amplifier) of the business cycle. Once this mechanism is introduced in the benchmark NCM model, this is shown to accurately reproduce US cyclical fluctuations since the mid-1970s (see, also, Gilchrist et al. 2009; Merola 2013). These results echo early work by Borio et al. (2001) and Borio (2006), which explored the effects of changes in the absolute level of financial risks over time. Borio and his colleagues show that, when incentives and potential mismeasurements by financial market participants are allowed, the underestimation of risks in booms and the overestimation in recessions become a realistic possibility. This has deleterious effects on bank provisions and capital ratios. In turn, this strengthens the pro-cyclicality of bank profits, thereby encouraging banks to increase lending in booms and to reduce it in recessions. Alternative recent ways of modelling of financial markets and financial frictions include the introduction of collateral constraints, currency risk premia in open economies, and Minsky-Fisher type of mechanisms (see, for useful surveys, Brunnermeier et al. 2012; Roger and Vlcek 2012). Other models have been obtained through the explicit inclusion of a heterogeneous, monopolistically competitive banking sector (e.g. Hafstead and Smith 2012). In summary, all recent attempts to improve or update the benchmark NCM model have tried in a way or another to model financial markets and financial frictions, and in this way they represent varieties of model (IV) presented in Table 4.1, namely, augmented FAM models.

The different augmented FAM models discussed above represent recent attempts by mainstream macroeconomists to improve or update the benchmark NCM model. For all interesting properties, these models share two problems, namely, a theoretical weakness and a policy inconsistent problem that seem to hinder further progress in mainstream macroeconomic theory and policy-making. Starting with the former, augmented FAM models assume that financial instability and long-lasting slumps are the result of exogenous market frictions, that is, imperfections, asymmetries or rigidities in the banking and financial sectors. They never allow for the possibility that financial instability is the endogenous by-product of the normal functioning of modern economies. In other words, augmented FAM models still assume like in the old benchmark NCM model that in the long run free market forces would drive the economy towards a unique exogenously given and socially optimal equilibrium. It was this hypothesis of a natural equilibrium, coupled with the REH, which had left early NCM scholars with no other choice but the adoption of ad hoc assumptions about the stickiness of prices to fit real-world data. As explained by De Grauwe (2010), '[w]hy is it that in a world where everybody understands the model and each other's rationality, agents would not want to go immediately to the optimal plan using the optimal price? […] Calvo pricing is an *ad hoc* assumption forced unto the model to create enough inertia so that it would fit the data better' (pp. 416–17). A similar consideration could be made against modern macroeconomists attempting to update the benchmark NCM model by modelling conditions in the banking and financial sectors, including the possibility of financial instability, via ad hoc assumptions about exogenous market frictions. This critical stance has in fact led some behavioural economists to explore the effects of different heuristics on the financial behaviour of agents (e.g. De Grauwe 2010), while other economists are experimenting with alternative macroeconomic modelling, including the 'stock-flow consistent' approach of Godley and Lavoie (2007; see also among others, Dos Santos 2006; van Treeck 2009; Caverzasi and Godin 2015; Greenwood-Nimmo 2014; Sawyer and Veronese Passarella 2017; Nikiforos and Zezza 2017).

Notwithstanding the theoretical weakness of old and new FAM models discussed above, these latter models lead to a different rule of central banking vis-à-vis the benchmark NCM model. This is the policy inconstant problem. Although seldom pointed out, this policy implication of FAM models should not be underestimated. The point is that once it is admitted that lending is driven by the creditworthiness of borrowers, and thereby by the soundness of their balance-sheets, it turns out that the stabilisation of the market value of financial assets, especially those used as collaterals by firms, should be the priority of the central bank. The policy implication of FAM models would also highlight another real-world feature of the last couple of decades, namely, that the vast majority of refinancing operations in the interbanking market are conducted through REPOs, with government bonds acting as collaterals. But, if this is the case, then the support of government bonds, and not price stability, should be regarded as the overriding concern of central banks, at least during periods of recessions and economic stagnation. Notice that replacing risky private assets with low-risk government bonds guaranteed by central banks would further strengthen the soundness of the balance-sheets of firms, thereby contributing to smooth the business cycle (e.g. Fontana et al. 2017). This is the 'portfolio effect' pointed out by Minsky (1986) and recently rediscovered by Eggertsson and Krugman (2012, p. 1471).

In short, as a result of the 2007–2008 financial crisis and the failure of the NCM benchmark model to explain it, let alone to predict it, many attempts have been made to amend mainstream macroeconomics. The introduction of volatile risk premia, collateral constrains, currency risk premia in open economies and Minsky-Fisher type of mechanisms are the most innovative financial frictions used to improve the long-established NCM model. A severe limitation of these models is that they never allow for the possibility that financial instability is the endogenous by-product of the normal functioning of modern economies. Notwithstanding this limitation, the introduction of financial frictions in the NCM model highlights the valuable role that central banks could play in stabilising the market value of financial assets, especially those used as collaterals by firms.

5 Summary and Conclusions

Early in the 1990s, a convergence of view emerged in mainstream macroeconomics. The NCM model quickly spread among academics and policymakers alike. The 2007–2008 financial crisis, resulting recession, and the current stagnation period have highlighted a problematic feature of the model. Banks and financial intermediaries, which have played a vital role in the start and unfolding of these dramatic real-world events, are not mentioned, let alone modelled in the NCM. During the last decade, several attempts have been made to improve and update the NCM model by adding to it a role for banks and financial intermediaries. This chapter has offered a preliminary and critical assessment of these efforts.

The chapter has started with a discussion of the set of three equations describing the benchmark NCM model in a closed economy, and has highlighted the role that the REH and the notion of 'natural equilibrium' play into it. It has also discussed an amended version of the model, the augmented NCM model, which allows for the possibility of interdependence between aggregate demand and aggregate supply, such that it could capture some prominent features of real-world economies. The chapter has also reviewed the original contributions to the FAM made by Bernanke and his colleagues in the early 1980s. These contributions have recently been rediscovered by scholars aiming to assign a greater role to banks and financial intermediaries in mainstream macroeconomics. The original benchmark FAM model together with the recent augmented FAM model has been discussed at great length in order to highlight the nature and role of financial instability in these models. The main conclusion of the chapter is that for all good intentions, there are two still main problems that seem to hinder progress in mainstream macroeconomic theory and policy-making. First, the policy implications of the recent theoretical innovations have not been fully explored. The augmented FAM suggests replacing price stability with financial stability as the main goal of central banks. Second, and more importantly, in the most recent mainstream macroeconomic models, financial instability is still modelled as the outcome of exogenous market frictions, rather than the endogenous by-product of the normal functioning of modern economies.

Notes

1. The authors acknowledge the support from the project Financialisation, Economy, Society and Sustainable Development (FESSUD), which has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant Agreement No. 266800. The authors are grateful to Philip Arestis and Malcolm Sawyer for useful comments and suggestions on an early draft of this chapter. The usual disclaimers apply.
2. The values of the 'deep parameters' of the NCM and related DSGE models, i.e. the parameters which are supposed not to be affected by policy, are

usually obtained through either 'calibration' methods or Bayesian estimation econometric techniques (Tovar 2009).

3. For a critical assessment of the monetary policy rules in the NCM, see Fontana and Palacio-Vera (2002), Brancaccio and Fontana (2013). See also Allington and McCombie (2005), for an analysis of the role of stock market prices in monetary policy rules.

References

Allington, N., & McCombie, J. S. L. (2005). Stock market prices and the conduct of monetary policy under the new consensus monetary policy. In P. Arestis, M. C. Baddeley, & J. S. L. McCombie (Eds.), *The new monetary policy. Implications and relevance*. Cheltenham: Edward Elgar.

Allington, N., McCombie, J. S. L., & Pike, M. (2012). Lessons not learned: From the collapse of long-term capital management to the subprime crisis. *Journal of Post Keynesian Economics, 34*(4), 555–582.

Allsopp, C., & Vines, D. (2000). The assessment: Macroeconomic policy. *Oxford Review of Economic Policy, 16*(4), 1–32.

Arestis, P. (2007). What is the new consensus in macroeconomics? In P. Arestis (Ed.), *Is there a new consensus in macroeconomics?* Houndmills: Palgrave Macmillan.

Arestis, P. (2009). New consensus macroeconomics: A critical appraisal. In G. Fontana & M. Setterfield (Eds.), *Macroeconomic theory and macroeconomic pedagogy*. Basingstoke: Palgrave Macmillan.

Arestis, P., & Sawyer, M. (2004). Can monetary policy affect the real economy? *European Review of Economics and Finance, 3*(3), 9–32.

Arestis, P., & Sawyer, M. (2006). The nature and role of monetary policy when money is endogenous. *Cambridge Journal of Economics, 30*(6), 847–860.

Arestis, P., & Sawyer, M. (2008). A critical reconsideration of the foundations of monetary policy in the new consensus macroeconomics framework. *Cambridge Journal of Economics, 32*(5), 761–779.

Bernanke, B. S. (1981). Bankruptcy, liquidity and recession. *American Economic Review Proceedings, 71*(2), 155–159.

Bernanke, B. S. (1983). Nonmonetary effects of the financial crisis in the propagation of the great depression. *American Economic Review, 73*(3), 257–276.

Bernanke, B. S., & Blinder, A. S. (1988). Credit, money, and aggregate demand. *American Economic Review*, Papers and Proceedings of the One-Hundredth Annual Meeting of AEA, *78*(2), 435–439.

Bernanke, B. S., & Gertler, M. (1989). Agency costs, net worth and business fluctuations. *American Economic Review, 79*(1), 14–31.

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1996). The financial accelerator and flight to quality. *Review of Economics and Statistics, 78*(1), 1–15.

Bernanke, B. S., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1). Amsterdam: Elsevier.

Borio, C. (2006, September). *Monetary and prudential policies at a crossroads? New challenges in the new century* (BIS working papers, no. 216).

Borio, C., Furfine, C., & Lowe, P. (2001, March). *Procyclicality of the financial system and financial stability: Issues and policy options* (BIS working papers, no. 1).

Brancaccio, E., & Fontana, G. (2013). Solvency rule versus Taylor rule. An alternative interpretation of the relation between monetary policy and the economic crisis. *Cambridge Journal of Economics, 37*(1), 17–33.

Brunnermeier, M. K., Eisenbach, T. M., & Sannikov, Y. (2012). *Macroeconomics with financial frictions: A survey* (NBER working paper series, no. 18102).

Buiter, W. (2009, March 3). The unfortunate uselessness of most 'state of the art' academic monetary economics. *Financial Times*.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics, 12*(3), 383–398.

Caverzasi, E., & Godin, A. (2015). Post-Keynesian stock-flow-consistent modelling: A survey. *Cambridge Journal of Economics, 39*(1), 157–187.

Christiano, L., Eichenbaum, M., & Rebelo, S. (2009). *When is the government spending multiplier large?* (NBER working paper series, no. 15394).

Christiano, L., Motto, R., & Rostagno, M. (2013). *Risk shocks* (NBER working paper series, no. 18682).

Clarida, R., Galì, J., & Gertler, M. (1999). The science of monetary policy: A new Keynesian perspective. *Journal of Economic Literature, 37*(4), 1661–1707.

Davidson, P. (1978). *Money and the real world*. London: Macmillan.

De Grauwe, P. (2010). The scientific foundation of dynamic stochastic general equilibrium (DSGE) models. *Public Choice, 144*(3–4), 413–443.

Del Negro, M., Giannoni, M. P., & Schorfheide, F. (2014). Inflation in the Great Recession and New Keynesian models. *Federal Reserve Bank of New York: Staff Reports*, No. 618 (May 2013; revised January 2014).

Dixon, H. D. (2008). New Keynesian macroeconomics. In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). Basingstoke/New York: Palgrave Macmillan.

Dos Santos, C. H. (2006). Keynesian theorising during hard times: Stock-flow consistent models as an unexplored 'frontier' of Keynesian macroeconomics. *Cambridge Journal of Economics, 30*(4), 541–565.

Eggertsson, G. B., & Krugman, P. (2012). Debt, deleveraging, and the liquidity trap. *Quarterly Journal of Economics, 127*(3), 1469–1513.

Fisher, I. (1933). The debt-deflationary theory of great depressions. *Econometrica, 1*(4), 333–357.

Flaschel, P. (2000). Keynes-Marx and Keynes-Wicksell models of monetary growth: A framework for future analysis. *Review of Political Economy, 12*(4), 453–468.

Foley, D., & Farmer, J. D. (2009). The economy needs agent-based modelling. *Nature, 460*, 685–686.

Fontana, G. (2007). Why money matters: Wicksell, Keynes, and the new consensus view on monetary policy. *Journal of Post Keynesian Economics, 30*(1), 43–60.

Fontana, G. (2009a). *Money, uncertainty and time*. London: Routledge.

Fontana, G. (2009b). Whither new consensus macroeconomics? The role of government and fiscal policy in modern macroeconomics. In E. Hein, T. Niechoj, & E. Stockhammer (Eds.), *Macroeconomic policies on shaky foundations – Whither Mainstream economics?* (pp. 187–208). Marburg, Metropolis Verlag. Republished in Wray, L. R. (Ed.). (2013). *Theories of money and banking, The international library of critical writings in economics series*. Cheltenham: Edward Elgar.

Fontana, G. (2009c). The transmission mechanism of fiscal policy: A critical assessment of current theories and empirical methodologies. *Journal of Post Keynesian Economics, 31*(4), 587–604.

Fontana, G. (2010). The return of Keynesian economics: A contribution in the spirit of John Cornwall's work. *Review of Political Economy, 22*(4), 517–533.

Fontana, G., & Palacio-Vera, A. (2002). Monetary policy rules: What are we learning? *Journal of Post Keynesian Economics, 24*(4), 547–568.

Fontana, G., & Palacio-Vera, A. (2007). Short-run output stabilisation all that monetary policy can aim for? *Metroeconomica, 58*(2), 269–298.

Fontana, G., & Setterfield, M. (2009). Macroeconomics, endogenous money and the contemporary financial crisis: A teaching model. *International Journal of Pluralism and Economic Education, 1*(1), 130–147.

Galì, J., López-Salido, J. D., & Vallés, J. (2004). Rule-of-thumb consumers and the design of interest rate rules. *Journal of Money, Credit, and Banking, 36*(4), 739–763.

Gilchrist, S., Ortiz, A., & Zakrajšek, E. (2009, June 4–5). *Credit risk and the macroeconomy: Evidence from an estimated DSGE model*. Paper prepared for the FRB/JMCB conference 'Financial Markets and Monetary Policy', held at the *Federal Reserve Board*, Washington, DC.

Godley, W., & Lavoie, M. (2007). *Monetary economics: An integrated approach to credit, money, income production and wealth*. Basingstoke: Palgrave Macmillan.

Goodfriend, M. (2004). Monetary policy in the new neoclassical synthesis: A primer. *Economic Quarterly, 90*(3), 21–45.

Goodfriend, M., & King, R. (1997). The new neoclassical synthesis and the role of monetary policy. In B. S. Bernanke & J. Rotemberg (Eds.), *NBER macroeconomics annual 1997* (Vol. 12). Cambridge, MA: MIT Press.

Goodhart, C. A. E. (2010). Money, credit and bank behaviour: Need for a new approach. *National Institute Economic Review, 214*, F73–F82.

Greenwald, B. C., & Stiglitz, J. E. (1993). Financial market imperfections and business cycles. *Quarterly Journal of Economics, 108*(1), 77–114.

Greenwood-Nimmo, M. (2014). Inflation targeting monetary and fiscal policies in a two-country stock-flow-consistent model. *Cambridge Journal of Economics, 38*(4), 839–867.

Hafstead, M., & Smith, J. (2012). Financial shocks, bank intermediation, and monetary policy in a DSGE Model. Available at: http://www.rff.org/Documents/HafsteadSmith_September2012.pdf

Hanngsen, G. (2006). The transmission mechanism of monetary policy: A critical review, in P. Arestis and M. Sawyer (eds.), A handbook of alternative monetary economics, Cheltenham/Northampton: Edward Elgar.

Hargreaves-Heap, S. P. (1980). Choosing the wrong natural rate: Accelerating inflation or decelerating employment and growth. *Economic Journal, 90*(1), 611–620.

Lavoie, M. (2006). A post-Keynesian amendment to the new consensus on monetary policy. *Metroeconomica, 57*(2), 165–192.

León-Ledesma, M. A., & Thirlwall, A. P. (2002). The endogeneity of the natural rate of growth. *Cambridge Journal of Economics, 26*(4), 441–459.

Lucas, R. (2009, August 6). In defence of the dismal science. *The Economist*.

McCombie, J. S. L., & Negru, I. (2014). On economic paradigms, rhetoric and the micro-foundations of macroeconomics. *European Journal of Economics and Economic Policies: Intervention, 11*(1), 53–66.

McCombie, J. S. L., & Pike, M. (2013). No end to the consensus in macroeconomic theory? A methodological inquiry. *American Journal of Economics and Sociology, 72*(2), 497–528.

Merola, S. (2013). *The role of financial frictions during the crisis: And estimated DSGE model* (Working paper research, no. 249). National Bank of Belgium.

Meyer, L. H. (2001). Does money matter? *Federal Reserve Bank of St. Louis Review, 83*(5), 1–15.

Minsky, H. P. (1977). A theory of systematic fragility. In E. I. Altman & A. W. Sametz (Eds.), *Financial crises*. New York: Wiley-Interscience.

Minsky, H. P. (1986). *Stabilizing an unstable economy*. New Haven: Yale University Press.

Nikiforos, M., & Zezza, G. (2017). *Stock-flow consistent macroeconomic models: A survey* (Levy Economics Institute of Bard College working paper, no. 891). Forthcoming in the *Journal of Economic Surveys*.

Palacio-Vera, A. (2009). Money wage rigidity, monopoly power and hysteresis. In P. Arestis & M. Sawyer (Eds.), *Path dependency and macroeconomics*. London: Palgrave Macmillan.

Roger, S., & Vlcek, J. (2012). *Macrofinancial modelling at central banks: Recent developments and future directions* (IMF working papers, no. 21).

Romer, D. (2000). Keynesian macroeconomics without the LM curve. *Journal of Economic Perspectives, 14*(2), 149–169.

Sawyer, M. (2013). Endogenous money, circuits and financialisation. *Review of Keynesian Economics, 1*(2), 230–241.

Sawyer, M., & Veronese Passarella, M. (2017). The monetary circuit in the age of financialisation: A stock-flow consistent model with a twofold banking sector. *Metroeconomica, 68*(2), 321–353.

Setterfield, M. (2002). Introduction: A dissenter's view of the development of growth theory and the importance of demand-led growth. In M. Setterfield (Ed.), *The economics of demand-led growth: Challenging the supply-side vision of the long run*. Cheltenham: Edward Elgar.

Setterfield, M. (2010). *Handbook of alternative theories of economic growth*. Cheltenham: Edward Elgar.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association, 1*(5), 1123–1175.

Smets, F., & Wouters, R. (2007, February). *Shocks and frictions in US business cycles. A Bayesian DSGE approach* (Working paper series, no. 722). European Central Bank.

Spaventa, L. (2009). Economists and economics: What does the crisis tell us? *Real-World Economics Review, 50*(September), 132–142.
Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*(1), 195–214.
Taylor, J. B. (1999). *Monetary policy rules*. Chicago: University of Chicago Press.
Taylor, J. B. (2000). Teaching modern macroeconomics at the principles level. *American Economic Review, 90*(2), 90–94.
Tovar, C. E. (2009). DSGE models and central banks. *Economics: Open Access/Open Assessment E-Journal*, *3*, 1–30, Available at: http://www.economics-ejournal.org
Van Treeck, T. (2009). A synthetic, stock-flow consistent macroeconomic model of financialisation. *Cambridge Journal of Economics, 33*(3), 467–493.
Veronese Passarella, M. (2014). Financialization and the monetary circuit: A macro-accounting approach. *Review of Political Economy, 26*(1), 107–127.
Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.
Woodford, M. (2009). Convergence in macroeconomics: Elements of the new synthesis. *American Journal of Economics: Macroeconomics*, *1*(1), 267–79 (PDF file downloadable from: http://www.columbia.edu/~mw2230/Convergence_AEJ.pdf. April 30th 2013).

5

Microeconomics, Mesoeconomics and Macroeconomics

Malcolm Sawyer

1 Introduction

I have long been puzzled over the accusation that macroeconomic analysis (of the Keynesian and Kaleckian forms) suffers from a lack of microeconomic foundations. The founders of macroeconomic analysis clearly provided microeconomic behaviour: Kalecki specifically in terms of pricing (based on the degree of monopoly) and investment, though there was a lack of explicit household behaviour with regard to consumption (most or all of wage income taken to be consumed) and labour (where the available labour force was treated as socially determined and little influenced by the level of real wages). Keynes provided an analysis of investment decisions, price setting and labour supply at the micro level. But, what Kalecki, Keynes and many others did not conform to was the acceptance of the dominance of a microeconomic analysis based on utility optimisation over a well-known future, and they focused on some essential macroeconomic relationships (in a way which is indicated below).

M. Sawyer (✉)
University of Leeds, Leeds, UK

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_5

For some, macroeconomic analysis sits uneasily with the idea of what economics covers. The well-known view of Lionel Robbins (1933) is that "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (p. 15). I do not intend to get into issues of whether economics is or could be a science (or what is meant by being a science). This definition focuses on human behaviour without mention of the interactions between individuals nor of how co-ordination between individuals is achieved. It says nothing about the overall levels of economic activity or of any macroeconomic relationships. As Joan Robinson (1972) observed, "It was just a coincidence that the book (Robbins 1933) appeared when means for any end at all had rarely been less scarce" (p. 1).

The American Economic Association cast the net a little, but not much, wider:

> Economics can actually be defined a few different ways: it's the study of scarcity, the study of how people use resources, or the study of decision-making. Economics often involves topics like wealth, finance, recessions, and banking, leading to the misconception that economics is all about money and the stock market….One of the central tenets of economics is that people want certain things and will change their behavior to get those things – in other words, people will respond to **incentives**. … Lower wages in another country provide an incentive for a factory to relocate overseas to cut down on costs. High taxes provide an incentive for people to look for ways to hide their income because they want to keep more of their money. …. [However] Economic study ranges from the very small to the very large. The study of choices by individuals (like how someone decides to budget their paycheck each month) is called **microeconomics**. … The study of governments, industries, central banking, and the boom and bust of the business cycle is called **macroeconomics.** (available at: https://www.aeaweb.org/resources/students/what-is-economics; bold in original) Macroeconomics is largely seen as relating to specific institutions (government, central bank).

It is necessary to locate macroeconomics within a much broader and inclusive perspective on economics—perhaps political economy would be a more appropriate term. Political economy would cover, inter alia, the generation and use of the surplus, the dynamics of capitalism, income distribution, growth and development. Within that perspective, the overall aim is to understand and analyse the workings of an economic system. Any system is based on how the individual components make decisions and seek to implement those decisions, and what motivates behaviour and decisions. At various levels within the system, individuals interact, and their interactions help to settle the outcomes. This may be undertaken at the level of a market, industry, and so on. The outcomes may be analysed in terms of a consistency analysis of some form; but at a minimum, there has to be mechanisms which in some sense reconcile the decisions of the individuals, even if that means some individuals not being able to fully implement their decisions. Individual behaviour is socially influenced and constrained. Individuals interact economically in many different ways, and notably through market interactions but also (and predominantly) within organisations and institutions—households, corporations, for example. Within organisations, economic (and other) activities are organised and co-ordinated, and economic power exercised. Economic analysis then involves investigations of individual behaviour, and (more importantly) the ways in which individuals interact and co-ordinate at what may be termed the meso level and the macro level.

This chapter has three main sections, in addition to the introductory and concluding sections. In Sect. 2, the focus is on what have been termed 'microeconomic foundations' of macroeconomics, and it presents a critique of that approach and indicates severe shortcomings. In Sect. 3, the nature of macroeconomic relationships is discussed—namely, the general (and obvious) proposition that there are macroeconomic conditions, which are not immediately derived from microeconomic considerations alone, but where consistency and sustainability considerations have to be brought in. In Sect. 4, some of the problems of undertaking macroeconomic analysis are considered. Section 5 summarises and concludes.

2 Microeconomic 'Foundations'

The relationships between microeconomics and macroeconomics are often discussed using the phrase 'microfoundations' of macroeconomics. As King (2012) remarks, "'Microfoundations' is a spatial analogy, taken from architecture, from the building trades or from constructional engineering. … Foundations have to come first, they must be solid and they must be reasonably extensive" (p. 22).

King's (2012) Chapter 2 has the title 'Microfoundations as a (bad) metaphor'. He argues that there are two essential reasons why the "microfoundations dogma" is "nearly all wrong" (p. 9). Also, "In sum, 'microfoundations' is a very bad metaphor, which has caused considerable confusion and has been used to justify some very bad decision by macroeconomic theorists" (King 2012, p. 26). These are the fallacy of composition and downward causation. "The fallacy of composition entails that an entire economy may behave in ways that cannot be inferred from the behaviour of its individual agents" (King 2012, p. 9). The best-known example relates to the 'paradox of thrift' in which it is argued that a decision by a single individual to seek to increase their savings may well lead higher savings by that individual, but a comparable statement does not hold for an increase in the overall level of savings. In the context of a given intended level of investment (in the context of a closed economy), and the consistency requirement that overall savings = overall investment, then overall savings would not increase. This conclusion is drawn within the context of a specific model, and one may query the workings of that model; for example, actual investment may differ from intended investment through inventory changes, or it may be argued that investment intentions alter in the face of changes in savings intentions.

"The principle of downward causation states that, in economics, causal processes operate in both directions, not only from the behaviour of individual agents to the behaviour of the entire economy, but also from the economy to the tastes, beliefs, expectations and actions of the individual agents" (King 2012, p. 9). This is a view, which I would share, but there is a line of argument, which can be further developed. There are many 'layers' within the overall economy, and it is often 'useful' to proceed through those 'layers'—engaging in what would often be regarded as 'partial equilibrium' analysis, though there is no presumption that equilibrium has to be involved. This could involve the grouping of individuals into a household and analysis of intra-household behaviour, the grouping of individuals within a set of employment relationships with corporations and other organisations, the ways in which producers and consumers interact within a market and an industry and so on. It has to be recognised, though, that these units of analysis, for example, a market and an industry can be constructs of economic theorising.

There has to be analysis of behaviour at the level of the individual, which could be labelled microeconomic. Using the term 'foundations' suggests essential building blocks, whereas behaviour and decision-making at the level of the individual is merely one component. Further, the term 'microfoundations' suggests that the direction of causation runs from the individual level to the aggregate level, whereas relationships and causation run in both directions. There are issues (as discussed further below) of how individual behaviour and decision-making is to be analysed and how extensive has to be the recognition that there is heterogeneity of behaviour within and between economic groups. There are two ways in which the analysis of individual decision-making has to incorporate what may be termed macroeconomic and systems influences. First, individuals are often portrayed as able to buy or sell what they wish at the parametric prices; and then mesoeconomic and macroeconomic complications arise since in general terms the amount demanded will not equal the amount which could be supplied. But there are many other influences, which reflect the macroeconomic conditions, such as the levels of employment and income, and the degree of credit rationing. Second, there are social influences, which mould individual decision-making. In the macroeconomic context, important influences here come from relationships between individuals and the degree to which relative income (whether relative to the income of others or to previous income levels) has influence on consumer behaviour.

The analysis of decision-making and activities at the isolated individual level would be rather uninteresting from an economic and social perspective. Indeed, it is difficult to think of many decisions and activities which do not have ramifications for others. It becomes rather like the 'economics of playing solitaire'—a solo activity but even then one with 'rules of the game', which are socially defined. A thought experiment such as how does an individual respond to different relative prices and so on in order to map out a demand curve is not of a great deal of interest. To use such information to make comments on economic events would require first some aggregation of the demand curves of relevant individuals and then understanding of how come price is now one unit lower, how do producers respond to wishing to buy more what are the interactions between sales and subsequent production decisions and so on. In a number of respects, economic analysis pays rather little attention to individual decision-making—though the development of behavioural economics and of experimental economics has led to more attention. A quick look at text books with microeconomics in the title would reveal much focus on markets and industries in which there are interactions between the decisions of individuals (broadly defined). Further, in this approach, decision-making is approached at the individual level, although there are often slippages into treating decision-making at the household level. In so far as individuals live in households with some sharing (e.g. of domestic arrangements) and some elements of joint decision-making, issues of aggregation from the individual to the household level are involved.

It is clear that using this metaphor of microeconomic 'foundations' and its implications are severely misleading. While much analysis starts from the micro level, it cannot finish there, and there are feedbacks from the meso level and the macro level, which have to be fully acknowledged. Denis (2016) raises the question as to whether macroeconomic analysis must be reducible to and derivable from microeconomic behaviour, and identifies such an approach as "expressing a reductionist or atomistic standpoint, such that the whole is just the sum of its parts" (p. 150).

The appeal for 'microeconomic foundations' is often associated, implicitly or explicitly, with the assertion that those foundations should be clearly based on the forward-looking utility-maximising individual operating with rational expectations, and then the 'representative agent' is invoked to enable a form of macroeconomic analysis to be conducted based on such microeconomic 'foundations'.

> Modern macroeconomics seeks to explain the aggregate economy using theories based on strong microeconomic foundations. This is in contrast to the traditional Keynesian approach to macroeconomics, which is based on ad hoc theorising about the relations between macroeconomic aggregates. In modern macroeconomics, the economy is portrayed as a dynamic general equilibrium system that reflects the collective decisions of rational individuals over a range of variables that relate to both the present and the future. These individual decisions are then co-ordinated through markets to produce the macroeconomy (Wickens 2008, p. 1).

As King (2012) notes, after citing this quote, "the reference that Wickens makes to 'the collective decisions of rational individuals' unwittingly points to the difficulty: it is, of course, *individual* and not collective decisions that are, supposedly, being aggregated. If they really were *collective* decisions, the aggregation would be unnecessary" (p. 1). It is also asserted that the co-ordination of those individual decisions has been effected: yet a part of macroeconomic analysis relates to whether decisions are indeed co-ordinated and the consequences of failures of co-ordination.

This use of utility-maximising individual as *the* acceptable foundation is presented as being unproblematic. Yet as Denis (2016) argues, "the assumptions which microfounded approaches make in connection with the representative agent and the notion of equilibrium at the heart of DSGE show a striking degree of ad-hocery—a failure to ground key assumptions required for tractability" (p. 150).

It must be recognised that invoking a utility analysis for the individual provides apparent links with the evaluation of changes and policies in economic welfare terms. Woodford (2003) argued that "an advantage of proceeding from explicit microeconomic foundation is that in this case, the welfare of private agents – as indicated by the utility functions that underlie the structural relations of one's model of the transmission mechanism [of monetary policy] – provides a *natural* objective in terms of which alternative policies should be evaluated" (p. 12; quoted by Denis 2016, p. 137; emphasis added). The limitations of this have to be acknowledged. The welfare criteria are built up from individual utility functions, and hence welfare is deemed to be enhanced if (using the Pareto criteria) some individuals' utility is increased, while the utility of others is not diminished. Using a representative agent approach, economic welfare is deemed to be enhanced if the utility of that agent increases. However, that means economic variables, which do not contribute to individuals' utility, are omitted from consideration in terms of economic welfare. Macroeconomic policies are then to be evaluated in terms of individual utility, and broader concerns are not considered. A notable omission would be inequality and the distribution of income. Wren-Lewis (2011) comments that "such derivations may result in policy objectives that are highly unrealistic, because the models from which they derive generally contain no unemployment and no bankruptcies" (p. 131).

The RARE (representative agent rational expectations) approach has three key elements: first, the use of the notion of representative agent—that issues of aggregation can in effect be ignored in that there is an agent, which is representative of all. Second, the representative agent is a forward-looking utility maximiser, subject to lifetime budget constraint. Third, the agent holds 'rational expectations' on the future—the future is essentially knowable such that the agent can foresee the probabilistic future. This RARE approach has come to dominate 'modern' macroeconomics notably in the dynamic stochastic general equilibrium (DSGE) framework and the 'new consensus in macroeconomics'. It can be critiqued in many ways, and here the focus is on three:

(i) It is a surprising feature that mainstream macroeconomic analysis relies entirely on utility maximisation over an indefinite time horizon with information on the future path of income and so on. Although utility maximisation still plays a significant role in microeconomic analysis, other forms of motivation and decision-making are frequently considered. It is also the case that corporations and firms are regarded in the mainstream macroeconomic analysis as expressions of the interests of their shareholders, who are in turn individuals. Thus, a corporation is the agent of individuals and is treated as maximising profits in the interests of its shareholders. There is then assumed to be a consensus of interests amongst a corporation's shareholders focused on profit maximisation. Further, there is no sense that the corporation, being a 'legal person' and an organisation, develops its own interests (such as survival, expansion) or that the key decision-makers within the corporation pursue their own interests.

With the representative agent approach, corporations and other organisations 'do not exist' as entities which have their own interests—it is rather that corporations are merely the agents of household and reflect the interests of the representative households. The representative agent approach is essentially based on an individualistic approach, albeit one in which the actions of diverse individuals can be summarised in terms of a single representative agent. What role is there then for organisations and institutions in this set-up? Implicitly (if not explicitly) a corporation is run in the interests of its shareholders; and a trade union in the interests of its members. Yet, a corporation is a legal person with rights and obligations.

As agents of households, firms do not act as employers of individuals. As King (2012) notes, "neither employment nor unemployment plays any significant role" (p. 1), in the model, and there is no index entry for unemployment in Wickens (2008). In a similar vein, firms make investment decisions as agents of households, and as such their investment decisions reflect the savings intentions of households and the intertemporal allocation of income desired by households (Blanchard and Fischer 1989; Woodford 2003).

The 'power' of the RARE approach is that (as exemplified by the representative agent) is based on its adoption of an institutional approach focused on the individual and where all individuals adhere to a uniform behaviour (that is utility maximisation). Economic system analysis has to include theorising on individual and institutional behaviour (as well as macroeconomic considerations). It can be readily recognised that industries, markets and corporations operate in diverse ways, which change over time and differ between countries. In the macroeconomic context, the ways in which price setting and determination, investment, production and employment decisions are made differ between industries, markets and so on. In a similar vein, wages are determined in a variety of ways through the economy. Wage determination can be used here to illustrate the issues involved. A first point to make is to what the wage determination relates. It has been a basic postulate of the mainstream models that in effect it is real wages, which are settled in the labour market; the demand for and the supply of labour are deemed to be functions of the real wage, and the interaction of demand and supply would settle the real wage in equilibrium. However, it is a general view of post-Keynesian economics that it is the money wage which is settled, though influenced by perceptions of what that money wage means in real terms. For workers, the real wage is their money wage adjusted for the price of goods and services which they buy, whereas for firms it is the relationship between money wage and price received for the goods produced. The second point is that it is generally recognised that there are different ways in which wages are settled, and economists and other social scientists have analysed and modelled wage determination in many ways such as bargaining models, efficiency wage considerations and competitive market determination, and within each of those types of models, there are many variants. The varieties of models and approaches illustrate issues of aggregate relationships which are returned to below and also illustrated the roles of institutions and departures from the utility-maximising approach of RARE.

> (ii) Paradoxically, the sort of macroeconomic models which claim to give a picture of economic reality (albeit a simplified picture) have almost no activity which needs coordination. This is because typically they assume that the choices of all the diverse agents in one sector – consumers for example – can be considered as the choices of one 'representative' standard utility maximizing individual whose choices coincide with the aggregate choices of the heterogeneous individuals. My basic point in this chapter is to explain that this reduction of the behaviour of a group of heterogeneous agents *even if they are all themselves utility maximizers*, is not simply an analytical convenience as often explained, but is both unjustified and leads to conclusions which are usually misleading and often wrong. Why is this? First, such models are particularly ill-suited to studying macroeconomic problems like unemployment, which should be viewed as coordination. (Kirman 1992, p. 117)

Kirman (1992) provides four reasons why it is untenable to argue that models using a representative agent "are not intended to study those problems which involve, in an essential way, questions of coordination but are designed to examine some central macroeconomic phenomena" (p. 118). First, "there is no plausible formal justification for the assumption that the aggregate of individuals, even maximizers, acts itself like an individual maximizer … Secondly, … [t]he reaction of the representative to some change in a parameter of the original model …. may not be the same as the aggregate reaction of the individuals he 'represents'. … Thirdly, … it may well be the case that in two situations of which the representative prefers the first to the second, every individual prefers the second to the first. Lastly, trying to exaplain[sic] the behavior of a group by that of one individual is constrainint[sic]. The sum of the behavior of simple economically plausible individuals may generate compicated[sic] dynamics, whereas constructing one individual whose behavior has these dynamics may lead to that individual having very unnatural characteristics".

(iii) The system analysis also has to be based on views on the 'human condition'. The RARE approach is based on a probabilistic view of the future in which the underlying forces of the economy operate. The alternative (post-Keynesian) 'vision' is based on fundamental uncertainty. Keynes (1937) drew the distinction between risk and uncertainty: "The sense in which I am using the term [uncertainty] is that in which the prospect of a European way is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know" (p. 214). Taking that view seriously (which I do) completely undermines the RARE approach—calculating expected utility in an uncertain world is not feasible and cannot be used to analyse individual behaviour. Further, as the future is uncertain, future outcomes will be moulded by actions and decisions en route to that future; in other words, there will be path dependency.

Within mainstream macroeconomics, notions of path dependence and hysteresis have been flirted with—notably with regard to the effects of the experience of unemployment on future labour supply and employment decisions. Growth models within the mainstream have been dominated by neo-classical and endogenous growth theories, which share the feature that growth is viewed in terms of supply side. In some contrast, heterodox economics has generally been aware of hysteresis.[1] The demand-driven approach lays down a path-dependency approach as compared with the mainstream supply-side approach as the path of demand and its structure impacts on investment, capital formation and sectoral developments.[2]

There is then a stream of macroeconomic analysis, which is focused on issues of co-ordination—how are economic activities co-ordinated? And why and how does co-ordination fail? Within macroeconomics, unemployment (of labour) has often been viewed as a failure of co-ordination—there are people willing to work, and there are people wishing to buy what could be produced. The literature coming from the 're-appraisal of Keynesian economics' through 'temporary equilibrium' emphasises this co-ordination approach. In effect a set of perfectly competitive and clearing markets would ensure full employment—after all the demand and supply of each type of labour would be brought into equality. But the question was posed as to the effects of trading out of equilibrium: there would still be a co-ordination of demand and supply, but that would be through the short side of the market dominating (actual trade = minimum of ex ante demand and ex ante supply). There is a failure of prices to adjust and ensure full employment equilibrium.

If by microeconomic foundations is meant the implementation of individual decision-making, then there is an obvious and immediate issue. Namely, that one individual's decision is not compatible with others' decisions in the sense of leading to inconsistent outcomes (e.g. the individual wishes to buy X, but the other individual is not willing to sell X). The analysis of perfect competition raised two related issues. First, as Arrow (1959) pointed out, if as assumed in the perfectly competitive model all economic agents are price takers, the question is: how do prices change? Attempts were made to overcome that issue ranging from invoking a Walrasian auctioneer whose role was to adjust prices through to some economic agents exploiting a limited monopoly to vary prices.

Second, how could the ability of economic agents to buy and sell as much as they wished at the prevailing price be compatible with economic agents being demand constrained as envisaged in the basic Keynesian macroeconomics story? This was in effect resolved by looking at a situation of non-market clearing where the minimum of demand and supply would be the amount traded. In a situation where the market price was above equilibrium, supply would exceed demand and suppliers (firms) would find themselves demand constrained.

These issues were often approached in the context of macroeconomics, and finding a consistency between microeconomics (as represented by individuals being price takers) and macroeconomics (where individuals are often seen as demand constrained). It should also be considered in terms of the relationship between microeconomics and mesoeconomics—decisions made by individuals (in this context with respect to relative prices) cannot in general be fully implemented and the effects of that have to be further considered.

Economic analysis operates at a number of levels: here the individual level of decision-making (what is often referred to as microeconomic foundations), the meso level (such as market, industry) and the macro level are distinguished. Although the question has been raised on the microeconomic foundations of macroeconomics, many of the issues also arise in relation with the microeconomic foundations of mesoeconomics. A pertinent example here concerns what is termed price rigidity/inflexibility. It is individual prices rather than the price level, which is deemed to be rigid/inflexible. Firms set prices at which they are willing to trade: the reassessment of price takes place non-instantaneously—it may be a matter of days or of months. Price flexibility is a meso-level issue though it can have macroeconomic implications on the path of the economy. At the level of the market/industry (and indeed firm), there is a question of how prices change, how frequently and in response to which forces. Similarly, there are 'fallacy of composition' issues at the meso level, which reflect that what may be (approximately) true at the individual level does not hold at the meso level (and then by extension at the macro level). An individual may be portrayed as able to purchase what she wishes at a prevailing market price. But, of course, there has to be a corresponding willing and able seller(s).

At the meso level, there are tests of consistency to apply—in a market, is there a consistency between what individuals wish to buy and what other individuals wish to sell? If there is not, what are the margins of flexibility (e.g. are sellers able to run down stocks)? In this example, which side of the market determines the outcome—a usual assumption being that the 'short side' of the market prevails and that it is the minimum of desired demand and desired supply which determines the amount actually traded.

3 Nature of Macroeconomic Relationships

There are clearly relationships which apply at the macroeconomic level (and similar remarks would apply to the mesoeconomic level), which do not have microeconomic underpinnings. The requirement for savings = investment is a notable one. Pasinetti (1974) argues that his investigation

> is not 'macro-economic' in the sense of representing a first simplified rough step towards a more detailed and disaggregated analysis. It is macro-economic because it could not be otherwise. Only problems have been discussed which are of a macro-economic nature; an accurate investigation of them has nothing to do with disaggregation. They would remain the same – i.e. they would still arise at a macro-economic level even if we were to break down the model into a disaggregate analysis. (p. 118)

King (2012) argues that there "are *macroeconomic* theories, which are consistent with a very wide of assumptions about individual behaviour and therefore also with a considerable variety of microeconomic models" (p. 24).

There is a partial but incomplete truth here; notably the relationship of the equality (in a closed private economy) between savings and investment (in terms of outcomes, and in terms of an equilibrium condition). Further, there is a 'split' between those who envisage that investment 'causes' savings (which may be termed the general Keynesian/Kaleckian approach) and those who envisage that there is a pool of savings which lead to investment (the neo-classical approach, now in the DSGE models). Within each of these broad groupings, there will be differences of view on how savings and investment are to be modelled. The ways in which savings and investment are approached will have implications for macroeconomic behaviour even though it is constrained by the requirements of savings equals investment as an outcome. Further, there have to be assumptions made on the way in which banks and the financial system operate—after all investment expenditure has to be financed.

There are relationships that hold at the macroeconomic level, which may involve individual behaviour and constraints, but which crucially involve a consistency requirement. The most well-known of these is the equality between savings (S) and investment (I) (for a closed private economy taken for simplicity). For any individual (person, corporation), considered as a balance-sheet constraint, borrowing/lending = savings minus investment. The requirement that one person's borrowing is another person's lending means that, in total, net borrowing/lending equals zero. Summing over individuals yields aggregate savings equals aggregate investment. Simply adding together the individual-level constraint would merely total borrowing/lending = savings minus investment. To arrive at the macro relationship requires noting a consistency requirement (one person's borrowing is another's lending) to yield savings equals investment. This is an aggregate/macro relationship. To add to it requires saying something on the determinants of savings and investment. The determinants of savings and investment may form a long list, and this part of the 'model' can make only a small contribution. If, for example, we have s.Y = I, with I given, which provides the determination of Y, there is then an associated adjustment mechanism where Y adjusts to fulfil that equilibrium condition. When there is a much longer list, all that can be said is that the equality between savings and investment has to be assured but how and when is left open.

Depending on how aggregate savings equals aggregate investment has been built up sets how the equality is interpreted. If at the level of the individual, borrowing/lending = savings minus investment is a balance-sheet outcome, then aggregate savings equals aggregate investment is the national income accounts identity. On the other hand, putting 'desired' into the equation at the individual level and combining a balance requirement (borrowing = lending) yields desired savings equals desired investment.

This relationship can then go on to provide the 'paradox of thrift', as mentioned above, in which from a simple representation of desired savings and investment, sY = I, a higher propensity to save does not lead to a higher level of savings. This 'paradox of thrift' is the best known of the paradoxes, which arise in post-Keynesian economics, as listed by Lavoie (2014). Other paradoxes including the 'paradox of costs' (higher real wages lead to higher rate of profit), paradox budget deficits (raise profits), paradox of debt (efforts to de-leverage might lead to higher leverage ratios), paradox of tranquillity (stability is destabilising), paradox of risk ('availability of individual risk cover leads to more risk overall'), paradox of liquidity ('new ways to create liquidity end up transforming liquid assets into illiquid ones') and paradox of profit-led demand ('generalized wage restrictions lead to a slowdown in growth even when all economies seem to be profit-led').[3] These 'paradoxes' are representative of macroeconomic relationships in the sense that they are not derived merely by the summation of an individual-level relationship. As illustrated by the 'paradox of thrift', interactions between the behaviour of individuals and adjustment processes have to be taken into account which cannot be solely the summation across individuals.

Money is a generally accepted means of payment, which is a credit relationship that depends on trust (notably that a 'piece of paper' will be accepted by others in payment). It is a macroeconomic concept in two ways. First, individuals accept money in payment only because they believe others will do so from them. Money is a social construct and one which could not yield any benefit to an isolated individual. Second, there are significant macroeconomic relationships involving money. A monetarist approach would invoke some form of MV = PT relationship. A post-Keynesian approach would note that the amount of money in existence has to be held by people and that stock of money becomes demand determined in the sense of the willingness of people to hold money.

Macroeconomic analysis also contains relationships and concepts which are macroeconomic in nature (in the sense of the quote from Pasinetti 1974, as above) and which are derived from some form of sustainability. The non-accelerating inflation rate of unemployment (NAIRU) provides an example. The NAIRU is a level of unemployment at which (according to the theory at hand) the rate of inflation would be constant. The rate of inflation and the rate of unemployment are macro concepts, and the NAIRU cannot be derived from summing individual experiences.

The NAIRU is a concept which I have found to be problematic (Sawyer 1999) and I prefer to refer to an inflation barrier (as in Arestis and Sawyer 2006). It should be noted that the NAIRU is a property of a specific theoretical model, and as such may not be a property of the real world. The NAIRU may not be a level of unemployment at which the economy generally operates—it may be a 'weak attractor'. There may also be forces at work which lead the inflation barrier to be being path dependent (Sawyer 2001). The NAIRU is a macroeconomic concept—that is, it only arises at the economy-wide level, and is more akin to a sustainability condition—if unemployment (according to the model) deviates from the NAIRU, then inflation will rise or fall continuously, imposing that sustainability condition of constant inflation yields the NAIRU. This is not to say that the NAIRU will be realised as it may not act as a 'strong attractor' for economic activity.

The 'natural rate of interest' provides a further example. The 'natural rate of interest' is again a macroeconomic phenomenon in the sense that it has no microeconomic counterpart, and is intended to correspond to a balance between savings and investment. It is also model dependent, and only has meaning in a group of models, but not in others. For example, a post-Keynesian/Kaleckian model of the economy in which savings and investment are insensitive to rates of interest would not generate a 'natural rate of interest'.

There are the many concepts and relationships which are macroeconomic in nature in the sense that they cannot be derived by the summation of individual microeconomic behaviour. These concepts and relationships are widely recognised even in mainstream economics and serve to show that macroeconomic analysis cannot be approached through mere aggregation from the individual.

4 Undertaking Macroeconomic Analysis

Macroeconomic analysis, whether in theoretical terms, for empirical forecasting or for pedagogical reasons, has generally proceeded by invoking relationships between macroeconomic aggregates. The use of aggregate functions could be seen as a reflection of a lack of human computing power. It is possible to shift the IS and LM curves around and derive predictions from them with a piece of paper. As individuals, there is a lack of computational ability to deal with say ten consumption functions and so on, though computer power would be able to do so. The IS-LM analysis, for example, is based on equations, which map equilibrium positions in terms of income and rate of interest based on an aggregate savings function (savings based on income), aggregate investment as a function of rate of interest and demand for money (function of income and rate of interest) and a given stock of money. In each case, the assumption made is that there is individual-level behaviour (in respect of savings, investment and demand for money) which can be aggregated to provide comparable behaviour at the aggregate level. But, as hinted at when discussing the representative agent above, the conditions under which the aggregate functional relationship exists and mimics the individual functional relationship are likely not to be met. The question then arises whether attempting the simplification of invoking an aggregate relationship may mislead.

First, consider the case of the consumption function. Take the simplest of consumption function $c_i = a_i + b_i.\, y_i$ for individuals $i = 1,2,\ldots n$; then summing across individuals yields $C = \sum_1^n c_i = \sum_1^n a_i + \sum_1^n b_i y_i = A + BY$; the last term equals $\sum_1^n B.\left(b_i / B\right).Y.\left(y_i / Y\right)$, which can be written as $B.Y$ if b_i/B, y_i/Y are constants. Specifically, if there are variations in the distribution of income (and hence y_i/Y vary), then there will be shifts in the consumption function. Introducing further variables would serve to complicate the picture. For example, the inclusion of individual wealth would involve similar distributional issues as those from income, but also raises issues of wealth valuation. Households whose consumption plans would exceed their income would be faced by credit constraints on their ability to borrow. The overall availability of credit (a macroeconomic phenomenon) would also need to be introduced.

The first conclusion to be drawn from this would be that the use of macroeconomic aggregate relationships, which mimic an individual-level relationship, may be misleading if some of the aggregation assumptions do not hold. In the example above, that could be if the distribution of income (between individuals) was also changing. A further example is coming from Steedman (1992), where he questioned the move from a relationship at the level of an industry under which it was postulated that the markup of price over unit costs depends on the 'degree of monopoly' to a comparable one at the macroeconomy. Thus, a rise in the degree of monopoly may not lead to a rise in the profit share.

A more severe example comes from the use of aggregate production functions (at the core of which aggregate output is related to aggregate employment and aggregate capital). Felipe and McCombie (2013) derive the subtitle of their book from "scientific idea is 'not even wrong' if it is so incomplete that it cannot be used to make predictions that could be compared to observations to see if the idea is wrong" (Peter Woit 2006, referring to some remarks by Wolfgang Pauli). Their book "shows that the aggregate production function suffers from this same problem, namely it is 'not even wrong'" (p. vi).

After noting the widespread use in macroeconomics and neo-classical growth theory, Felipe and McCombie (2013) state that there are numerous methodological problems in the use of aggregate production functions. Notable amongst these are the Cambridge 'capital controversy' issues ("theoretical problems of aggregating heterogeneous capital goods into a single index that could be taken as a measure of 'capital' as a factor input", p. 3) and general aggregation issues ("this shows that the conditions under which it is possible to sum micro-production functions to give an aggregate production function are so restrictive as to make the concept of the aggregate production function untenable", p. 4).

The work of Felipe and McCombie (2013) raises some significant issues. They show that the econometric estimation of what appears to be an aggregate production function (e.g. regressing output on labour, capital stock) may well provide satisfactory estimates (relationship statistically significant). Yet, the regression estimates do not represent an aggregate production function. Insofar as the distribution of income between wages and profits is little changing, then a Cobb-Douglas production function will appear. The first derivatives of the production function cannot then be used to provide estimates of the marginal productivities of the factors.

This discussion suggests that the uncritical use of aggregate functions, which mimic corresponding micro/meso functions, can often led to misleading conclusions. In the case of the aggregate production function, not only is there the problematic nature of 'aggregate capital' but also the assumption being made that there is technical efficiency assumed, whereas it is well-known that firms differ substantially in terms of technical inefficiency. For the aggregate consumption function, it is often forgotten that households differ in terms of how far they are credit constrained. Using a set of aggregate functions may be the first convenient step for macroeconomic analysis, but it has to be backed up through using relationships, which hold at the individual or group level. The developments of simulation and agent-based modelling now provide ways of undertaking such analysis.

There is, though, a further issue, namely, that important relationships arise at the aggregate level, which must be captured in the analysis. The particular example would be the equality between savings and investment at the aggregate level (for closed private economy). However savings decisions and investment decisions are arrived at, there is still that requirement. This may though only be a reinforcement of issues at say the market level. It may not be possible to derive a demand for X curve and a supply of Y curve summed from individual demand and supply curves. Yet it would still be required that demand equals supply (whether as an actual equilibrium condition or in terms of outcomes).

5 Summary and Conclusions

I conclude by echoing the sentiments of Vercelli (2016) when he writes that

> the only way to reduce macroeconomics to *Homo-economicus* microeconomics is to kill macroeconomics as an autonomous discipline, denying its inner life rooted in its emergent properties. We believe, on the contrary, that a vital and lively macroeconomics is needed: autonomous but with sound methodological and institutional foundations. To this end, we need non-dogmatic microfoundations in the sense – different from that supported by the MIF [microfoundations of macroeconomics] – of a clarification to assumptions about individuals' features and behaviour and how the interaction between individuals causes emergent properties. (p. 164)

The interpretation here is that macroeconomic analysis requires a pluralistic and realistic microeconomic basis—that is not one based on utility-maximising individuals with rational expectations. The microeconomic basis has to reflect the institutional arrangements in the economy being analysed—how do corporations behave particularly with regard to investment, pricing and employment? How are wages determined, and how to incorporate the heterogeneity of institutions and their behaviour? Simple aggregate relationships are unlikely to exist, which poses major issues for the techniques to be deployed by macroeconomic analysts. There has to be full respect for consistency and sustainability criteria, which provide much of macroeconomic analyses. There has to be behavioural underpinnings of individuals and organisations where the interactions between individuals and organisations set the path of the economy. The macroeconomic conditions in turn mould the behaviours of individuals and organisations.

Notes

1. For discussion, see Arestis and Sawyer (2008), Sawyer (2010).
2. See, for example, Setterfield (2002).
3. Quotes in this paragraph are from Lavoie (2014, p. 18).

References

Arestis, P., & Sawyer, M. (2006). Aggregate demand, conflict and capacity in the inflationary process. *Cambridge Journal of Economics, 29*(6), 959–974.

Arestis, P., & Sawyer, M. (2008). Path dependence and demand-supply interactions in macroeconomic analysis. In P. Arestis & M. Sawyer (Eds.), *Path dependency and macroeconomics* (pp. 1–36). Basingstoke: Palgrave.

Arrow, K. J. (1959). Towards a theory of price adjustment. In M. Abramovitsz (Ed.), *The allocation of economic resources*. Stanford: Stanford University Press.

Blanchard, O. J., & Fischer, S. (1989). *Lectures on macroeconomics*. Cambridge, MA: MIT Press.

Denis, A. (2016). Microfoundations. *Review of Political Economy, 28*(1), 134–152.

Felipe, J., & McCombie, J. S. L. (2013). *The aggregate production function and the measurement of technical change: 'Not even wrong*. Cheltenham: Edward Elgar.

Keynes, J. M. (1937). The general theory of employment. *Quarterly Journal of Economics, 51*(2), 209–223.

King, J. E. (2012). *The microfoundations delusion*. Cheltenham: Edward Elgar.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives, 6*(2), 117–136.

Lavoie, M. (2014). *Post-Keynesian economics: New foundations*. Cheltenham: Edward Elgar.

Pasinetti, L. L. (1974). *Growth and income distribution: Essays in economic theory*. Cambridge: Cambridge University Press.

Robbins, L. (1933). *An essay on the nature and significance of economic science*. London: Macmillan.

Robinson, J. (1972). The second crisis of economic theory. *American Economic Review, 62*(1), 1–10.

Sawyer, M. (1999). The NAIRU: A critical appraisal. *International Papers in Political Economy, 6*((2).) Re-printed in Arestis, P., & Sawyer, M. (Eds.). *Money, finance and capitalist development* (pp. 220–254). Aldershot: Edward Elgar.

Sawyer, M. (2001). The NAIRU, aggregate demand and investment. *Metroeconomica, 53*(1), 66–94.

Sawyer, M. (2010). Path dependency and the interdependence of demand and supply in macroeconomics. *European Journal of Economics and Economic Policies: Intervention (EJEEP), 8*(2), 277–293.

Setterfield, M. (Ed.). (2002). *The economics of demand-led growth: Challenging the supply-side version of the long run*. Cheltenham: Edward Elgar.

Steedman, I. (1992). Questions for Kaleckians. *Review of Political Economy, 4*(2), 125–151. https://doi.org/10.1080/09538259200000011.

Vercelli, A. (2016). Microfoundations, methodological individualism and alternative economic visions. *Review of Political Economy, 28*(1), 153–167.

Wickens, M. (2008). *Macroeconomic theory: A dynamic general equilibrium approach*. Princeton: Princeton University Press.

Woit, P. (2006). *Not even wrong: The failure of string theory and the continuing challenge to unify the laws of physics*. London: Jonathan Cape.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.

Wren-Lewis, S. (2011). Internal consistency, price rigidity and the microfoundations of macroeconomics. *Journal of Economic Methodology, 18*(2), 129–146.

6

A Coherent Approach to Macroeconomic Theory and Economic Policies

Philip Arestis

1 Introduction

It has been a huge pleasure to have been a colleague of John McCombie in the Department of Land Economy, University of Cambridge. John has been an excellent colleague in every aspect of our work. It is, therefore, a great pleasure to edit and contribute to this volume in honour of John.

This contribution relies mainly on the notion that there is often inadequacy of aggregate demand relative to what would be required for full employment of the factors of production. The level and distribution of productive capacity can often be inadequate to underpin full employment. Fiscal and monetary policies are of course important and we suggest that co-ordination of them is a way forward. In this contribution, we briefly discuss the theoretical framework that underpins the relevant economic policies discussed subsequently. In terms of the latter, we argue that two 'new' important policy dimensions, which have been ignored in

P. Arestis (✉)
Department of Land Economy, University of Cambridge, Cambridge, UK

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_6

the past, namely, distributional effects and financial stability, should seriously be taken on board in relevant discussions and implementation of economic policies. Such economic policies are thereby urgently required to avoid crises similar to the 'Great Financial Crisis' (GFC) of 2007/2008.[1] We discuss such economic policies and also, but briefly, the current 'unorthodox' monetary policies.

We proceed as follows. After this short introduction, Sect. 1, Sect. 2 deals with our theoretical background that comprises of five blocks.[2] Section 3 discusses the 'new' economic policies that emerge from the theoretical framework of this contribution. Section 4 discusses the current 'unorthodox' monetary policies. Finally Sect. 5 summarises and concludes.

2 A Coherent Theoretical Macroeconomic Model

The purpose of this contribution is to discuss our theoretical framework that underpins our proposed economic policies. Clearly, economic policy formulation is heavily conditioned by the underlying theoretical framework that should underpin it. We, thus, begin with the essential elements of such a theoretical framework. The overall focus of economic analysis should be: sustainable and equitable economic development and growth at full employment. Achieving such objective requires the maintenance of a high level of aggregate demand and sufficient productive capacity. The general background to this theoretical framework relates to an economy, which has degrees of instability and is prone to crisis. It also relates to a monetary production economy in which finance and credit play a significant role. This theoretical framework draws on five blocks as summarised in the appendix. In the rest of this section, we discuss briefly the main elements of each block. We begin with block I.

Block I (Eqs. 6.1, 6.2, 6.3, 6.4, 6.5 and 6.7 as in the appendix). This is based on the demand side of the economy, which relates to expenditure, income and employment, and also on the supply side. The level of economic activity is set by aggregate demand. No market-based mechanism exists to propel the level of aggregate demand to any specific level of output. Distributional effects are paramount and are seriously taken on board. Changes in economic activity affect the rate of change of prices and wages, and consequent changes in the distribution of income between wages and profits emerge. Changes in the distribution of income have effects on the level of aggregate demand, with the nature of the effects depending on whether there is a wage-led or a profit-led regime. Aggregate demand has a dual characteristic in this model: it is a relatively volatile component; and it is also a creator of productive potential. This establishes interdependence between demand and supply. The supply side of the economy is viewed in terms of the following characteristics: the interaction between production decisions of firms in the light of the (expected) level of aggregate demand and the consequent decisions on employment. An important part of aggregate demand is investment, which influences the supply side of the economy since it is the principal determinant of potential output and labour productivity in the long run. Not only does investment spending add to the stock of capital available per worker, but also determines the extent to which the capital stock embodies the latest and most efficient technology (Blinder and Zandi 2010). It is also viewed in terms of the relationship between prices and wages, and their setting. Clearly, this approach denies the validity of the NCM approach that portrays the long run as characterised by a supply-side equilibrium (at NAIRU), with aggregate demand having no impact whatsoever.

Block II (Eqs. 6.8, 6.9, 6.10, 6.11, 6.12 and 6.13 as in the appendix). It relates to the distributional aspects and the inflationary process. The range of factors, which impact on the distributional aspects and the rate of inflation, includes struggle over income shares, the level and rate of change of the level of aggregate demand and cost-push factors emanating notably from the foreign sector (changes in import prices and the exchange rate). It is also the case that the sources of inflationary pressures vary over time.

Block III (Eqs. 6.14, 6.15, 6.16, 6.17 and 6.18 as in the appendix). This block relates to the money, credit and finance aspects. Money is essentially endogenously created within the private sector with loans initiated by banks, thereby generating bank deposits. The behaviour of banks and related credit institutions become important for the economy. Their willingness or otherwise to create loans and the terms upon which they are provided impact on the level and structure of demand. The central bank sets the key policy interest rate, which governs the terms upon which the central bank provides the 'base' money to the banking system. Monetary policies, however, such as credit-rationing by the authorities, which can control the financial sector, are also important. These are what are labelled as 'financial stability' policies as discussed in Sect. 3.3.

Block IV (Eqs. 6.19, 6.20, 6.21, 6.22 and 6.23 as in the appendix). There is also the government sector with its expenditure and taxes along with the public sector borrowing requirement, which are taken on board and examined, as well as endogenised as necessary. When government spending is treated as exogenous, the main reason for such treatment is because "legislative and administrative decisions do not respond predictably to economic conditions" (Blinder and Zandi 2010).

Block V (Eqs. 6.24, 6.25, 6.26 and 6.27 as in the appendix). Finally, the open economy aspects are examined as in this block. The openness of the economy means that the domestic economy is buffeted by events in the rest of the world. A relevant and significant aspect of the foreign sector is that imports and exports are included in the aggregate demand equation and endogenised in this block. This inclusion also reflects the effects on demand (and hence employment) of variations in the exchange rate.

The model just presented is cyclical and could potentially produce periods of instability. It is, thus, paramount that economic policies to stabilise the economy and lead it to high levels of employment and output are vitally necessary. This is undertaken in the section that follows, where we concentrate mainly on 'new' economic policies rather than on the traditional ones.

3 Economic Policies

3.1 Prolegomena

The overall objective of economic policies should be sustainable and equitable economic development and growth, along with the achievement of full employment of the labour force. Maintenance of a high level of aggregate demand and provision of sufficient productive capacity are important prerequisites for such objectives. It is clear from the analysis in Sect. 2 that traditional fiscal and monetary policies employed in a co-ordinated manner (see, also, Arestis 2012, 2013, 2015) could potentially help on this score. Our theoretical analysis, however, suggests that further economic policies for the achievement of the above mentioned objectives are paramount. These are relevant economic policies for a fair distribution of income and financial stability, which have not been sufficiently considered previously,[3] especially so by policy makers. We elaborate on these two types of economic policy in the rest of this section.

3.2 Distributional Policies

Distributional effects should be a major objective of policy as this is clear from our theoretical analysis (see, also, Arestis and González-Martinez 2016). Recent evidence of a steady but sharp rise in inequality is also very supportive of this proposition.[4] Inequality had risen prior to the emergence of the GFC of 2007/2008 and the 'Great Recession' (GR) that followed, and has continued since then. Galbraith (2012) suggests that "inequality was the heart of the financial crisis. The crisis was about the terms of credit between the wealthy and everyone else, as mediated by mortgage companies, banks, rating agencies, investment banks, government sponsored enterprises, and the derivatives markets" (p. 4). Arestis and Karakitsos (2011, 2013, see, also, Arestis 2016) argue that inequality was one of the main causes of the GFC. Stiglitz (2013) suggests that income inequality in the USA has been an important cause of the economic and financial troubles in the last 20 years. According to Stiglitz (op. cit.) in 2007, the income of the top 0.1% in the USA was 220 times larger than the average of the bottom 90%. Piketty (2014) also shows that the income share of the top 1% in English-speaking countries (especially in the USA) rose after 1980; the same 1% appropriated 60% of the increase in US national income between 1977 and 2007. Kumhof and Rencière (2010a, b, 2011) make the point that restoring equality through redistribution of income from the rich to the poor could very well save the global economy from another crisis similar to the GFC. Atkinson et al. (2011) show that the share of US total income going to top income groups had risen dramatically prior to 2007. The top pre-tax decile income share had reached almost 50% by 2007, the highest level on record; the share of an even wealthier group, the top 0.1%, more than quadrupled from 2.6% to 12.3% over the period 1976 to 2007.

Clearly, the declining wage and rising profits share were compounded by the increasing concentration of earnings at the top, especially in the financial sector. An important piece of evidence in the case of the USA is the share of the financial sector to GDP, which almost doubled in size between 1981 and 2007 and subsequently accounted for 8% of the US GDP (Philippon 2008). Between 1981 and 2007, the US financial sector, as measured by the ratio of private credit to GDP, grew from 90% to 210%; also, a sharp, nearly sixfold increase occurred in their profitability as from 1982 and beyond. Similar but less pronounced financial shares are relevant in many other countries. Germany, China and the UK are three examples but many more can be cited (see OECD 2008, for a relevant discussion and empirical evidence on these economies). Turner (2010), the ex-Chairman of the UK Financial Services Authority, made the point in the case of the UK: "there has been a sharp rise in income differential between many employees in the financial sector and average incomes across the whole of the economy". A recent contribution to the UK inequality has been the pursuit of 'quantitative easing' (QE) type of policy. The Bank of England (2012) report shows that its QE programme increased the value of the relevant financial assets by 26% with 40% of the gains having gone to the richest 5% of British households.

Similar results are expected for the US economy, where the top 5% of wealthiest households owns 82% of all individually held stocks and more than 90% of the individually held bonds. Even more recent evidence (US Census Bureau, September 2013) shows that US household incomes have been falling for the fifth consecutive year; the typical US family earned less in 2013 than in 1989. In fact the medium household income is now 8.3% below its pre-GR peak in 2007. Still, the share of the wealth accruing to the top 1% grew by 31% in the three years to 2012, while the rest rose by just only 0.4%. The top 1% is close to full recovery since the emergence of the 2007/2008 GFC, while the bottom 90% hardly started recovering, as Saez (2013) reports. Cynamon and Fazzari (2014) also argue that rising income inequality as from 1980 reduced income growth for the bottom by 95% of the US income distribution. The debt to income ratio of the bottom 95% increased dramatically. The income growth for the top 5% increased by contrast. The consumption-income ratio of the bottom 95% rose dramatically unlike that of the top 5%. The end of that borrowing boom prior to the 2007/2008 GFC caused household spending to collapse, which was the proximate cause of the subsequent GR. However, during the GR and subsequently, the end of the borrowing boom and the higher inequality with the associated demand drag provide an explanation of the slow US recovery following the GR.

These are clear empirical examples, in addition to the theoretical premise as discussed in Sect. 2, of the importance of distributional effects as a clear and vital objective of economic policy; if not accounted for and proper action initiated, it can produce serious problems. A clear message then follows from both our theoretical framework and the evidence discussed in this sub-section: distributional effects should be a major objective of economic policy. We have also argued in Arestis and Sawyer (2011) for the importance of accounting for 'distributional effects' in both economic theory and policy, which have been fatally ignored recently (see, also, Arestis 2016; and Arestis and González-Martinez 2016). Consequently, it is vital that not only economic policies should focus on achieving full employment but should also be geared towards reducing inequality. These theoretical propositions are supported by empirical evidence, as, for example, the findings of Onaran and Galanis (2013), and other relevant contributions as in Lavoie and Stockhammer (2013).[5]

It clearly is the case then that pro-labour distributional policies that promote wage policies, strengthening the status of labour unions and collective bargaining, are important and relevant policies. Such a strategy should be complemented by fiscal and monetary policies, along with proper co-ordination of them. The objective should be full employment. Fiscal policy in particular is an important dimension in this regard (Arestis 2012, 2015). The study by Muinelo-Gallo and Roca-Sagalés (2011) employs an endogenous growth model that incorporates fiscal policy and economic growth along with their effects on income inequality. Pooled-panel estimations are undertaken for 43 upper-middle and high-income countries for the period 1972–2006 to conclude that increases in public investment expenditure reduce inequality without harming output, regardless of whether they are financed through direct or indirect taxes.

Targeting social spending, including people's investment in skills and education, is paramount from the government spending point of view. Reforming taxes to make them fairer is another important aspect of fiscal policy. Indeed, Berg and Ostry (2011) show that a redistributive tax system is associated with higher and more adurable economic growth. Raising the minimum wage and indexing it to inflation is another important tool to fight inequality (see, e.g. The Economist 2014). A further example, and priority, is the removal of subsidies for the 'too-big-to-fail' financial institutions. Such policy initiatives would help to remove, to a large extent, one of the main contributory factors to the surge in wealth at the top of income distribution and to the financial sector in particular. This inequality, as we have argued elsewhere (Arestis and Karakitsos 2011, 2013; see, also, Arestis 2016), was one of the main causes of the GFC of 2007/2008. In the latter sense, addressing the problem of income inequality is even more important today with the background to which we have just referred. A recovery led by domestic demand and an increase in the wage share in the global economy would help to reverse the major factor of inequality.

To summarise, a combination of economic policies is needed to tackle inequality; progressive taxation and public expenditure policies, social welfare, industrial relations[6] are all relevant and important. Most important of it all is the suggestion by Atkinson (2015) that "a more progressive structure for the personal income tax" (p. 290) should be introduced. Atkinson (op. cit.) also suggests that it is of paramount importance to have in place proper distributional policies along with wage policies if a viable growth regime is to emerge and be sustained. We would go a step further, though, and argue that to reduce inequality significantly proper policies as discussed above are necessary, but also with appropriate co-ordination of monetary and fiscal policies, along with financial stability, would be the best way forward. Monetary and fiscal policies should be directed at reducing inequality through appropriate expenditure and progressive tax policies, which should be supported by financial stability type of policies. The latter should be concerned with reforms in an attempt to regulate the financial sector and avoid the type of financial architecture that led to the 2007/2008 GFC.

3.3 Financial Stability Policies

Financial stability should be the priority of central banks. The focus of financial stability should be on proper control of the financial sector so that it becomes socially and economically useful to the economy as a whole and to the productive economy in particular. Financial deregulation entails redistribution effects in favour of the financial sector by allowing for greater risk-taking and higher expected profits. Redistribution of welfare thereby emerges from workers to bankers (Korinek and Kreamer 2013). Greater risk-taking, though, can lead to losses sufficient to cause a credit crunch. As suggested above (see, also, Arestis 2016), this process had been one of the major causes of the 2007/2008 GFC. Clearly, then, banks should serve the needs of their customers rather than provide short-term gains for shareholders and huge profits for themselves. This requires the development of alternative policy instruments alongside the downgrading of interest rate policy, as the only instrument of monetary policy, and of any notion of price stability as the only objective of economic policy.

Financial stability policies have emerged as particularly important in view of the GR. In Arestis and Karakitsos (2013), it is argued that in the past, especially prior to the GFC and GR, a variety of regulatory policies were in place, which were intended to maintain financial stability. However, those policies were focused merely on the stability and viability of individual banking institutions, the microprudential type of policies, rather than embracing also the whole of the financial system, the macroprudential type of policies. The key point here is to bring to the forefront a form of financial policy, which is focused on proper financial stability. Such a policy has to be comprehensive in its coverage, in terms of both the range of financial institutions covered and its international coverage. Such policy needs to act in a counter-cyclical manner and to differentiate between assets. Central banks should go beyond the traditional single objective of targeting inflation; they should monitor price fluctuations of assets, especially housing. Of equal importance is the further proposition, discussed below, which suggests that co-ordination of macroprudential and monetary policies along with fiscal policy is paramount, as argued in Arestis (2015).

Financial stability should incorporate both microprudential and macroprudential instruments. Microprudential instruments relate to the structure and regulation of individual banks. Banks that are 'too big to fail' should be reduced in size; guarantees to retail depositors should be limited to banks with a narrower range of investments; risky banks to taxpayers and economy should face higher capital requirements; large and complex financial institutions can be wound down in an orderly manner; and large banks should not be allowed to combine retail banking with risky investment business. The macroprudential toolkit should account for the potential failures of the system: low levels of liquid assets, inadequate levels of capital with which to absorb losses, too big a financial sector and too leveraged a sector with high risks to the taxpayer and the economy. Thus, macroprudential financial instruments should be able to control the size, leverage, fragility and risks of the financial system. And to quote the Bank of England (2009), "In general terms, the goal of financial stability policies should be the stable provision of financial intermediation services to the wider economy—payment services, credit intermediation and insurance against risk. They should seek to avoid the type of boom and bust cycle in the supply of credit and liquidity that has marked the recent financial crisis" (p. 9). Most importantly from our theoretical framework perspective, macroprudential policy should be linked to other relevant policies that affect cyclical fluctuations, and in particular monetary policy, which affects asset prices and bank credit. The latter variables are also affected by macroprudential policies and thereby can influence the transmission mechanism of monetary policy in terms of avoiding excessive liquidity and containing cyclical fluctuations. In this sense, macroprudential authorities should take credit as an important indicator of financial stability as suggested by our theoretical framework. Possibly all the elements just suggested could be combined so that both micro- and macroprudential instruments would be under the banner of the policy makers and properly co-ordinated with the central bank to avoid conflicting policies and results (see, also, Angelini et al. 2012). Policy makers should avoid rules and employ instead judgement and thus discretion.

Clearly, serious interest in financial stability requires the development of a range of policy instruments. There is an important ingredient in the idea of financial stability as the key objective of the monetary authorities. This relates to the notion of independent central banks, based on the idea of price stability being the single objective of economic policy, to be pursued by inflation-averse central bankers. Such notion substantially weakens the idea of financial stability. With multiple objectives pursued by multiple instruments, there is a need for co-ordination between the macroeconomic authorities (e.g. the Ministry of Finance/Economics and the Central Bank), which is precluded by the independent central bank notion. This suggestion also reinforces the argument for co-ordination between monetary and fiscal policies (Arestis 2015). With financial stability policies in place, the new economic policies will work best if co-ordinated with other areas of economic policy outside the remit of the central bank. The GR has clearly demonstrated the need for more than one objective of economic policy and indeed proper co-ordination of the relevant economic instruments to achieve the objectives.

Such co-ordination is supported by empirical evidence, which suggests that under fiscal and monetary policy (including financial stability) co-ordination, fiscal multipliers are higher than when no policy co-ordination prevails (even bigger than the Keynesian ones). This is possible so long as the fiscal and monetary authorities have a common objective, for example, maximisation of social welfare. The multiplier under fiscal and monetary policy co-ordination, and in the case of deficit spending, is found to be of the order of 3.8 (Eggertsson 2006; see also, Arestis 2015). When there is no policy co-ordination, that is, when the central bank is 'goal independent', the deficit spending multiplier is zero. This large difference in fiscal multipliers is explained by the expectations channel, which is thought to work via inflation expectations. Fiscal expansion increases expectations about future inflation, the real rate of interest is reduced (provided the central bank collaborates with the fiscal authority) and spending is stimulated. Expectations of future income also improve, thereby stimulating spending further (Eggertsson 2011). These results suggest that macroeconomic stability is the joint responsibility of the monetary and fiscal authorities: potentially destabilising behaviour by one authority can be offset by an appropriate stance of the other authority. These results are particularly important in view of the current New Consensus Macroeconomic (NCM) theory and practice that sees fiscal policy better divorced from monetary policy.

It is the case that efforts to establish a financial stability framework have been undertaken as shown in Arestis and Karakitsos (2013). Perhaps the most promising initiative on this score is the establishment in the UK of a Financial Policy Committee (FPC) created in April 2013, which "is charged with a primary objective of identifying, monitoring and taking action to remove or reduce systemic risks with a view to protecting and enhancing the resilience of the UK financial system".[7] The FPC agreed at its meeting of June 2013 on the creation of two further committees: the Financial Conduct Authority (FCA), which is accountable directly to the Treasury and the Parliament. Its purpose is to regulate "financial firms providing services to consumers and maintain the integrity of the UK's financial markets. It focuses on the regulation of conduct by both retail and wholesale financial services firms".[8] And the Prudential Regulation Authority (PRA), which is part of the Bank of England, and whose responsibility is "the prudential regulation and supervision of banks, building societies, credit unions, insurers and major investment firms. In total the PRA regulates around 1,700 financial firms".[9] There is also the Financial Policy Committee (FPC), which is an official committee of the Bank of England; this is a body responsible for macroprudential measures. It focuses on the macroeconomic and financial issues that may threaten long-term growth prospects. And as the Bank of England (2016) confirmed: "The FPC is focused on promoting a financial system that dampens, rather than amplifies, the impact of uncertainty and adjustment on the real economy. This means reducing any pressure on firms to restrict the provision of financial services, including the supply of credit and support for market functioning". The FPC cooperates and coordinates action with PRA and FCA. These committees are in addition to the Monetary Policy Committee created in May 1997.

However, despite this recent UK experience, and other financial stability schemes around the globe,[10] progress around the world on financial reform is extremely slow and worrying poverty of action is evident. And as the IMF managing director (Lagarde 2014) suggests "the behaviour of the financial sector has not changed fundamentally in a number of dimensions since the financial crisis"; the IMF managing director proceeds to complain that "The bad news is that progress is still too slow, and the finish line is still too far". It is the case that financial stability still remains a work in progress across the world; positive action is desperately needed along the lines suggested in this contribution. The annual central bank gathering at Jackson Hole, Wyoming (24–26 August 2017), reinforced this conclusion. The central banks' preoccupation was financial stability; however, mounting tension in terms of ensuring financial stability was a big theme. A relevant factor on this score was the US administration's efforts to loosen bank regulations because, in this view, of their negative impact on bank lending and growth. The response of the Fed Chair, in her speech at the Jackson Hole,[11] was that financial regulation would not hold back bank lending and growth; on the contrary, it strengthens them (see, also Dragui 2017).[12]

4 Current 'Unorthodox' Monetary Policies

Monetary policy since the GFC of 2007/2008 in effect has abandoned the main policy instrument, namely, manipulation of the rate of interest to achieve the central bank's Inflation Target (IT). This was the only policy instrument to achieve the only policy objective, namely, price stability, which had been very fashionable prior to the GFC. In view of the rate of interest reduced to nearly zero in many countries, and even to negative interest rates in some cases after the GFC and GR (and has stayed there ever since in most cases), monetary-policy makers introduced unconventional policies to still achieve an IT. QE has been introduced along with near-zero interest rates and negative interest rates in some cases. A new, and additional, objective has been introduced, namely, financial stability, as argued above,[13] but IT is still around to be achieved through the new 'unorthodox' instruments of monetary policy, namely, QE and near-zero or negative interest rates.[14]

These policies, however, entail serious problems. The most serious problem of the current QE and near-zero/negative interest rate policies is the potential negative impact on the banking system. Commercial bank profitability is determined by the difference between the interest rates they pay on deposits and receive on loans. If lending rates fall more than deposit rates, in view of the fear of the banks that depositors would respond by withdrawing their cash, then the profitability of the commercial banks suffers and the performance of the financial sector is undermined. This would be more likely to happen when negative interest rates emerge as in the case of Japan and the Economic and Monetary Union (EMU) for example. This implies that under such policies the performance of the financial sector would suffer.

A further problem with the current 'unorthodox' monetary policies relates to the main aim of these policies, which is to stimulate investment. But investment has been unresponsive. Our investment equation (see Eq. 6.3 in the Appendix) contains the variables of profitability and income in addition to other variables. If expectations of profitability and income are not robust, investment would not be forthcoming. Central banks need to think in terms of their monetary policies in a way that they would have a direct impact on investment. Financial stability and fiscal policy thereby become paramount. A related problem is that when interest rates of all debt maturities are zero, "then money and long-term government bonds become perfect substitutes (they are both government promises to pay, which offer zero interest), and the creation of one by buying the other makes no difference" (King 2016, p. 183). It is thereby highly unlikely for productive investment to emerge in view of investors would prefer to hold more cash than investing; this is so in view of poor growth expectations and uncertainty. It is also the case that those who rely on bonds for their income, such as banks, insurance and pension companies, suffer substantially. UK pension fund deficits have emerged in view of the QE; also, and according to the IMF, as reported in the Observer (16 October 2016), pension funds have suffered particularly badly across the developed world. As a result a pension's crisis could follow.

Another problem with negative interest rates is that in those cases where there is a strong 'savings culture', negative interest rates would hurt savers and smaller banks that rely heavily on interest income. The European Central Bank (ECB) charges banks across the euro area of

0.4% on their deposits held at national central banks. It is the case that banks in some euro-area countries, Germany in particular, resist to pass negative interest rates on to retail customers in view of the uncertainty of the latter's reaction to such a move. Retail customers might begin to store cash outside the banking system in response to such a move. Banks initially responded by introducing quietly fees for services that were completely free previously (Financial Times, 30 June 2016). More recently, though, and as reported in the Financial Times (10 August 2017) on findings of a survey by Ifo (Munich-based think tank), based on 4000 companies, 20% of German companies have had negative interest rates on their deposits with banks. A number of companies shifted their deposits to other financial assets or changing their banks. It is also stated in the same publication that a small number of lenders in Germany have imposed negative interest rates on individual large savings. In other countries, such as the UK, for example, National Westminster Bank, Royal Bank of Scotland, HSBC, and Lloyds Banking Group have warned business customers that negative interest rates on current accounts could be introduced (personal customers are not to be affected), if the Bank of England base rate was reduced below 0%.

In more general terms, negative interest rates have put financial institutions, investors and savers, under strain. As reported in the Financial Times (21 May 2016), the Fitch credit rating agency estimates show that $10 trillion negative-yielding government bonds cost investors annually around $24 trillion. It is also the case that German banks have accused the ECB for punishing savers and their business model with negative interest rates; and Japanese banks raised the issue of ending their sales of government debt to the central bank (Financial Times, 9 June 2016). All these fragile consequences of negative interest rates have been particularly harmful in the case of Germany as reported in the Financial Times (21 and 22 April, 16 May, and 01 June 2016). Also reported in the Financial Times (21 July 2016), and based on data from the Bank of America Merrill Lynch, more than 50% of German bonds eligible for the ECB's QE have become too expensive (with an interest rate lower than the −0.4% ECB's deposit rate charged on bank reserves) for the Central Bank of Germany to purchase. The ECB president, however, has defended negative interest rates arguing that without them and the ECB's QE, serious deflation would have emerged along with substantially lower euro-area growth.[15]

A further problem is that negative interest rates could produce reductions in the velocity of circulation of money. Economic agents may very well take their money out of the banking sector, and keep it in 'home safes', and in more general terms money could be kept out of circulation in the economy. Such a reduction in money velocity of circulation increases deflationary pressures. Savers, especially by getting low returns on their savings, may be forced to save more rather than spend and stimulate the economy in an attempt to increase savings to make up for what is perceived permanent loss of returns. This would lead to lower consumption and lower GDP growth, thereby making the negative interest rate policy counterproductive. This would be especially so for those savers who are prevented from getting the returns they need for retirement. It is also the case that negative interest rates can cause disruption by jeopardising the insurance companies and pension funds sectors through lowering their incomes. Under such circumstances both insurance companies and pension funds may shift the composition of their portfolios to risky assets, thereby adding to asset price bubble pressures. Another impact of negative interest rates could be on the fragile banking sectors, especially in the EMU. Those institutions that are unable to increase lending or pass the costs of negative interest rates on to their depositors face a serious reduction of their profits, thereby facing constrains on their ability to provide credit.

Our main conclusion in terms of this section's analysis is that the unorthodox instruments have not been really effective in terms of achieving their objectives, especially that of boosting the level of nominal economic activity. In terms of financial stability, and although proposals have been put forward to achieve it (see Sect. 3.3 and footnote 12), not much is evident in terms of implementing these proposals, and thereby avoiding a future crisis of the 2007/2008 GFC type. It is true, though, that central banks managed to avoid a complete collapse of their financial systems and their real economies after the emergence of the GFC. However, monetary policies have been very ineffective in restoring a robust recovery as the proponents expected. The enormous expansion of the monetary base has had little effect on the broader monetary and credit aggregates, let alone on the level of nominal economic activity. No wonder the IMF (2016) World Economic Outlook suggested that the then worldwide poor economic situation, especially in developed countries, risked getting into a full-blown deflation gap. Governments should avoid their continuing over-reliance on central banks and monetary policies, which are increasingly constrained, to single-handedly stimulate economic growth. Governments need to join central banks to undertake more economic policies to boost growth rates. Our suggestion on this aspect is that proper co-ordination of monetary and fiscal policies along with financial stability is the best and probably the only way forward to produce and maintain healthy growth in the economy.

5 Summary and Conclusions

We have put forward a theoretical framework, which entails 'new' aspects of economic policies. In terms of the theoretical framework of this contribution, it is constructed by putting together five building blocks. The focus of all five blocks is on five propositions: (i) aggregate demand is always important for the level of economic activity; (ii) the supply side of the economy has to be fully incorporated; (iii) distributional effects are very important for they do matter, but are not always acknowledged as such; (iv) money is endogenous and credit driven, with financial stability being of primary importance; and (iv) co-ordination of economic policies is paramount.

A number of economic policies follow from our theoretical construct, but two 'new' economic policies emerge. The pursuit of distributional policies, which should be properly and fully considered and proper economic policies should be implemented; such policies, should focus on an attempt to reduce inequality. Inequality is strongly correlated with less economic growth over time; inequality, therefore, matters for it does have an important impact on economic growth (see Berg and Ostry 2011; see, also, Stiglitz 2013). It clearly is the case, then, that reducing inequality and promoting economic growth are "two sides of the same coin" (Berg and Ostry 2011, p. 3). Of equal importance is the pursuit and implementation of financial stability policies by the central banks, a rather slow process currently; the central bank role, however, should be to achieve financial stability. In terms of current monetary policies, QE and low/negative interest rates, they have not been successful in terms of achieving their objectives. A sustainable recovery, with achieving the targeted inflation rate, remains elusive.

The banking sectors around the world are still fragile. A relevant contribution in this respect is the Sarin and Summers' (2016) study, which, using financial market information on a number of relevant variables, shows that despite regulatory changes initiated in the wake of the GR, such as higher capital requirements, increases in bank liquidity and bank stress tests, major financial institutions in the USA and around the world are more vulnerable to adverse shocks than they were before the GFC. The decline in financial institutions' ratio of the market value of their equities to total assets on both risk-adjusted and risk-unadjusted basis, promoted, at least in part, by the new regulations, is the major cause. Proper regulatory policies need to be urgently introduced.

Appendix: Summary of Blocks

In what follows, the sign under the variables indicates the partial derivative of the dependent variable with respect to the relevant independent variable. In all equations in this contribution, lower case letters stand for the rate of change of the relevant variable; otherwise letters indicate the level of the relevant variable.

Block I: Aggregate Demand and Supply

$$Y = C + I + G + \left(X - Q\right) \tag{6.1}$$

where Y is national income, C is consumption, I is investment, G is government expenditure, X is exports and Q is imports and thus $(X - Q)$ is net exports (NE).

$$C = C[\left(WE\left(1 - tw\right), \Pi\left(1 - t\pi\right), R, \Delta \mathrm{BLP_h}\right] \tag{6.2}$$

where W is wages, E total employment so WE is the wage bill, tw is the tax rate on wages, Π is total profits, $t\pi$ is the tax rate on profits, R is the rate of interest on loans to households and $\Delta \mathrm{BLP_h}$ is changes in bank lending to households.

$$I = I\left(\Pi / K, Y / \mathrm{Ya}, R, \Delta \mathrm{BLP_f}\right) \tag{6.3}$$

where the symbols are as above with the exception of K, which is capital stock, $\Delta \mathrm{BLP_f}$ that stands for changes in bank lending to firms, and Ya, which is a measure of capacity output and corresponds to the 'desired level' of operation.

$$\mathrm{Ya} = \mathrm{Ya}\left(\underset{+}{\mathrm{E}}, \underset{+}{\mathrm{K}}, \underset{+}{\mathrm{ST}}\right) \tag{6.4}$$

so that Ya would change over time in the same direction as changes in employment, capital stock and state of technology (ST).

There is a level of employment that corresponds to the capacity-output measure (Ea):

$$\mathrm{Ea} = E\left(\underset{+}{Y / \mathrm{Ya}}, \underset{+}{K}, \underset{+}{ST}\right) \tag{6.5}$$

Ya is taken as a benchmark for firms' investment decisions and employment.

Aggregate supply of output is:

$$Ys = Ys\left(\underset{+}{E}, \underset{+}{K}, \underset{+}{ST}\right) \tag{6.6}$$

where the aggregate supply output (Ys) is determined by employment, capital stock and state of technology, with the symbols defined as above.

There is a level of employment that corresponds to output:

$$E = E\left(\underset{+}{Y}, \underline{K}, \underline{ST}\right) \tag{6.7}$$

Block II: Distributional Aspects and the Inflationary Process

$$\pi = \pi\left[\left(P / \underset{+}{U}LC\right), Y / \underset{+}{Y}a, \underline{R}, \underline{dR_f}\right] \tag{6.8}$$

where the variables, in addition to the ones defined as above, are: π is the profit rate, P is the level of prices, ULC is unit labour cost and dR_f is the debt ratio of firms, defined as total debt to total assets of firms.

$$\text{ULC} = \frac{W}{PR} \tag{6.9}$$

where W is the level of wages, and PR is productivity.

$$w = w\left\{\left[\left(W / P\right)^d \underset{+}{-} \left(W / P\right)\right], \left(Y / \underset{+}{Y}a\right), \underset{+}{p}, \underline{U}, \underset{+}{\pi}, \underset{+}{w^e}\right\} \tag{6.10}$$

where the variables are as above with the exception of the bargaining position of workers, which is defined as the difference between their desired real wage [$(W/P)^d$] and the actual real wage (W/P); p, which is inflation, U is unemployment, where unemployment is taken as percent of the labour force and w^e that stands for expectations of the wage rate.

$$U = U\left[\left(\underline{Y / Ya}\right), \underline{PR}\right] \tag{6.11}$$

where the variables are as defined above.

$$p = p\left[\underset{+}{w},\ \underset{+}{\left(Y / Ya\right)},\ \underline{q},\ \underline{er},\ \underset{+}{p_{\mathrm{rm}}},\ \underset{+}{p^{\mathrm{e}}}\right] \tag{6.12}$$

where the variables are as defined above, with the exception of q, which is the rate of change of productivity, er is the rate of change of the nominal exchange rate, p_{rm} is the rate of change of the prices of raw materials and p^{e} that stands for price expectations.

$$p_{\mathrm{rm}} = p_{\mathrm{rm}}\left(\underline{er}, \underset{+}{WT}\right) \tag{6.13}$$

where the variables are as defined above with the exception of WT that stands for world trade.

Block III: Money and Credit

$$\Delta M = \Delta \mathrm{BDGC} + \Delta \mathrm{BDP} \tag{6.14}$$

where ΔM is changes in the money supply, namely, the sum of changes in bank deposits to the government including currency (ΔBDGC) and changes in bank deposits to the public (ΔBDP). In view of the small proportion of ΔBDGC of the total money supply, we treat it as the residual in the following identity:

$$\Delta \mathrm{BDGC} = \Delta \mathrm{BLP} + \Delta \mathrm{BLG} + \Delta \mathrm{BLES} - \Delta \mathrm{BDP} \tag{6.15}$$

Identity (Eq. 6.15) then defines ΔBDGC as the sum of changes in bank lending to the public (ΔBLP) and to the government (ΔBLG) as well as of changes in bank lending to the external sector including other non-bank lending (ΔBLES), minus ΔBDP. ΔBLES is treated as an exogenous variable and with ΔBLG endogenised in Block IV (see Eq. 6.19), the remaining variables in Eq. (6.15) are endogenised as follows:

$$\Delta \mathrm{BLP} = \Delta \mathrm{BLP}\left(\underset{+}{\Delta Y}, \underline{\Delta R}, \underline{MP1}\right) \tag{6.16}$$

where the variables are as defined above, with the exception for the variable $MP1$, which stands for monetary policy variables such as credit-rationing by the authorities; this is of course in addition to changes in the rate of interest, which as Eq. (6.18) below shows, it is influenced by changes in the bank rate.

$$\Delta \mathrm{BDP} = \Delta \mathrm{BDP}\left(\underset{+}{\Delta Y}, \underline{\Delta R}, \underline{MP2}\right) \tag{6.17}$$

with the variables as defined above, with the exception of $MP2$ that stands for monetary policy variables.

Both MP1 (Eq. 6.16) and MP2 (Eq. 6.17) can be thought of as financial-stability policy variables as discussed in Sect. 3.3.

$$\Delta R = \Delta R\left(\underset{+}{\Delta BR}, \underset{+}{\Delta EF}, \underset{+}{\Delta PDC}\right) \tag{6.18}$$

where in addition to the variables as defined above, ΔBR stands for changes in the bank rate, ΔEF stands for changes in external financing and ΔPDC that stands for sales of public debt to the non-bank public.

Block IV: Government Sector

$$\Delta \mathrm{BLG} = \mathrm{PSBR} + \Delta EF - \Delta \mathrm{PDC} \tag{6.19}$$

where the variables are defined as above with the exception of PSBR that stands for the public sector borrowing requirement.

PSBR, as portrayed in Eq. (6.20), is simply defined as the difference between government expenditure (G) and tax revenues (T) along with other government revenues (OGR).

$$\mathrm{PSBR} = G - T - \mathrm{OGR} \tag{6.20}$$

We treat OGR as exogenous and hypothesise G and T to be determined as shown in Eqs. (6.21) and (6.23), respectively.

$$G = P_{\mathrm{G}} Q_{\mathrm{Q}} + WE_{\mathrm{G}} + UU_{\mathrm{B}} + ID \tag{6.21}$$

where the symbols are defined as follows, with the exception of G, W and U that are defined as above: Q_{Q} denotes the amount of goods and services bought by the government, with P_{G} being their prices, E_{G} stands for the number of employees in the government sector, U_{B} is unemployment benefits and ID stands for interest payments on government debt. E_{G} is defined as in Eq. (6.22):

$$E_{\mathrm{G}} = E - E_{\mathrm{P}} - U \tag{6.22}$$

where E is total working population, as defined above for the purposes of Eq. (6.2), and E_{P} is employment in the private sector. Clearly $E_{\mathrm{G}} + E_{\mathrm{P}} = E$ that is total employment.

$$T = T\left(\underset{+}{Y}\right) \tag{6.23}$$

Block V: Open Economy Aspects

$$\Delta EF = \mathrm{CB} + \Delta KM - \mathrm{OEF} \tag{6.24}$$

where ΔEF is equal to the current balance of international payments (CB) plus changes in capital movements (ΔKM) minus other external financing (OEF); the latter variable includes external lending to the public sector plus domestic bank lending to the public sector in foreign currencies. We treat OEF as exogenous and endogenise CB and ΔKM.

$$\mathrm{CB} = NE + OCB = X\left(\underset{+}{WT}, \underline{RER}\right) - Q[\left(WE\left(1 - \underset{+}{tw}\right),, \left(1 - \underset{+}{t\pi}\right),, \underset{+}{RER}\right) + OCB \tag{6.25}$$

where CB and NE are as above, OCB stands for other earnings on foreign investments minus payments made to foreign investors and cash transfers, WT is world trade, RER is the real exchange rate (where the exchange rate is defined as foreign to domestic currency), with $WE(1 - tw)$ and $\Pi(1 - t\pi)$ being the income distribution terms as they influence imports; all these variables are in real terms.

Finally, Eqs. (6.26) and (6.27) define ΔKM and RER:

$$\Delta KM = \Delta KM\left[\left(\underset{+}{R / R_{\mathrm{W}}}\right), \underline{(er)^{\mathrm{e}}}\right] \tag{6.26}$$

where the variable, ratio of domestic interest rates (R) to world interest rates (R_{W}), is included, along with the expected rate of change of the nominal exchange rate variable, $(er)^{\mathrm{e}}$.

$$\mathrm{RER} = \mathrm{RER}\left[\left(\underset{+}{R / R_{\mathrm{W}}}\right),,, \underline{Y},,, \underset{+}{WT},,, \underset{+}{(er)^{\mathrm{e}}}\right] \tag{6.27}$$

where the variables are as defined above.

Notes

1. It should be noted that this contribution goes well beyond the New Consensus Macroeconomics (NCM) and its policy implication, namely, inflation targeting (see also Arestis 2007, 2009, 2010).
2. The five blocks discussed in Sect. 2 were originally suggested in Arestis (2013)—see, also, Arestis (2010) and Arestis and Sawyer (2010). The current contribution, though, relies more extensively on the relevant theoretical framework and economic policies.
3. There are of course exceptions as highlighted below. See, for example, Onaran and Galanis (2013) and Lavoie and Stockhammer (2013).

4. It is interesting to note that at the G20 meeting in Hangzhou, China (September 2016), the managing director of the IMF praised the group's commitment to reduce excessive inequality and to ensure growth would be widely shared.
5. See, also, The Economist (2014) where it is argued that redistribution does help to increase national income.
6. An important aspect on this score is the change in labour markets over the recent years, where the role of trade unions has diminished substantially. Re-strengthening the role of trade unions is crucial in terms of reducing inequality.
7. Available at: http://www.bankofengland.co.uk/financialstability/pages/fpc/default.aspx
8. Available at: http://en.wikipedia.org/wiki/Financial_Conduct_Authority
9. Available at: http://www.bankofengland.co.uk/PRA/Pages/default.aspx
10. The Dodd-Frank Act of July 2010 in the USA (especially its relevant rule that does not allow banks to use secured deposits for speculative activities), the UK Vickers Report (which proposes ring-fencing retail bank deposits), the European Liikanen Report (which proposes ring-fencing the commercial bank activities), the IMF proposal to tax banks and the Basle III proposal to increase banks' equity in relation to their risk-weighted assets (RWA) are additional proposals around the globe.
11. Available at: https://www.federalreserve.gov/newsevents/speech/yellen20170825a.htm
12. Available at: https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp170524_1.en.html
13. However, see above on the slow progress on the financial stability 'new' objective.
14. The Bank of Japan pledged (end of September 2016) to overshoot its 2% inflation target and has also adopted a new tool to achieve it. This is to cap the ten-year bond yields at zero, in addition to its quantitative and qualitative and negative interest rate policies.
15. The ECB president has also argued at the conference of the European Systemic Risk Board in Frankfurt (available at: https://www.esrb.europa.eu/news/speeches/date/2016/html/sp160922.en.html) that the euro-area banks have become too big relative to the needs of the economy. Bank lending should flow to productive projects if the economy is to prosper. But bank lending in the euro area tends to be procyclical: growing too fast in the upswing and insufficiently in the downswing. Thereby banks in the euro area have not been helpful to the economy in the aftermath of the GFC.

References

Angelini, P., Nery, S., & Panetta, F. (2012). Monetary and macroprudential policies (Working paper series no. 1449). Frankfurt: European Central Bank.

Arestis, P. (2007). What is the new consensus in macroeconomics? In P. Arestis (Ed.), *Is there a new consensus in macroeconomics?* (pp. 22–42). Basingstoke: Palgrave Macmillan.

Arestis, P. (2009). New consensus macroeconomics and Keynesian critique. In E. Hein, T. Niechoj, & E. Stockhammer (Eds.), *Macroeconomic policies on shaky foundations – Whither mainstream economics?* (pp. 165–185). Marburg: Metropolis-Verlag.

Arestis, P. (2010). Economic policies after the new consensus macroeconomics. In S. Dullien, E. Hein, A. Truger, & T. van Treeck (Eds.), *The world economy in crisis – The return of Keynesianism?* (pp. 271–294). Marburg: Metropolis-Verlag.

Arestis, P. (2012). Fiscal policy: A strong macroeconomic role. *Review of Keynesian Economics, 1*(1), 93–108.

Arestis, P. (2013). Economic theory and policy: A coherent post-Keynesian approach. *European Journal of Economics and Economic Policies: Intervention (EJEEP), 10*(2), 243–255.

Arestis, P. (2015). Co-ordination of fiscal with monetary and financial stability policies can better cure unemployment. *Review of Keynesian Economics, 3*(2), 233–247.

Arestis, P. (2016). Main and contributory causes of the recent financial crisis and economic policy implications. In P. Arestis & M. Sawyer (Eds.), *Emerging economies during and after the Great Recession. International papers in political economy* (Annual ed.). Basingstoke: Palgrave Macmillan.

Arestis, P., & González-Martinez, A. R. (2016). Income inequality: Implications and relevant economic policies. *Panoeconomicus, 63*(1), 1–24.

Arestis, P., & Karakitsos, E. (2011). An analysis of the causes of the 'Great Recession' and some policy implications. In T. Niechoj, Ö. Onaran, E. Stockhammer, A. Truger, & T. van Treeck (Eds.), *Stabilising an unequal economy? Public debt, financial regulation, and income distribution*. Marburg: Metropolis-Verlag.

Arestis, P., & Karakitsos, E. (2013). *Financial stability in the aftermath of the 'Great Recession'*. Basingstoke: Palgrave Macmillan.

Arestis, P., & Sawyer, M. (2010). 21st century Keynesian economic policies. In P. Arestis & M. Sawyer (Eds.), *21st century Keynesian economics, international papers in political economy* (Annual ed.). Basingstoke: Palgrave Macmillan.

Arestis, P., & Sawyer, M. (2011). Economic theory and policies: New directions after neoliberalism. In P. Arestis & M. Sawyer (Eds.), *New economics as mainstream economics, International papers in political economy* (Annual ed.). Basingstoke: Palgrave Macmillan.

Atkinson, A. B. (2015). *Inequality: What can be done?* Cambridge, MA: Harvard University Press.

Atkinson, A. B., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature, 49*(1), 3–71.

Bank of England. (2009, November 19). *The role of macroprudential policy* (Bank of England discussion paper).

Bank of England. (2012, July 12). The distributional effects of asset purchases. *Bank of England Report*.

Bank of England. (2016, July). *Financial stability report: Executive summary*. Available at: http://www.bankofengland.co.uk/publications/Documents/fsr/2016/fsrjul16sum.pdf

Berg, A. G., & Ostry, J. D. (2011). *Inequality and unsustainable growth: Two sides of the same coin?* (IMF Staff discussion note 11/08). Washington, DC: International Monetary Fund.

Blinder, A., & Zandi, M. (2010). *How the Great Recession was brought to an end*. Available at: https://www.economy.com/mark-zandi/documents/End-of-Great-Recession.pdf

Cynamon, B. J., & Fazzari, S. M. (2014). *Inequality, the great recession and slow recovery*. Available at: https://doi.org/10.2139/ssrn.2205524

Eggertsson, G. B. (2006). *Fiscal multipliers and policy co-ordination* (Federal Reserve Bank of New York Staff Reports, No. 241). New York: Federal Reserve Bank of New York.

Eggertsson, G. B. (2011). What fiscal policy is effective at zero interest rates? *NBER Macroeconomics Annual 2010, 25*, 59–112.

Galbraith, J. K. (2012). *Inequality and instability*. Oxford: Oxford University Press.

International Monetary Fund (IMF). (2016, October). World economic outlook.

King, M. (2016). *The end of alchemy: Money, banking and the future of the global economy*. London: Little, Brown.

Korinek, A., & Kreamer, J. (2013). *The redistributive effects of financial deregulation* (NBER working paper 19572). Cambridge, MA: National Bureau of Economic Research.

Kumhof, M., & Rencière, R. (2010a). *Inequality, leverage and crises* (IMF working paper 10/268). Washington, DC: International Monetary Fund.

Kumhof, M., & Rencière, R. (2010b). Leveraging inequality. *Finance and Development, 47*(4), 28–31.

Kumhof, M., & Rencière, R. (2011). Unequal = indebted. *Finance and Development, 48*(3), 25–27.

Lagarde, C. (2014, May 27). *Economic inclusion and financial integrity*. Speech given at the conference on Inclusive Capitalism, London. Available at: http://www.inc-cap.com/IC_ESSAY_Book_Introduction_Keynotes.pdf

Lavoie, M., & Stockhammer, E. (Eds.). (2013). *Wage-led growth: An equitable strategy for economic recovery*. Basingstoke: Palgrave Macmillan/International Labour Organization.

Muinelo-Gallo, L., & Roca-Sagalés, O. (2011). Economic growth and inequality: The role of fiscal policies. *Australian Economic Papers, 50*(2–3), 74–97.

OECD. (2008). *Growing unequal? Income distribution and poverty in OECD countries*. OECD: Paris. isbn:978-92-64-044180-0.

Onaran, O., & Galanis, G. (2013). Is aggregate demand wage-led or profit-led? A global model, Chapter 3 in Lavoie and Stockhammer (2013).

Philippon, T. (2008, November). *The evolution of of the U.S. financial industry from 1860 to 2007* (Working paper). New York: New York University.

Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: Harvard University Press.

Saez, E. (2013, January 24). *Income inequality: Evidence and policy*. Lecture given for the ethics and wealth series. McCoy Family Center for Ethics in Society, Stanford University.

Sarin, N., & Summers, L. H. (2016, September 15–16). *Have big banks gotten safer?* Brookings Papers on Economic Activity, BPEA Conference Draft.

Stiglitz, J. E. (2013). *The price of inequality: How Today's divided society endangers our future*. London: Allen Lane/Penguin Books Ltd.

The Economist. (2014, March 1). *Social mobility: A memo to Obama, and inequality v growth*.

Turner, A. (2010, March 17). *What do banks do? What should they do and what public policies are needed to ensure best results for the real economy?* Speech given at the CASS Business School. Available at: http://www.fsa.gov.uk/pubs/speeches/at_17mar10.pdf

7

Is the Share of Income of the Top One Per cent Due to the Marginal Product of Labour or Managerial Power?

Marta R. M. Spreafico

1 Introduction

For a long time, there has been little interest among most neoclassical economists about issues of income and wealth inequality. As Milanovic (2013) dramatically put it, "Before the global crisis, income inequality was relegated to the underworld of economics. The motives of those who studied it were impugned. According to Martin Feldstein, the former head of Reagan's Council of Economic Advisors, such people have been motivated by envy. Robert Lucas, a Nobel prize winner, thought that nothing [is] as poisonous to sound economics as 'to focus on questions of distribution.' " Yet recently, issues posed by income (and wealth) inequality have been highlighted in a number of recent books, inter alios, by Stiglitz (2012), Deaton (2013), Piketty (2014) and Atkinson (2015).

Undoubtedly, Piketty (2014) has attracted the most attention and controversy. One of the reasons for the notable impact of *Capital in the Twenty First Century* is that it focuses on the great increase in the share of

M. R. M. Spreafico (✉)
Institute of Economic Policy, Università Cattolica del Sacro Cuore, Milano, Italy

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_7

the top one per cent in income and wealth over the last 30 years, or so, in the USA, the UK, Australia and Canada. The increase has been much less in the other developed European countries. This emphasis on the change in share of the top one per cent presents a much more dramatic picture of the increase in inequality than, say, a substantial change in inequality recorded by the Gini coefficient.

The figures for the increase in income inequality for the USA over the last three decades are remarkable. The labour compensation of the top one per cent over the period 1979–2007 accounted for 60 per cent of the growth of market-based incomes (38 per cent of post-tax incomes) (Bivens and Mishel 2013). The income of the top one per cent is largely driven by the earnings of chief executive officers (CEOs), not only because they comprise a substantial proportion of the top one per cent but also because there is a comparability effect on the salaries of the other top earners. So, consequently I shall largely concentrate on the pay of the CEOs.

In the USA, over the period 1965–2013, the remuneration of the average CEOs' annual income increased from just over $800,000 to $15.3 million in 2013 (Mishel and Davis 2014). The ratio of the pay of the average CEO to that of the average worker was 20:1 in 1965, peaking at 383:1 in 2000 and is nearly 300:1 in 2013. In the UK, the FTSE 100 senior executives today earn 150 times that of their average employees; in 1998 the figure was about 50.

One reason why there has been little attention paid to issues of income inequality, with some exceptions, is that pay, including that of CEOs, is seen as being driven by market forces (Mankiw 2013). Individuals are paid their marginal products. Hence, both the salaries of individuals and the share of income going to labour are largely determined by the technological parameters underlying the aggregate production function. Hence, there is little need to consider the role of institutional factors such as how salaries are determined or the influence of sociological factors or social norms.

However, there are severe theoretical and empirical problems underlying the aggregate production function that vitiates the marginal productivity theory of distribution. This will be discussed after considering Mankiw (2013). Piketty (2014) is sceptical of the relevance of the marginal productivity theory for the determination of the salaries of the top one per cent. Nevertheless, he still considers it applicable for explaining the pay of those undertaking 'replicable' work, such as a fast-food server. I show, by means of a simple example that this still concedes too much to the marginal productivity theory of distribution. I next briefly discuss the problems surrounding the existence of the aggregate production, especially the remarkable work of John McCombie and his colleague, Jesus Felipe.[1] Next, given the rejection of the marginal productivity theory of distribution, I consider how CEO pay is determined in practice. I discuss the way that the attempt to solve the principal–agent problem has paradoxically substantially increased the relative income share of the top one per cent. Finally, I analyse the 'managerial power approach' associated principally with the work of Bebchuk and Fried (2004). The last section summarises and concludes.

2 Why Should Income Inequality Be a Matter of Concern? Are Not CEOs Paid Their Marginal Products?

The neoclassical standard explanation of how factors of production are rewarded has been developed from Ricardo's model of distribution by applying the marginal principle to all factors of production and not just to land (Kaldor 1955–1956). Although the early models concerned themselves with homogeneous labour, it is a small step to apply this methodology at the microeconomic level to individuals.

Consequently, in a nutshell, those workers with higher productivities earn higher incomes that reflect their greater contribution to society. This is determined solely by the technical conditions of production and factors affecting the supply of labour. As Clark (1899) wrote many years ago, "[i]t is the purpose of this work to show that the distribution of income to society is controlled by a natural law, and that this law, if it worked without friction, would give to every agent of production the amount of wealth which that agent creates" (p. v). While Clark's statement does not imply that this is what every agent necessarily *ought* to get, it is often implicitly assumed that this is the case (Mankiw 2013). Moreover, the implication is that any attempt to alter the free market distribution of earnings will lead to a 'great contradiction', as Okun (1977) termed it, namely a trade-off between equity and efficiency. As altering the distribution of income is likely to reduce the efficiency of the allocation of resources, it, therefore, comes at an economic cost.

A recent statement defending the present distribution of the income of the top one per cent along these lines, albeit with some minor qualifications, is that of Mankiw (2013). Mankiw believes that in a competitive economy individuals are paid their marginal products. For example, in outlining what he sees as the criticism of what he describes as the 'left', he writes as follows: "In the standard competitive labor market, a person's earnings equal the value of his or her marginal product" (p. 32). The normative implications of this are made explicit when he attempts to defend the earnings of the top one per cent along the following lines of the ethical argument of 'just deserts'. "If the economy were described by a classical competitive equilibrium without any externalities or public goods, then every individual would earn the value of his or her marginal product, and there would be no need for government to alter the resulting income distribution" (Mankiw 2013, p. 32).

Consequently, this may be taken as the neoclassical benchmark. The key, Mankiw (op. cit.) continues, is whether the earnings of the top one per cent reflect their higher (marginal) productivity or represent the extraction of rents. Indeed, he concedes that if the increase in the share of the top one per cent were attributable to successful rent-seeking, he would deplore it. He asserts that on his own reading of the evidence the earnings of the top one per cent, and their rapid growth over the last 30 years, is due to their increased productivity.

The evidence Mankiw (2013) offers in support of this is not compelling. He invokes the superstar theory that "changes in technology have allowed a small number of highly educated and exceptionally talented individuals to command superstar incomes in ways that were not possible a generation ago" (Mankiw 2013, p. 13). As an example of this, he cites Steve Jobs of Apple and the authoress J.K. Rowling. However, their large incomes are heavily dependent on institutions set up by governments in the form of patents, copyright monopolies and, in the case of Jobs, US state expenditure on R&D (Mazzucato 2013), all of which are the antithesis of the free market. Moreover, such huge salaries are not necessary to persuade individuals to make substantial contributions to society. Just think of the unsung heroes who developed the internet and indeed the role of the US government in facilitating it. Then, there is the Genotype project, which makes the results freely available to all, compared with the smaller project of the Celera Corporation, whose aim was to appropriate the private rents from advances in this area. One could go on almost indefinitely. Finally, the share of the top one per cent is dominated by CEOs and the finance sector, not talented innovators.

The second line of reasoning is that Mankiw argues that the increase in the share is due to the 'race between education and technology'. This is the hypothesis that skill-biased technical change has increased the demand for skilled relative to unskilled labour and has led to a college premium. This, according to the hypothesis and which is Mankiw's view, has led to rising income inequality, which has nothing to do with rent-seeking, but is simply the operation of supply and demand for labour. Mankiw argues that, while Goldin and Katz (2009) concentrate on the full distribution of income rather than the top one per cent, 'it is natural to suspect that similar forces are at work'. The share of the top one per cent is considered to follow a similar U-shaped pattern over time similar to the skill–unskilled wage differential. However, unfortunately for this explanation, the college premium flattened out in the 1990s, while the growth of the share of the top one per cent was accelerated and bears little resemblance to the path of the college premium. Moreover, the skill-biased explanation cannot explain the fact that there has also been a rapid increase in the share of the top one per cent in capital income (Mishel and Davis 2014). The hypothesis of skill-biased technical change is predicated upon the existence of a well-behaved CES production function and the indirect measure of different types of technical change.[2] I shall question the foundations of the aggregate production function below.

However, for neoclassical economists, the existence of the concept of the marginal product of labour and the necessary adjunct of the (aggregate) production function is taken as axiomatic. In the language of Lakatos (1970), the latter is part of the 'hard core' or, in Kuhnian (1970) terms, it is a paradigmatic heuristic. The role of the marginal product of labour in determining pay is taken for granted and is deemed untestable by fiat. Consequently, the mainstream view has been that income inequality and its changes are not major issues. The former merely reflects differences in the marginal productivities of labour. Moreover, the decline in labour's aggregate share, which has been observed in many advanced countries, is explained solely in terms of the aggregate production function and the value of the elasticity of substitution, together with changes in the capital-output ratio.

3 On Piketty's 'Illusion of Marginal Productivity'

It is difficult to discuss changes in wealth or income inequality without mentioning Piketty's (2014) influential *Capital in the Twenty-First Century.* Piketty (2014) is rightly extremely sceptical of the concept of marginal productivity as an explanation for the determination of wages and salaries of the top one per cent. The hedge fund manager, for example, Paulson earned $3.7 billion in 2007 (Rajan 2010, p. 80). Was this his marginal product? How do we test this proposition? Should the marginal products of a handful of CEOs of the banks that precipitated the Great Recession be regarded as substantially negative over this period? It is worth citing Piketty (2014):

> To my mind, the most convincing explanation for the explosion of the very top US incomes is the following. As noted, the vast majority of top earners are senior managers of large firms. It is rather naïve to seek an objective basis for their high salaries in individual "productivity". When a job is replicable, as in the case of an assembly-line worker or fast food server, we can give an approximate estimate of the "marginal product" that would be realized by adding one additional worker or waiter (albeit with a considerable margin of error in our estimate). But when an individual's job functions are unique, or nearly so, then the margin of error is much greater. Indeed, once we introduce the hypothesis of imperfect competition into standard economic models (eminently justifiable in this context), the very "individual marginal productivity" becomes hard to define. *In fact, it becomes something*

close to a pure ideological construct on the basis of which justification for higher status can be elaborated. (pp. 330–331; emphasis added)

What is interesting here is that although Piketty dismisses the concept of marginal productivity for senior managers and executives, he seems to consider that theoretically it can be measured for those doing 'replicable' jobs, albeit imprecisely. This seems a somewhat contradictory position. As the top one per cent took the vast majority of the increase in income over the last 30 years in the USA, and this had nothing to do with their marginal productivity (which, as Piketty notes, cannot be independently measured), how could the remainder of the labour force be paid their marginal products? Nevertheless, it is a short step from Piketty's statement to assuming that for these employees with replicable jobs, competitive markets will ensure that they are paid the contribution they make to the economy. However, while the evidence discussed later provides support for Piketty's arguments regarding CEOs' pay, I shall argue that even for replicable jobs, the marginal productivity theory, *qua* a theory, is logically problematical.

To show what, in retrospect, may be seen to be a straightforward point, let us, following Piketty, take the example of a small restaurant managed by the owner. The manager has no idea of the elasticity of demand for his meals, and so undertakes a mark-up pricing policy, a là Kalecki. Prices are determined by a mark-up on the unit costs of labour (the salaries of the waiters and chefs) and the ingredients of the meals together with the other capital costs (energy, rates, etc.). Consequently, total revenue is given by:

$$p_M M \equiv R \equiv (1+\pi)(wL+I) \tag{7.1}$$

where p_M is the price of a meal (M), R is total revenue and I is the value of the ingredients. The operating profit is equal to $\Pi \equiv \pi(wL + I)$. The mark-up is determined by the state of competition from other restaurants, the overall level of affluence in the local area and it is also influenced by a target for the level of profits. Nominal wages are assumed to be determined by the state of the local labour market. The contribution of value added of the restaurant to output as reported in the national income and product accounts (NIPA) is given by:

$$R - I \equiv Y \equiv wL + \Pi \equiv wL + \pi\left(wL + I\right) \tag{7.2}$$

Suppose the restaurant is flourishing and the manager considers it desirable to hire a new waiter to speed up the service, but for the sake of argument, the same number of meals is served. Under this pricing policy, the increase in value added (Y) in adding an extra employee, from Eq. (7.2), is *definitionally* equal to $\partial Y/\partial L = (1 + \pi)w$. So, if we interpret $\partial Y/\partial L$ as the marginal product of labour, we can see that it is less than the wage rate. This is because the hiring of the extra waiter, through the pricing policy, automatically increases profits at the same time. Consequently, Π is not held constant as L changes and as the neoclassical marginal productivity theory assumes. Of course, if the manager merely passes on the increased labour cost in the form of an increased price of the meal, then, from Eq. (7.2) and holding Π constant, by definition, $\partial Y/\partial L \equiv w$ (it should be noted that the greater price of the meal reflects its increased quality, which includes a better speed of service). But this is not the result of optimization using a well-behaved production function subject to a cost constraint. In fact, changes in the local labour market conditions (such as an increase in the minimum wage) that affect the wage rate of the waiter will also cause his/her supposed marginal productivity to change. But the causation runs from the wage rate to the putatively marginal productivity.[3]

It should be noted that this applies to a firm that is selling a marketed product to the private sector. But what about the large (public) sector of the economy where there is no independent measure of aggregate output? Much depends upon the way it is calculated. In the early national accounts, the output was just taken to be equal to the total labour compensation with an arbitrary adjustment for capital costs. In many cases, there are measures of physical outputs (such as the number of operations in hospitals, or number of trials in the judicial system, which can be used), but the problem still arises as to how to price or value them.

Attempts in the UK have been made to revise the output measures of government services after the Atkinson Review (2005), but insurmountable problems remain for the testing of marginal productivity.

It should be noted that the accounting identity, $Y \equiv wL + rK$, where Y is income, holds irrespective of the degree of state of competition, whether or not there are well-defined production functions and whether or not firms optimise. If this accounting identity is partially differentiated with respect to labour, we obtain $\partial Y/\partial L = w$ and $(\partial Y/\partial L)(L/Y) = wL/Y = a$ where a is labour's share. The expression $(\partial Y/\partial L)(L/Y) = \alpha$ is the neoclassical definition of labour's output elasticity and, under neoclassical production theory, is equal to the wage share if there are competitive markets, a well-behaved aggregate production function and factors are paid their marginal products. But from the definition of the national accounts, α must be *definitionally* equal to the wage share, a. This led Phelps Brown (1957) to comment that labour's output elasticity of the production function and the wage share "will be only two sides of the same coin" (p. 557).

On a more pragmatic note, Thurow (1975) in his 'A Do-it-Yourself Guide to Marginal Productivity' (pp. 211–230) raises some further problems that occur even if output can be valued independently of the inputs. Other questions include the problems posed by disequilibrium, uncertainty, the presence of increasing returns to scale, whether governments can in principle ever pay their employees according to their marginal productivity and to what extent income benefits influence monetary remuneration. As Adam Smith long ago pointed out, production is characterised by the division of labour. The decisions of, say, a CEO will be influenced by the quality of the decisions of his subordinates, and indeed the outcome of different views in the decision-making process. It makes little sense to try to identify the output of an individual in these and similar circumstances. Clearly, even ignoring the problems of the measurement of the monetary value of output independently of the value of wages, there are many other insuperable difficulties noted by Thurow (1975) in the way of providing an adequate test of the marginal productivity theory. These concerns are shared, inter alios, by Stiglitz (2012, p. 97).

4 The 'Illusion of the Aggregate Production Function'

It is somewhat paradoxical that Piketty, in spite of his reservations about the marginal productivity theory in explaining the wage rate, nevertheless at times explains the changes in the shares going to capital and labour in terms of an aggregate CES production function. Piketty notes that over the last 30 years or so, capital's share of income has risen in many countries while the ratio of capital to income has also increased. In terms of conventional neoclassical production theory, this change is simply explained in terms of an aggregate production function where capital and labour are paid their marginal products and the elasticity of substitution is greater than unity.[4] After discussing the effect of bargaining power on factor shares, this is soon ignored and Piketty considers the role of technology and the production function as an explanation for the changes in the functional distribution of income between capital and labour. However, Piketty's estimates of the capital stock, which are broadly defined, seem to be overstated and the capital-output ratio has fallen. This implies an elasticity of substitution of less than unity, which empirically seems to be the case (Chirinko 2008; Rowthorn 2014). However, in this approach, there is no role for changes in labour market polices, globalization and so on, to affect the functional distribution of income. It is all down to the technology of production. But is it?

4.1 The Cambridge Capital Theory Controversies and the Aggregation Problem

Although the aggregate production function is now used in neoclassical economics, there is a fundamental problem as to whether or not it exists. First, there is the question as to even when there are well-defined micro-production functions, these can be aggregated to give an aggregate production function. Fisher (2005), who has done more work on this problem than most, comes to the conclusion that micro-production functions cannot be successfully aggregated.[5]

Related to this, are the Cambridge capital theory controversies of the 1950s and 1960s. This debate was largely between Cambridge UK and Cambridge Massachusetts (MIT). The first issue centred on whether the theoretical concept of 'capital' as a factor of production had any meaning outside the highly restrictive one-commodity world. The upshot was that the answer was 'no'. This important debate between Cambridge, UK, and Cambridge, Massachusetts, has long been relegated to the history of economic thought, forgotten or treated as an esoteric debate in theory (Birner 2002). Samuelson (1962) published a paper where he purported to show that a production system with more than one technique of production could be represented by a one-commodity aggregate production function. The capital theory controversies, and they were entirely a matter of theory, proved that this construct was untenable. It was also shown that outside a one-commodity world, an increase in the wage rate was not necessarily associated with an increase in the capital-labour ratio ('capital reversing'). 'Reswitching' can also occur, which is when the same technique of production can be the most profitable at two different interest rates.[6] While even theoretical debates are rarely conclusive in economics, the force of the Cambridge (UK) critique was conceded by Samuelson (1966). However, results and implications of this debate have long been forgotten by most economists.

So why are aggregate production functions still so widely used?

4.2 Why Aggregate Production Functions 'Work'?

One reason is that aggregate production functions 'work', in that statistical estimations of them give plausible estimates of the parameters. As Solow once remarked to Fisher, "had Douglas found labor's share to be 25 per cent and capital's 75 per cent instead of the other way around, we would not now be discussing aggregate production functions" (Fisher 1971, p. 305).

Most neoclassical economists accept Friedman's (1953) methodological stance that the realism of the assumptions of a model does not matter, what is important is its predictive ability. Ever since Cobb and Douglas's (1928) seminal paper, many estimations of aggregate production functions have found good statistical fits with the estimated output elasticities close to the factor shares. This has been taken to show that the aggregation problem and the Cambridge capital controversies are empirically irrelevant. Furthermore, this statistical result is interpreted as an indirect confirmation that factors are paid their marginal products.[7]

However, a difficulty arises from the fact that the aggregate production function is an engineering relationship and should be expressed in *physical* terms (see, e.g., Ferguson 1971, p. 250). However, aggregate production functions are estimated using constant price value data for output and the capital stock, where the output and the capital stock are a constant-price value measure and the 'price' is a price deflator. The (erroneous) implication is that the results of the physical one-sector production function still follow through unaffected. The problem is that in practice the aggregate production function has to be estimated using constant-price value data for both output (confusingly, sometimes called the 'volume' of output) and the capital stock. The accounting identity $Y = wL + rK$ must hold for any state of competition, whether or not there are constant returns to scale and, importantly, even if the aggregate production function does not exist. If the identity is differentiated and then integrated at any point of time, then the result is a Cobb–Douglas relationship given by:

$$Y \equiv wl + rK \equiv Bw^{a} r^{1-a} L^{a} K^{(1-a)} \equiv AL^{a} K^{(1-a)} \tag{7.3}$$

where B is the constant of integration and a and $(1 - a)$ are the factor shares.[8] Equation (7.3) has no behavioural content at all. However, when cross-sectional observations are used in the statistical estimation of the Cobb–Douglas, a, $(1 - a)$, w and r may all differ. But generally if one were to estimate a putative Cobb–Douglas production function, the 'output elasticities' would be close to the factor shares, which would be misleadingly interpreted as confirming that factors of production are paid their marginal products. If the factor shares differ in the cross-sectional data, then the use of a Box–Cox transformation may suggest that a more flexible functional form, such as the CES relationship, may give a better statistical fit and approximation to the accounting identity.

What about estimates of aggregate production functions using time-series data? Following Felipe and McCombie (2013), we can express the argument as follows where the 'direction of causation' runs from the identity to the putative production function:

$$Y_t \equiv w_t L_t + r_t K_t \Rightarrow \hat{Y}_t \equiv a_t w_r + a_t \hat{L}_t + \left(1 - a_t\right) r_t + \left(1 - a_t\right) \hat{K}_t \Rightarrow$$

$$Y_t = F\left(K_t, L_t, t\right) \Rightarrow \text{Cobb Douglas; CES; translog production functions} \tag{7.4}$$

Expressing the accounting identity in growth rates may yield a variety of functional forms, depending upon how the factor shares vary over time, if in fact they do. For expositional ease, if the factor shares are constant, then the accounting identity may be expressed as:

$$Y_t \equiv A_0 e^{\lambda t} K_t^a L_t^{(1-a)} \tag{7.5}$$

where $\lambda = ar + (1 - a)w$, that is, the weighted growth of the rate of profit and the wage rate are constant. If this is the case, estimating the accounting identity will give a perfect fit to the supposed Cobb–Douglas production function. More generally, the identity will give a good fit to time-series data, provided the weighted logarithm of the wage rate and profit rate can be accurately proxied by a time trend. This will often have to be a non-linear function as the wage rate and the profit rate have a strong cyclical component. The use of a linear time trend can give such poor statistical results that it often gives the impression that a behavioural equation is being estimated. It should be noted that this critique does not just apply to the Cobb–Douglas production function. If the identity has changing factor shares due to, say, the relative change in the bargaining power of firms and workers due to globalization, a better transformation of the accounting identity may be given by a CES relationship as in Eq. (7.4) (Felipe and McCombie 2001; Simon 1979). What are the implications? The use of the aggregate production function to determine the output elasticities and, hence, indirectly test and often supposedly confirm the marginal productivity theory of distribution by comparing them to the factor shares is without foundation.

Piketty is aware of the limitations of the aggregate production function and the role of the paradigm in determining what the legitimate questions are.

> All economic concepts, irrespective of how 'scientific' they pretend to be, are intellectual constructions that are socially and historically determined, and which are often used to promote certain views, values or interests. […] In particular, the notion of the aggregate capital stock K and of an aggregate production function $Y = F(K, L)$ are highly abstract concepts. From time to time I refer to them. But I certainly do not believe that such gross oversimplified concepts can provide an adequate description of the production structure and the state of property and social relations for any society. (Piketty 2015, p. 70)

Given these conclusions, the logical step is to examine how the pay of, say, the top one per cent is determined in practice, looking at the institutional framework within which these salaries are determined. This involves using a completely different framework and discarding the neoclassical paradigm.

5 The Determination of the Pay of CEOs

The increase in overall inequality in incomes has generally been explained in terms of labour market forces; the increasing wage premium for college graduates, the effect of technical change on the increased demand for skills, the effect of globalization, and the weakening of labour and product market policies and institutions (OECD 2011; Autor 2014). But these explanations, such as those based on the supply and demand for skills, are not adequate to explain the rapid rise of the extreme top end of the earnings distribution. Table 7.1 shows the extraordinary increase in the ratio of CEO pay to the average worker's pay for the USA over the period 1965–2015. There is the rapid rise in the ratio from 1990 to 2000, followed by a sharp dip associated with the bursting of the dot.com bub-

Table 7.1 CEO-to-worker compensation ratio, 1965–2015 (selected years)

Year	CEO-to-worker compensation ratio
1965	20.0
1970	23.2
1975	25.1
1980	33.8
1985	45.9
1990	71.2
1995	122.6
2000	376.1
2005	308.0
2010	229.7
2013	303.1
2015	275.6

Source: Mishel and Schieder (2016)

ble, the recovery and then the short-term decline with the Great Recession. The first obvious problem with the marginal productivity explanation is that the rapid growth of CEO salaries since 1990 is not matched by any increase in the efficiency of firms or the growth of total output. According to the Bureau of Labor Statistics, the growth of US labour productivity over the period 1990–2000 was 2.2 per cent per annum, 2000–2007, 2.6 per cent per annum and 2007–2016, a mere 1.2 per cent.

The evidence seems to point to the fact that the increase in the share of the top tail of the distribution has been the result of rent extraction and the pay-setting institutions and not the working of competitive markets (Bivens and Mishel 2013).

Compelling evidence that these high salaries are largely rents is that the increase in the top one per cent in the USA has been mirrored in the UK, Australia and Canada, but not to such an extent in the other advanced countries, such as continental Europe, Korea and Japan. The experiences of Japan, Germany and Sweden, where the share of the top one per cent since the 1930s either depicts an L-shaped curve or is flat, are very different from those of the USA, and the UK, where the pattern of inequality follows a U-shaped curve. Alvaredo et al. (2013) suggest that different institutional arrangements and policies may be the reason why similar countries exhibit 'such diverging patterns' in inequality. They maintain that "purely technological stories based solely upon the supply and demand of skills can hardly explain such diverging patterns" (Alvaredo et al. 2013, p. 5).

Arguments in support of the contention that CEOs are paid their marginal products in competitive markets are unconvincing. Kaplan (2012) asks how is it that other groups such as private corporate lawyers, hedge fund investors and private equity investors have achieved equal significant increases? He further argues that CEO compensation has risen slower than the average incomes of the top *households*, an argument quoted with approval by Mankiw (2013). But as Bivens and Mishel (2013) and Mishel and Davis (2014) have shown, if one uses the earners and not households as the comparator, CEO compensation has risen faster. But even if Kaplan (2012) is correct, how does this necessarily demonstrate that top incomes are determined in a competitive market for talent? The rapid growth of their income could be largely the result of comparability with CEOs' remuneration and influenced by the fact that the pay determination of the top earners has changed since the mid-1970s.

Furthermore, in the USA and the UK, the rapid increases in the size and profits of the financial sector have driven up top salaries in this sector. In 2008, in the USA, the finance sector earned a quarter of GDP and 40 per cent of profits.[9] Philippon and Reshef (2012) have estimated that the most significant factor in determining wages in this sector just prior to the subprime crisis was deregulation. This led for a short time to an increase in this sector's profits, before the subprime crisis, through a rapid increase in leverage and risk taking, the latter caused by the extensive use of financial instruments such as Residential Backed Securities, Collateralized Debt Obligations and Credit Default Swaps on the Collateralized Debt Obligations. Philippon and Reshef (2012) find that the excess wage in finance, the difference between the amount employees earned in this industry, compared with the amount they are predicted to make, reached 40 per cent, which can largely be attributed to rents.

But clearly, to understand why CEOs' income has risen so dramatically, it is necessary to examine how their salaries are determined in practice. There is now great deal of evidence as to how top executives' pay is set in reality. As Bebchuk and Fried (2004, 2005) have shown, CEOs' salaries are determined by supposedly independent remuneration committees and directors on behalf of the shareholders. These committees, which can hardly be described as independent (Bebchuk and Fried 2004), are responsible for not only setting the base salary but also bonus schemes, such as stock options and restricted stock, to incentivise the CEO to act in the best interests of the shareholders (Conyon 2006).

There are basically two competing explanations as to whether this is successful. One view is that 'optimal contracts' have been introduced for CEOs, and other highly paid executives, and have largely solved the principal–agent problem. The other view is articulated by Bebchuk and Fried (2003, 2004, 2005) who dismiss the optimal contracts literature, referring to it disparagingly as the 'official story'.[10] Their central hypothesis is that the determination of executive pay is the result of a process of remuneration committee capture, whereby the CEOs succeed in setting their own compensation. Bebchuk and Fried (2004) call this process 'the managerial power approach', which is presented as a more convincing alternative to the optimal contracting theory.

According to the optimal contracting approach, CEOs earn what is termed their 'reservation utility', which is the remuneration that prevents them from quitting and going somewhere else. According to the managerial power approach, the CEO compensation is set as high as possible, subject to an 'outrage factor', which has changed for some reason over time. According to the principal–agent approach, the use of options and restricted shares, as a substantial part of a CEO's salary package, is seen as incentives given to solve an agency problem. CEOs' compensation is linked to the financial performance of their firms as reflected in their share valuation. According to the managerial power approach, whatever their rationale, options and restricted stock only transfer rents to executives and do *not* act as an incentive to get value-maximising strategies adopted.

Much of the impetus for the rapid increase in the use of stock options as a substantial part of CEOs' remuneration came from the work of two influential business economists, Jensen and Murphy (1990a, b). Under the standard belief that the best judge of the performance of corporations are financial markets, they encouraged the remuneration committees of companies to award CEOs high compensation (they thought that, at the time, CEOs were underpaid), using stock options in order to attract and retain the best and most talented individuals and to use monetary incentives to align the conflicting interests. This 'pay for performance' was seen as the best solution to the principal–agent problem. It aligns shareholders' and CEOs' interests because, so the argument goes, CEOs are rewarded only if they pursue the principals' interests, which will be reflected in the firms' share price.

This 'optimal contracting', which is aligned to the 'maximizing shareholder value' approach, has been widely adopted in the USA. The success of the management of the firms was to be judged largely, or solely, in terms to the share price of the firm. Typically, top executives have been given options to buy shares not at the then prevailing price, but at some time in the future, when the share price is likely to be higher, supposedly due to CEOs' efforts. It is notable that in 2004, on the basis of evidence of the actual effect of the stock options, Jensen et al. had a complete *volte face* and completely changed their minds.[11] However, by then, it was too late.

Consequently, we have an answer to the question posed above: what was the cause of the dramatic rise in CEOs' pay over the last 30 years or so? If one were to search for an, or indeed the most, important proximate factor in the growth of CEO pay relative to the mean wage, one need look no further than the widespread use of stock options. The use of stock options was introduced in addition to CEOs' salary as there was no corresponding reduction in the latter when the stock options were introduced. Starting from the 1980s, there is a high correlation between CEO's remuneration and stock prices. Table 7.1 shows the consequences of the move towards a much greater part of the remuneration of CEOs being tied up with stock options and, hence, being closely correlated in the value of the stock prices.

Table 7.2 reports the results of regressing the logarithm of CEO annual compensation on the logarithm of S&P Index series over the period 1965–2014. The regression results reveal the strong and statistically significant impact of the growth of the S&P index on that of the top executives' pay, with over 80 per cent of the variation of the latter explained.[12] The regression analysis starts by assessing the estimated impact of the lagged level of the S&P Index on the CEO's annual compensation both without and with a time trend (columns I and II). It is found that the time

Table 7.2 CEO's annual compensation and S&P 500 index (1965–2014). OLS regressions

	Dependent variable: *(ln) CEO compensation (in millions of 2014 dollars)*				
	I	II	III	IV	V
(ln) S&PIndex$_{t-1}$	1.7563*** (17.70)	0.8040*** (8.22)	0.5498*** (4.27)	0.6501*** (4.52)	0.5424*** (3.77)
*(ln) S&PIndex*tbreak$_{t-1}$*					0.0921*** (4.76)
Time trend		0.0492*** (15.52)	0.0419*** (8.25)	0.0501*** (9.33)	0.0501*** (8.88)
Structural break dummy			0.5275** (2.16)	59.58** (2.66)	74.20*** (3.58)
Time trend Structural break dummy*				−0.0296** (−2.64)	−0.0371*** (−3.56)
Constant	−10.22*** (−15.08)	−102.03*** (−17.46)	−86.04*** (−8.34)	−102.80*** (−9.29)	−102.14*** (−8.74)
N	49	49	49	49	49
R-squared	0.8184	0.9496	0.9574	0.9643	0.9692
F-test p value	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Own construction

Notes: Superscripts */**/*** denote 10, 5, and 1 per cent significance levels

Figures in parentheses are the t-statistics. The regressions are controlled for heteroskedasticity. *Structural break* is a dummy that is zero before 1993 and 1 otherwise; *time trend* structural break dummy* is an interaction variable that creates a dummy counter of 0 before the break and time-period number after the break.

trend is statistically significant and that the S&P index has affected positively the level of the CEO's pay, and is statistically significant. The same occurs even when we control for the structural break. Empirical tests reveal that there has been a structural break in 1993: before and after that, the autonomous growth of CEO compensation is positive and significant, and equal to five and two per cent per annum, respectively (columns III and IV). Finally, it is investigated whether there had been any change in the slope coefficient of the S&P index. It is found that the slope has changed and has increased after 1993, but by a small amount (column V).

Bebchuk and Grinstein (2005) run regressions attempting to explain the rapid rise of CEO compensation, over the period 1993–2003, but solely in terms of standard industry variables. They conclude that "the growth in pay levels has gone far beyond what can be explained by the changes in market cap and industry mix" (p. 302).

Why did performance-related pay prove ineffective, and merely led to rapid increases in CEOs' remuneration? The answer is that in the USA, the structure of a corporation is such that CEOs have enormous influence over the board of directors, who are supposed to be independent and to supervise the CEOs' conduct and remuneration. Directors often receive large direct and indirect benefits, which are largely at the CEOs' discretion. Moreover, there are often interlocking pay committees with CEOs being on each other's remuneration committees, even if at several times removed. Consequently, the CEOs' remunerations are effectively mutually determined. There are spillover effects into the public sector where large pay increases of the top managers are justified by reference to comparable private-sector pay, often judged merely by the size of the organization rather than any reference to its profitability (Bebchuk and Fried 2003).

Bebchuk and Fried (2004) analyse in detail the performance-related pay schemes, with a view to determining whether these resemble more the optimal contracting approach (according to the principal–agent theory) or the so-called managerial power approach. They found that the structure of the compensation schemes provides compelling evidence for the managerial power approach. Performance pay in the private sector is often linked to the *overall* increase in the value of the company's shares, not how the company performs *relative* to the stock market overall. Ideally, CEOs' compensation should reflect only the degree to which the company performance that has been affected by their actions. If the value of all shares increases, as it happens during a stock market boom, then additional compensation should go only to the CEOs of those companies whose stock prices rose more rapidly than the average. But this never occurs in practice. CEOs receive stock options with a fixed price and can achieve considerable payments for these, even if their stock increases less than the market (Bebchuk and Fried 2004). Moreover, many of the arrangements for CEOs' pay are far from transparent, which is the opposite of what one would expect if the principal–agent problem was to be minimised.

The remarkably small number of financial linkages that connect most of the world's international firms has been demonstrated by Vitali et al. (2011). They used complex network analysis to trace the cross-holdings between 43,060 transnational corporations and found that 147 of these companies had control of 40 per cent of the value of transnational corporations, and 737 had control of 80 per cent. It can be seen that this close interrelationship not only poses severe economic stability problems but also how a very small network of top managers could come to set their own salaries based on a circularity notion of comparability.[13]

In other words, according to the evidence, the rapid increase in CEOs' remuneration has been driven more by rent extraction than the result of a well-functioning competitive market for senior executives. Moreover, while changes in income distribution need not be a zero-sum game, there is overwhelming evidence that the rise in the share of the top one per cent has been at the expense of the remaining 99 per cent. The relationship between work effort and pay in the neoclassical schema (work is seen merely as a disutility) is over simplistic. Many CEOs and top earners gain a great deal of utility through the power and prestige of their positions, and it is doubtful whether their work effort would decrease if their earnings were taxed more or their salaries were less.

6 Summary and Conclusions

The last three or four decades have seen an explosion in the pay of not only the CEOs but also of managers in the non-private sector. What was once considered an unacceptable salary for the top earners compared to the average remuneration has now become commonplace. The whole question of the remuneration of top executives and managers is one that involves a consideration of how these payments are determined and social norms about what is acceptable. These social norms are not those of the society as a whole, but rather those involved in the determination of these salaries. Clearly, an important question is how these social norms (or moral outrage) are determined and how and why they change over time.

What is clear, however, is that any defence of the rapid increase in the earnings of the top one per cent based on the notion of marginal productivity by neoclassical economists and the concept of 'just deserts' is untenable. I have highlighted the theoretical and insurmountable problems concerning the marginal productivity theory of factor pricing and the related concept of the aggregate production function. But what is also telling is that for the neoclassical approach, grounded in the need for microfoundations and using extensively the individual representative agent, it is impossible to test whether the remuneration of a specific individual represents his or her contribution to society. The chapter has considered the way that CEOs are remunerated. It is clear that the rapid increase in their pay, and that of the top one per cent, represents a change in societal values and their managerial power; a concept that fits uncomfortably within neoclassical economics. In fact, the debate over the pay of the CEOs merely serves to emphasise the fact that the neoclassical approach, in relying on the marginal productivity theory of distribution, does not have a coherent theoretical explanation of wage determination.

Notes

1. See Felipe and McCombie (2013) for a compendium of their research.
2. See Solow (2014) for compelling criticisms of some of Mankiw's arguments.
3. Note that if prices are determined by a mark-up on unit labour costs, labour's share is given by $1/(1+\pi)$. The mark-up will be determined by the state of competition in both the product and the labour market.
4. In terms of the aggregate CES production function with constant returns to scale and factors paid their marginal products, capital's share equals $(1 - a) = \delta(K/Y)^{(\sigma - 1)/}\sigma t$ and σ is the elasticity of substitution.

5. For a more detailed discussion of the aggregation problem, see Fisher (1992) and Felipe and Fisher (2008).
6. See Cohen and Harcourt (2003) and Pasinetti and Scazzieri (2008) for useful summaries.
7. See, for example, Mankiw and Taylor (2008, p. 69) and Hoover (2012, pp. 326–331) for textbook justifications of this approach.
8. Note that this is different from the identity derived from neoclassical production theory where the value of output is pQ where p is the price in, say, £s per unit output. It is theoretically possible to recover the physical volume of output from this and theoretically estimate the production function in terms of physical units.
9. However, as Haldane et al. (2010) suggested, the conventional way that output of the finance sector is calculated in the NIPA is likely to have provided an overestimate in the run up to the subprime crisis.
10. See also the review of Bebchuk and Fried (2004) by Weisbach (2007).
11. See also Stout (2014).
12. The regression results pass all the usual diagnostic texts. Tests for structural breaks (Clemente–Montañés–Reyes and Zivot–Andrews unit root tests) reveal that a breakpoint in the (ln) CEO's annual compensation series occurred in 1993. Both the augmented Dickey–Fuller test (t-statistic equal to −3.462, 5% critical value being −2.955) and the Johansen tests for co-integration (t-statistic equal to 16.1067 for the null of no co-integration, 5 per cent critical value being 15.41; t-statistic equal to 0.4939 for the null hypothesis of at most one co-integrating equation, 5 per cent critical value being 3.76) reject the null hypothesis of no co-integration.
13. Bivens and Mishel (2013, pp. 63–71) and Alvaredo et al. (2013, pp. 9–11) present a more detailed discussion of linkages between individual CEOs remuneration.

References

Alvaredo, F., Atkinson, A. B., Piketty, T., & Saez, E. (2013). The top 1 percent in international and historical perspective. *Journal of Economic Perspectives, 27*(3), 3–20.

Atkinson, A. B. (2005). *The Atkinson review: Final report. Measurement of government output and productivity for the National Accounts*. London: Palgrave Macmillan.

Atkinson, A. B. (2015). *Inequality. What can be done?* Cambridge, MA: Harvard University Press.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the 'other 99 percent'. *Science, 344*(6186), 843–851.

Bebchuk, L. A., & Fried, J. M. (2003). Executive compensation as an agency problem. *Journal of Economic Perspectives, 17*(3), 71–92.

Bebchuk, L. A., & Fried, J. M. (2004). *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge, MA: Harvard University Press.

Bebchuk, L. A., & Fried, J. M. (2005). Pay without performance: Overview of the issues. *Journal of Applied Corporate Finance, 17*(4), 8–23.

Bebchuk, L. A., & Grinstein, Y. (2005). The growth of executive pay. *Oxford Review of Economic Policy, 21*(2), 283–303.

Birner, J. (2002). *The Cambridge controversies in capital theory: A study in the logic of theory development*. London: Routledge.

Bivens, J., & Mishel, L. (2013). The pay of corporate executives and financial professionals as evidence of rents in top 1 percent incomes. *Journal of Economic Perspectives, 27*(3), 57–78.

Chirinko, R. S. (2008). σ: The long and short of it. CESIFO working paper no. 2234. http://www.ifoinstitut.de/DocDL/cesifo1_wp2234.pdf

Clark, J. B. (1899). *The distribution of wealth*. New York: Macmillan.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review, 18*(1), 139–165. (Supplement).

Cohen, A. J., & Harcourt, G. C. (2003). Retrospectives: Whatever happened to the Cambridge capital theory controversies? *Journal of Economic Perspectives, 17*(1), 199–214.

Conyon, M. J. (2006). Executive compensation and incentives. *The Academy of Management Perspectives, 20*(1), 25–44.

Deaton, A. (2013). *The great escape: Health, wealth, and the origins of inequality*. Princeton: Princeton University Press.

Felipe, J., & Fisher, F. M. (2008). Aggregation (production). In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). New York: Palgrave Macmillan.

Felipe, J., & McCombie, J. S. L. (2001). The CES production function, the accounting identity, and Occam's razor. *Applied Economics, 33*(10), 1221–1232.

Felipe, J., & McCombie, J. S. L. (2013). *The aggregate production function and the measurement of technical change: 'Not even wrong'*. Cheltenham: Edward Elgar.

Ferguson, C. E. (1971). Capital theory up to date: A comment on Mrs Robinson's article. *The Canadian Journal of Economics/Revue Canadienne d'Economique, 4*(2), 250–254.

Fisher, F. M. (1971). Aggregation production functions and the explanation of wages: A simulation experiment. *The Review of Economics and Statistics, LIII*(4), 305–325.

Fisher, F. M. (1992). In J. Monz (Ed.), *Aggregation. Aggregate production functions and related topics*. London: Harvester Wheatsheaf.

Fisher, F. M. (2005). Aggregate production functions – A pervasive, but unpersuasive, fairytale. *Eastern Economic Journal, 31*(3), 489–491.

Friedman, M. (1953). The methodology of positive economics. In M. Friedman (Ed.), *Essays in positive economics*. Chicago: University of Chicago Press.

Goldin, C. D., & Katz, L. F. (2009). *The race between education and technology*. Cambridge, MA: Harvard University Press.

Haldane, A., Brennan, S., & Madouros, V. (2010). What is the contribution of the financial sector: Miracle or mirage? In A. Turner (Ed.), *The future of finance, the LSE report*. London: London School of Economics.

Hoover, K. D. (2012). *Applied intermediate macroeconomics*. Cambridge: Cambridge University Press.

Jensen, M. C., & Murphy, K. J. (1990a). Performance pay and top-management incentives. *Journal of Political Economy, 98*(2), 225–264.

Jensen, M. C., & Murphy, K. J. (1990b). CEO incentives: It's not how much you pay, but how. *Harvard Business Review, 68*(3), 138–149.

Jensen, M. C., Murphy, K. J., & Wruck, E. G. (2004). Remuneration: Where we've been, how we got to here, what are the problems, and how to fix them. ECGI working paper series in finance no. 44. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.642.7257&rep=rep1&type=pdf

Kaldor, N. (1955–1956). Alternative theories of distribution. *Review of Economic Studies*, *23*(2), 83–100.

Kaplan, S. N. (2012). *Executive compensation and corporate governance in the U.S.: Perceptions, facts and challenges* (NBER working paper series, no. w18395). Available at: http://www.nber.org/papers/w18395

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: Chicago University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.

Mankiw, N. G. (2013). Defending the one percent. *Journal of Economic Perspectives, 27*(3), 21–34.
Mankiw, N. G., & Taylor, J. P. (2008). *Macroeconomics* (European ed.). New York: Worth Publishers.
Mazzucato, M. (2013). *Debunking public versus private sector myths*. Institute for New Economic Thinking. Available at: https://ineteconomics.org/ideas-papers/blog/the-entrepreneurial-state-debunking-public-vs-private-sector-myths-1
Milanovic, B. (2013). Why income inequality is here to stay. *Harvard Business Review*. Available at: https://hbr.org/2013/01/why-income-inequality-is-here
Mishel, L., & Davis, A. (2014). CEO pay continues to rise as typical workers are paid less. Economic Policy Institute, Issue brief no. 380. http://www.epi.org/files/2014/ceo-pay-continues-to-rise.pdf
Mishel, L., & Schieder, J. (2016). Stock market headwinds meant less generous year for some CEOs. *Economic Policy Institute Report*. Available at: http://www.epi.org/files/pdf/109799.pdf
OECD. (2011). *Divided we stand. Why inequality keeps on rising*. Paris: OECD Publishing.
Okun, A. M. (1977). *Equality and efficiency: The big tradeoff*. Washington, DC: Brookings Institution.
Pasinetti, L. L., & Scazzieri, R. (2008). Capital theory (paradoxes). In S. N. Durlauf & L. E. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed.). London: Palgrave Macmillan.
Phelps Brown, E. H. (1957). The meaning of the fitted Cobb-Douglas function. *Quarterly Journal of Economics, 71*(4), 546–560.
Philippon, T., & Reshef, A. (2012). Wages and human capital in the US finance industry: 1909–2006. *Quarterly Journal of Economics, 127*(4), 1551–1609.
Piketty, T. (2014). *Capital in the twenty-first century*. Cambridge, MA: Harvard University Press.
Piketty, T. (2015). Putting distribution back at the center of economics: Reflections on 'capital in the twenty-first century'. *Journal of Economic Perspectives, 29*(1), 67–88.
Rajan, R. (2010). *Fault lines: How hidden fractures still threaten the world economy*. Princeton: Princeton University Press.
Rowthorn, R. (2014). A note on Piketty's capital in the twenty-first century. *Cambridge Journal of Economics, 38*(5), 1275–1284.
Samuelson, P. A. (1962). Parable and realism in capital theory: The surrogate production function. *The Review of Economic Studies, 29*(3), 193–206.

Samuelson, P. A. (1966). A summing up. *Quarterly Journal of Economics, 80*(4), 568–583.

Simon, H. A. (1979). On parsimonious explanations of production relations. *Scandinavian Journal of Economics, 81*(4), 459–474.

Solow, R. M. (2014). Correspondence, "the one percent". *Journal of Economic Perspectives, 28*(1), 243–244.

Stiglitz, J. E. (2012). *The price of inequality*. New York: W. W. Norton & Company.

Stout, L. A. (2014). Killing conscience: The unintended behavioral consequences of 'pay for performance'. *Journal of Corporation Law, 39*(3), 525–561.

Thurow, L. C. (1975). *Generating inequality*. London: Macmillan.

Vitali, S., Glattfelder, J. B., & Battiston, S. (2011). The network of global corporate control. *PLOS One*. Available at: http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0025995&representation=PDF

Weisbach, M. S. (2007). Optimal executive compensation versus managerial power: A review of Lucian Bebchuk and Jesse Fried's "Pay without performance: The unfulfilled promise of executive compensation". *Journal of Economic Literature, 45*(2), 419–428.

8

Macroeconomic Lessons from the Financialisation Process

Jesús Ferreiro

1 Introduction[1]

Although the financialisation process can be considered a historical and structural phenomenon, mainly in the most developed economies, it is true that in the more recent decades, mainly since the 1990s, it has suffered an unparalleled increase. This is due to the policies of domestic and international liberalisation and deregulation of financial markets. In this sense, a rising number of contributions argue, as we argue below, that the rising size of finance is the main cause of the financial and economic crises that emerged in the world economy, mainly in the developed economies, since 2007. This contribution, which is mainly based on the research carried out along the FESSUD research project, tries to collect, with the obvious limitation of space, the main macroeconomic lessons that can be extracted from the macroeconomic consequences generated

J. Ferreiro (✉)
Department of Applied Economics V, University of the Basque Country UPV/EHU, Bilbao, Spain

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_8

by the financialisation process in most developed and emerging market economies, with a special focus on the case of European economies.

This chapter is structured into three main sections. The first section will focus on the definition of the financialisation processes. The second section will focus on the consequences of the financialisation process on economic activity in general and on the activity carried out by particular sectors and agents. The third section will deal with the Great Recession (GR) in terms of an extended consensus on the key role played by the excessive growth of finances on the burst of the crisis. This study pays attention to the different impacts of the economic and financial crises on European countries and to the consequences generated on the management of macroeconomic policies, mainly in developed and European countries. The final section will be devoted to the consequences of financialisation on the European integration process. Finally, we summarise and conclude.

2 Definition of the Financialisation Process

The last decades have witnessed a fast growth of financial sectors not only in developed but also in emerging and developing economies. This expansion of financial sectors, financial institutions and financial products has given rise to what is labelled as the "financialisation" process. Not only does this concept encompass the rising size of the financial sector but, mainly, the rising influence of finances in non-financial agents' decision-making: "financialisation means the increasing role of financial motives, financial markets, financial actors and financial institutions in the operation of the domestic and international economies" (Epstein 2005, p. 3).

However, although we refer to a recent and typical of modern economies' phenomenon, "capitalist economies have always relied heavily on finance" (Brown et al. 2015, p. 6). Therefore, what actually defines, and is characteristic of, the current financialisation process is the fact that the influence of finances in the economic process, that is, in the processes of decision-making of private (financial and non-financial) and public agents, and in the political and social arenas, is significantly larger than in the past (Sawyer 2015, 2017).

For this reason, most studies about the financialisation process begin with a definition of this process that is mainly focused in the description (and further explanation) of the consequences of financialisation instead of a precise definition of this process. Thus, for instance, for Hein and van Treeck (2010), the main consequences of financialisation would be their effects on the objectives and constraints of corporations (financial and non-financial, mainly large corporations), with the consequent impact on corporate investment, the creation of new opportunities and risks for families due to the larger influence of wealth and indebtedness on households' decisions about consumption, and, lastly, the impact generated in income distribution due to the change in the power relations among shareholders, managers and workers.

Fine (2013) emphasises that the financialisation process has involved

> the phenomenal expansion of financial assets relative to real activity (…); the proliferation of types of assets, from derivatives to future markets (…); the absolute and relative expansion of speculative as opposed to or at the expense of real investment; as shift in the balance of productive to financial imperatives within the private sector whether financial or not; increasing inequality in income arising out of weight of financial rewards; consumer-led booms based on credit; the penetration of finance into ever more areas of economic and social life such as pensions, education,, health, and provision of economic and social infrastructure; the emergence of a neo-liberal culture of reliance upon markets and private capital and corresponding anti-statism despite the extent to which the rewards to private finance have in part derived from state finance itself (…) the continued role of the US dollar as world economy (…) And however financialisation is defined, its consequences have been perceived to be: reductions in overall levels and efficacy of real investment (…); prioritizing shareholder value, or financial worth, over other economic and social values; pushing of policies towards conservatism and commercialization in all respects; extending influence of finance, more broadly, both directly and indirectly, over economic and social policy; placing more aspects of economics and social life at the risk of volatility from financial instability and, conversely, places the economy and social life at risks of crisis from triggers within particular markets. (p. 6)

In a shorter, and more operative, way we can state that the main elements that define the financialisation process are (i) the rising weight and size of financial activities, sectors, institutions and products, in modern market economies; (ii) the rising size of indebtedness of private agents (families and financial and non-financial corporations); and (iii) the rising influence of financial variables on the non-financial private agents' decisions on their resources allocation processes.

Having said that, it is evident that although financialisation is a common process to all developed economies, among them the European ones, and to many emerging and developing economies, and that the main reasons of this process have been the widespread policies of liberalisation and deregulation of the financial system (Sawyer 2011, 2017; Stockhammer 2011; Tyson and McKinley 2014; Hein 2015), it must be recognised that the intensity and the consequences of this global process differ markedly among countries, leading to a variegated process of financialisation (Brown et al. 2015, 2017; Hein et al. 2016; Sawyer 2015, 2017).

The existence of a variegated financialisation process is a key element to study and define, as it was stipulated at the Description of Work (DoW) of the FESSUD Project the "main policy implications of the financial crisis of 2007/09 and the events leading to that crisis, and the policy recommendations coming from the work of FESSUD". Throughout the research developed in the FESSUD project, it was clear that the origins of the current financial and economic crisis must be found in the financialisation process. Also, and consequently, we cannot correctly understand the origins and the consequences of the GR if we do not pay the necessary attention to the huge development of finances that has taken place in the last decades and the consequent larger size of financial activities and the unparalleled larger size of the financial sector (Detzer and Herr 2014; Hein et al. 2015; Hein and Dodig 2014).

Although the elements that have triggered the financialisation of modern economies are common to all of them (the deregulation and liberalisation of financial sectors) and the consequences are similar (in terms of an unparalleled rising size of the financial sectors and agents), in each case country a number of different and specific elements converge that have contributed to defining the particular model of development and working of the financial systems and the relationships between the financial sector and the non-financial private agents, thus implying that the "processes of financialisation are not uniform across countries and time" (Sawyer 2015, p. 4).

3 Macroeconomic Implications of the Financialisation Process

Although financialisation processes are country- and time-specific, most studies share the opinion that these processes, mainly when the size and influence of financial markets exceeds a certain threshold, generate a set of negative macroeconomic consequences, with all of them operating in the same direction, although the intensity of these effects may differ.

It is commonly argued that one of the main features of the financialisation process is the huge increase in the size of financial sectors, and, consequently, in the size of financial assets and liabilities, usually measured as a percentage of gross domestic product (GDP). Ferreiro and Gómez (2016) analysed the evolution between the years 1999 and 2014 of the size of financial assets and liabilities in the countries that belong to the Eurozone. The analysis was carried out not only for the whole economy of each euro country but also for the main agents, namely, non-financial corporations, financial corporations, households, general government and the rest of the world. The objective of that paper was twofold. The first was to detect the existence of significant differences in the size of the financial balance sheets of the different institutional agents in the euro countries. The second was to analyse whether the differences in the macroeconomic performances of these countries were associated with the differences in the size and evolution of those financial balance sheets. Thus trying to check the hypothesis that the financialisation process (the larger sizes and increases in financial assets and liabilities) was associated to a better (or worse) macroeconomic performance.

The main findings of this study can be summarised in the following four conclusions. First, larger financial balance sheets are not associated with higher growth of economic activity in the euro countries. The authors analysed the existence of a relationship between the variation of GDP and the change recorded in the net financial assets, in the financial assets or in the financial liabilities, both for the whole period 1999–2014 and for the two sub-periods 1999–2008 and 2008–2014. The authors did not find any significant relationship between the aforementioned variables neither when a linear relationship was analysed nor when a quadratic relationship between the dependent and the explanatory variable was analysed. This result implies that the financialisation process was not a significant determinant of the expansion registered before the onset of the great financial crisis (GFC) and of the decline in economic activity registered during the GR.

The second conclusion was that more finance was not associated to a more intense fixed capital formation. The authors analysed whether the change (measured as a percentage of real GDP) of net financial assets, financial assets or financial liabilities recorded for each period in the euro countries had a significant effect on the evolution of gross fixed capital formation (GFCF) in the total economy. They did not find any significant relationship between the variation of the GFCF and the change recorded in the net financial assets, in the financial assets or in the financial liabilities, neither in the period 1999–2008 nor in the years 2008–2014.

A similar analysis was carried out for the investment of non-financial corporations. Checking whether the changes in the main components of the financial balance sheets of non-financial corporations in the Eurozone countries (i.e., the change measured as percentage of the GDP of net financial assets, financial assets and financial liabilities of non-financial corporations) had an impact on the investment made by this sector.

In this case, the conclusions reached in the study differed depending on the period analysed. Regarding the period 1999–2008, Ferreiro and Gómez (2016) did not find a significant relationship between the change in the size of real investment of non-financial corporations and the change in the size of net financial assets, financial assets or financial liabilities of these corporations. This result implies that the financialisation of non-financial corporations between 1999 and 2008 would have not been a significant determinant of the average increase of investments of non-financial corporations in the euro countries.

However, the authors detected a significant impact of the change of financial assets and liabilities in the years 2008–2014 on the investment of non-financial corporations. The analysis showed the existence of a quadratic relationship between the change of investment and the change in the size of financial assets of non-financial corporations. The sign of the coefficients implied that the change of financial assets had a positive but decreasing effect on the change of the gross capital formation made by the non-financial corporations, with the inflexion point being an increase of financial assets equivalent to 189.3 per cent of GDP. This result implied that over this figure, the rise of financial assets had a negative impact on GFCF of non-financial corporations. The sign and the values of the constant and the coefficients of financial assets implied that the increase in the size of financial assets contributed to compensating the declining trend in the investment of non-financial corporations. Indeed, the investment of non-financial corporations would have only grown in the countries where the financial assets of these companies increased above 39 per cent of GDP, something that happened during those years only in Ireland and the Netherlands.

In the case of the financial liabilities, the impact of this variable on the gross capital formation of non-financial corporations would have been positive, although the investment of non-financial corporations would have been positive only in those countries where liabilities grew above 114 per cent of GDP, something that happened only in Ireland (where the increase amounted to 308 per cent of GDP).

The above results imply that the deleveraging process carried out by non-financial corporations during the crisis was associated with a strong decline of GFCF by non-financial corporations.

Finally, Ferreiro and Gómez (2016) also analysed the influence of financialisation of the consumption and savings decisions made by households in the euro countries, concluding that larger financial balance assets in booms have led to larger households consumption but deleveraging (decline of financial liabilities) has led to a significant decline in consumption.

During the years of the GR, they did not find any significant relationship between the change in the size of households' private consumption and the change in the size of net financial assets, financial assets or financial liabilities of these agents. However, for the period 1999–2008, they did find a significant impact on consumption of the change of financial assets and liabilities. Ferreiro and Gómez (op. cit.) argued that the change of financial assets had a positive and rising effect on change of private consumption. Therefore, before the GR the declining rise of households' financial assets would have contributed to moderate the growth of private consumption. However, the evolution of households' financial liabilities would have had a larger impact on private consumption. Therefore, that result reinforced the idea that before the GR private consumption had been fuelled by the larger households' borrowing.

The authors also analysed the impact of the changes in the components of the financial balance sheets of households (i.e., the change measured as percentage of GDP of net financial assets, financial assets and financial liabilities of non-financial corporations) on households' savings rate. Before the year 2008 only a significant direct relationship between the change in households' gross savings rate and the change in net financial assets prevailed. The study concluded that, by itself, the decline in households' net financial assets would have generated a fall in the savings rate amounting to 1.3 percentage points (higher than the registered fall). However, when the determinants of the change in gross savings rate during the GR were analysed, the only significant relationship was with the change of households' financial assets. During the GR, only the change of the size of financial assets had a significant impact on the change of households' gross savings rate. Thus, the larger size of financial assets in the euro countries would have led to a decline of the gross savings rate in the years 2008–2014.

Financialisation processes are also associated with the existence of significant changes in income distribution, both in the personal and the functional distribution. Thus, financialisation is related to significant declines in the wage shares, that is, in the size of wages as a percentage of GDP. But, financialisation is also generating a more inegalitarian income distribution with a decline in the share of low earnings coming from a rising share of ten per cent or even one per cent of population.

It is important to emphasise that the causation relationship between an inegalitarian income distribution and the financialisation process is bidirectional. In this sense, it is frequently argued that the decline in real wages, mainly in low-paid workers, is leading to a rise in the size of borrowing held by low-paid earners. Thus, households would be funding a significant part of their consumption, including here the purchase of a house, on external indebtedness; thus contributing to rising the size of the liabilities held by households, and consequently, the financial balance sheets of households and financial (mostly, banking) institutions.

Not only does financialisation generate changes in primary income distribution but also in the secondary income distribution by affecting redistributive public policies. The last decades have witnessed a process of privatisation of services formerly supplied by public institutions, like housing or pensions. As a result, financialisation has made those public institutions lose leverages to modify secondary (available) income distribution.

This process has occurred in parallel to a decline in taxation and public spending. In the case of public revenues, more than an absolute decline in total revenues, financialisation has come with a change in the composition of total public revenues, in the form of a decline in direct taxation and a rise in indirect taxation. In the case of the public expenditures, there has been a decline in the items directly related to the welfare state and social policies. All in all, these changes have implied a decline in the capacity of states to alter the income distribution in a more egalitarian way.

In recent years, a number of studies have argued that the relationship between finance and economic growth is not a linear one; certain thresholds exist, above which a higher size of finances would exert a negative impact on economic activity and growth (e.g., Arcand et al. 2015; Bouis et al. 2013; Creel et al. 2014; Dabla-Norris and Srivisal 2013; Law and Singh 2014).

The final result of the financialisation process would have been the break of the presumed positive nexus between economic growth and financialisation; and, thus, nowadays, mainly in the case of developed countries, which are the economies with the largest size of finance, the excessive size (and growth) of the financial sector would have a negative impact on economic activity and economic growth (Arestis 2016).

The financialisation process, fuelled by financial liberalisation and deregulation, would also have come in parallel with a rising financial instability and a higher occurrence of banking and financial crisis (Saidi et al. 2017). This financial instability would have increased the possibility that episodes of stress generated in certain segment of the financial sector would get the category of systemic risks, affecting the whole financial sector. Moreover, the liberalisation and deregulation of international capital flows would have boosted the contagion effects of crisis arisen in certain countries. It is important to note that, as Creel et al. (2014, 2015) have shown, there is a negative relationship between financial stability and economic performance, and thus financial instability would be an element exerting a negative impact on economic activity and growth.

4 Lessons from the Great Recession

There is a widespread consensus, mainly among heterodox economists, on the idea that the excessive growth of the financial sector is the main determinant of the financial crisis that burst in 2007 and the subsequent GR and the later period of low economic growth that is affecting most developed economies, mainly in Europe (García-Arias et al. 2017). In this sense, this long-lasting period of stagnation offers a set of lessons that should be taken into account not only to get the needed economic recovery but, mainly, to avoid the existence of a new financial crisis of similar dimensions and consequences to that of the GFC.[2]

Perhaps, the main lesson from the GFC and the GR is that finance matters in all economies. Although in the 1970s and 1980s, when there was a wave of banking crisis in a high number of developed economies, the belief that these kind of crises were exclusive of developing countries was widely extended. Indeed, the recommendations and policy prescriptions from economists and international institutions focused on measures to deregulate and liberalise (domestically and internationally) the financial sector and institutions with the final objective of making them similar to those existing in developed economies, mainly in the USA, whose financial sector was considered the most efficient one in the whole world.

The GFC, however, proved to be a global phenomenon, affecting both developed and developing and emerging economies. However, its origin was in the financial sector of the USA, the most developed country in the world, and its impact, not only on the financial sector but also at the whole economic activity, has been more intense in the developed countries (mainly in Europe) than in emerging and developing countries (Carrasco et al. 2016; Ferreiro and Serrano 2011; Esteban et al. 2010). As Table 8.1 shows, only the GFC (with the exception of the Commonwealth of Independent States) led to a decline in economic activity in the group of advanced economies, mainly in the case of the European Union and the euro area.

In this sense, it is important to notice that the euro economies only returned to the level of GDP registered in the year 2007 in the year 2015. As a result, the GDP of the euro economies in the year 2016 was only 3.2 percentage points higher than in the year 2007. By comparison, the real GDP of the emerging markets and developing economies in 2016 was 56.4 percentage points higher than in 2007, and in the case of the emerging and developing Asian economies, the increase was even much higher: 89.1 percentage points higher.

Another lesson from the GR is that large financial and banking crises have an enormous impact on public finances. Besides the impact on public expenditures resulting from the public assistance to troubled financial and banking institutions, serious financial crises exert a significant impact on public budget balances and on the fiscal policy stance, usually measured as the changes in the public budget balances. Thus, the impact on public finances of the episodes of financial crises determines to a great extent the orientation of fiscal policies and in many cases leads to the implementation of measures of fiscal austerity directed to the adjustment of fiscal imbalances, what in turn has a depressing impact on the level of economic activity (Ferreiro et al. 2015; Ferreiro et al. 2016a).

Directly related to the former point, it must be emphasised that large financial and banking crises have a deep and long-lasting negative impact on economic activity. There are a number of different channels that explain the impact on the GDP of the systemic financial and banking crisis and the collapse of financial and credit markets through the consequences generated on private consumption and investment. Moreover, as mentioned in the previous paragraph, these effects can be exacerbated by the implementation of restrictive fiscal policies, which also contribute to dampen the economic activity. As a whole, these negative effects do not only imply a temporary decline in the real economic activity, but also

Table 8.1 Gross domestic product, constant prices (2007=100)

Group of countries	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
World	100	103.0	102.9	108.5	113.1	117.1	121.1	125.3	129.5	133.5
Advanced economies	100	100.1	96.7	99.7	101.4	102.6	103.9	106.0	108.3	110.1
Euro area	100	100.4	95.9	97.9	99.4	98.5	98.2	99.4	101.4	103.2
Major advanced economies (G7)	100	99.7	96.0	98.7	100.3	101.6	103.0	104.9	106.9	108.5
Other advanced economies (advanced economies excluding G7 and euro area)	100	101.7	100.8	106.7	110.3	112.7	115.4	118.8	121.2	123.8
European Union	100	100.6	96.3	98.3	100.1	99.7	100.0	101.7	104.1	106.1
Emerging market and developing economies	100	105.8	108.8	116.9	124.3	131.0	137.6	144.0	150.2	156.4
Commonwealth of Independent States	100	105.3	98.6	103.2	108.0	111.8	114.2	115.4	112.9	113.3
Emerging and developing Asia	100	107.2	115.3	126.4	136.3	145.9	156.0	166.6	177.8	189.1
Emerging and developing Europe	100	103.1	100.1	104.7	111.5	114.2	119.8	124.5	130.3	134.3
ASEAN-5	100	105.4	107.9	115.4	120.8	128.3	134.8	141.0	147.8	155.1
Latin America and the Caribbean	100	104.0	102.1	108.4	113.4	116.8	120.3	121.7	121.8	120.6
Middle East, North Africa, Afghanistan and Pakistan	100	104.8	106.3	111.4	116.3	122.5	125.4	128.8	132.3	137.5
Middle East and North Africa	100	104.8	106.3	111.6	116.6	123.0	125.6	128.9	132.2	137.2
Sub-Saharan Africa	100	105.9	110.1	117.7	123.6	129.0	135.8	142.7	147.5	149.6

Source: Our calculations based on World Economic Outlook Database, April 2007, accessed at 29 July 2017 (available at the IMF's website: http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/index.aspx) they can imply a decline in the long-term rates of economic growth—or, in other words, in the rates of growth of potential output.

A further relevant lesson from the GR has to do with the general strategy of macroeconomic policy. Since the 1990s, the economic authorities in most developed economies, mainly in Europe, adopted the axiom that price stability, in the form of a low and stable inflation rate, was a sufficient condition to achieve both financial stability and macroeconomic stability. The latter is understood as the achievement of a rate of economic growth equal to the rate of growth of potential output, in other words, as the absence of output gaps.

With this objective a rising number of central banks, not only in developed countries but also in developing and emerging market countries, adopted the strategy of the so-called inflation targeting. In this framework of monetary policy, central banks gained an increasing degree or independence from the political authorities, while price stability, at least in the medium term, was considered the main objective of the monetary policy. Moreover, the management of short-term interest rates by central banks adopted the role of the main tool of monetary policies.

The key role given to monetary policy implied the parallel downgrading of fiscal policies, being subordinated to monetary policies (Ferreiro et al. 2011). Thus, fiscal policy focused in this strategy on the removal of fiscal imbalances, that is, unsustainable levels of fiscal deficits and public debt, with the objective of generating a balanced public budget, or even a fiscal surplus, over the business cycle.

In any case, it must be noted that the very concept of macroeconomic policy has experienced a significant change. The final objective of macroeconomic policy is no longer to achieve a precise (high) rate of growth of GDP or a full employment level of economic activity. Macroeconomic policies are now implemented to avoid cyclical fluctuations of economic activity. However, these cyclical fluctuations are now defined, as previously mentioned, as the differences between the current levels of economic activity and the potential outputs.

Macroeconomic policies, or demand-side policies, by affecting aggregate demand only have a temporary impact, affecting economic activity on a short-time basis. In other words, the impact of these policies in the long term is minimal. Long-term economic activity, the potential output, is determined by factors related to the supply-side elements (existing levels of capital and labour inputs, technology, and certain institutions like those of the labour markets). These supply-side elements can only be influenced by structural policies, but not by fiscal or monetary measures. Consequently, macroeconomic policies do not have an impact in the long term on variables such as the rate of growth of GDP, the level of employment or full employment.[3]

As mentioned earlier, according to this theoretical framework, business cyclical fluctuations are generated by changes in the aggregate demand, basically by fluctuations in private consumption and investment decisions. In a world formed by rational agents, these fluctuations were explained by the mistakes in the inflation expectations of private agents (households and firms). Price stability, in the form of a low and stable inflation rate, favours the generation of correct inflation expectations, that is, inflation expectations of private agents, and in the absence of any inflationary surprise or unforeseeable events, would equal the inflation rate set as target by the central bank. In other words, price stability, by anchoring inflation expectations, would guarantee the absence of deviations of current economic activity from the potential output (zero output gap). This implies that price stability guarantees the macroeconomic stability, with the economy growing at the rate of growth of potential output and unemployment rate being that of the NAIRU (Carrasco and Ferreiro 2011, 2013a, b, 2014).

Furthermore, price stability would also guarantee financial stability, avoiding the existence of financial bubbles or an excessive growth of credit or monetary aggregates. As Montanaro (2016) argued, before the GFC there was the "prevailing belief that financial markets were naturally efficient and resilient, the pre-crisis consensus was that a low and stable inflation, together with 'light touch' micro-prudential supervision, was also the best way to deliver financial stability" (p. 4). Consequently, "monetary policy should not react to asset prices bubbles, except to the extent that they affect price stability, and should only intervene after the bubble had burst" (p. 4).

However, the GFC and the GR took place in a context of very low inflation, thus proving that, contrary to the widespread belief in the

Table 8.2 Inflation, end of period consumer prices (percentage change)

Group of countries	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
World	4.8	4.0	4.0	3.6	4.1	3.9	3.9	5.3	4.6	3.3	4.3	4.7	3.9	3.3	2.9	2.9	3.1
Advanced economies	2.5	1.4	2.1	1.6	2.4	2.5	1.9	3.1	1.6	1.1	1.8	2.7	1.7	1.3	0.7	0.5	1.5
Euro area	2.6	2.1	2.3	2.0	2.4	2.3	1.9	3.1	1.6	0.9	2.2	2.8	2.2	0.8	−0.2	0.2	1.1
Major advanced economies (G7)	2.3	1.1	1.9	1.5	2.4	2.4	1.8	3.0	1.2	1.1	1.6	2.6	1.6	1.3	0.8	0.5	1.6
Other advanced economies (advanced economies excluding G7 and euro area)	2.6	1.5	2.2	1.4	2.1	2.3	1.9	3.4	3.0	1.7	2.4	2.8	1.8	1.5	1.0	0.7	1.4
European Union	3.4	2.7	2.5	2.2	2.4	2.4	2.2	3.2	2.2	1.4	2.6	3.1	2.4	1.0	0.0	0.2	1.1
Emerging market and developing economies	8.2	7.5	6.6	6.2	6.2	5.5	6.2	7.5	7.8	5.3	6.6	6.5	5.8	5.0	4.7	4.7	4.4
Commonwealth of Independent States	21.7	16.9	13.2	11.3	11.3	10.3	9.2	13.0	13.8	8.6	8.8	8.7	6.3	6.1	11.4	14.0	6.5
Emerging and developing Asia	3.0	3.2	2.6	3.6	3.7	3.9	4.7	6.5	4.6	4.7	6.0	5.3	4.5	4.5	3.1	2.6	3.0
Emerging and developing Europe	25.9	29.6	14.0	10.1	7.1	5.6	6.0	7.0	6.7	4.9	5.4	6.7	4.9	3.8	3.4	3.9	4.2
ASEAN-5	5.6	6.3	5.5	3.4	5.4	9.1	5.2	5.1	8.2	3.3	5.5	4.9	3.5	5.5	4.8	2.0	2.6
Latin America and the Caribbean	8.3	6.2	8.6	6.9	6.5	4.8	3.7	4.9	6.7	3.3	5.0	5.3	4.4	4.5	5.0	6.2	4.6
Middle East, North Africa, Afghanistan and Pakistan	2.3	3.7	5.2	5.7	7.9	6.4	10.2	9.7	12.8	5.6	8.9	9.2	11.4	7.1	6.7	5.0	5.9
Middle East and North Africa	1.9	3.9	5.1	6.2	7.8	6.1	10.5	10.0	11.8	5.1	8.5	8.7	11.5	7.3	6.6	5.2	6.2
Sub-Saharan Africa	17.2	11.8	12.5	10.4	7.9	7.9	7.7	7.6	13.4	9.2	7.7	10.0	8.2	6.1	6.1	8.2	12.7

Source: World Economic Outlook Database, April 2007, accessed at 29 July 2017 (available at the IMF's website: http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/index.aspx)

1990s and 2006, price stability was neither a guarantee for macroeconomic stability nor for financial stability. Thus, as Table 8.2 shows, the inflation rate in the 2000s was very low, mainly in the case of the developed economies, whose inflation rate was slightly above the figure of 2 per cent.

This result implies that inflation targeting (or similar monetary policy strategies focused on price stability as the single or main objective of monetary policy) must be abandoned. Consequently, monetary authorities must adopt objectives of real macroeconomic variables (rate of growth of GDP, employment, unemployment, income distribution, etc.) and financial stability ones. In this sense, it is important to emphasise that fiscal and monetary policies must pay attention not only to real and monetary variables and (domestic and external) imbalances but also to financial imbalances: paying attention, first, to the evolution of the size of financial balance sheets of financial and non-financial corporations and of households and, second, to the size and growth of their components (Ferreiro and Gómez 2016).

As mentioned above, the relationship between the size of the financial system and the impact on macroeconomic performance is not a linear one (Arcand et al. 2015; Cecchetti and Kharroubi 2015; Cournède et al. 2015). Therefore, we need to know whether the size of the financial system has exceeded the threshold since from that point on a larger size of finances exerts a negative impact on economic activity and growth (and, obviously, welfare). If we focus on the case of the European Union countries, the answer would be that such a threshold has been exceeded in most of these economies, mainly in the most developed ones. Thus, in the year 2012, the (unweighted) mean of financial liabilities in the EMU-11 countries reached 1088 per cent of GDP, 672 per cent of GDP in the EMU-6 countries and 1421 per cent of GDP in the EU-10 countries. Moreover, this size of financial liabilities had increased during the GR, and thus, since the year 2008, the size of financial liabilities had increased in 114 percentage points of GDP in the EMU-11 countries and in 111 percentage points of GDP in EMU-6 countries, and had only declined in the case of the EU-10 countries where the size of financial liabilities had fallen in 70 percentage points of GDP (Carrasco et al. 2016).

Creel et al. (2015) have concluded that the financialisation process (identified with the financial depth) has not had a positive impact on economic growth in EU economies. For the authors, this result is explained by the fact that the level of financial depth in the European Union is so high that it has stopped generating positive effects on economic growth.

Accepting this conclusion implies that economic growth in the European Union would rise if the size of finances in the EU shrinks. In the short run, a widespread de-leveraging process and reduction of the size of financial balance sheets of households and financial and non-financial corporations can negatively affect economic activity, mainly if no offsetting measures are adopted. This impact is highly probable in the current situation. Thus, as Hein (2015) argues, "stagnation after big financial crises becomes likely when the balance sheets of economic units are not quickly cleaned, when the nominal wage anchor breaks, and when there is no big and longer stimulus by the government" (p. 9).

Nonetheless, from the above conclusion, it cannot be automatically inferred that a de-financialisation process implies a stimulus to economic growth. In other words, we are not defending the hypothesis of an expansionary de-leveraging or de-financialisation process. On the contrary, as Hein (op. cit.) argues, unless it comes with the proper offsetting measures, this process will unavoidable have a negative impact on economic activity in the short and perhaps medium term. Indeed, there would be doubts about the size of this offsetting impact. What we are actually arguing is that the de-leveraging process is a necessary condition to recover the path of high and sustained economic growth that allows the reach and maintenance of a level of economic activity compatible with full employment.

It is important to emphasise that large de-leveraging processes (i.e., the decline in the size of financial balance sheets), after a big financial crisis, have an even greater negative impact on economic activity unless offsetting measures are adopted, for instance, fast cleaning of financial balance sheets, nominal wage anchors or big fiscal stimulus.

5 Financialisation and European Integration

As far as international capital flows are concerned, they are at the origin of some national imbalances and they are a transmission channel of domestic (real and financial) shocks; consequently, there must be a coordinated strategy to reduce the size and volatility of international capital flows. The experience since the 1980s with financial crises in certain countries has proven that international capital flows are a powerful transmission mechanism of economic shocks. Thereby, the real and/or financial crisis episodes that emerged in certain countries become systemic. Moreover, it is widely argued that, for instance, in the case of the Eurozone, certain macroeconomic imbalances (like the surge and increase of external, i.e., current account imbalances, or the generation of fiscal imbalances) or even financial imbalances (e.g., the housing bubbles in countries like Ireland or Spain) generated in the peripheral countries are associated and explained by the huge capital inflows coming from third countries, in particular in the Eurozone from the core countries, like France, Germany or the Netherlands (Carrasco and Serrano 2014; Hein and Truger 2014; Carrasco and Peinado 2015; Dodig and Herr 2015). Therefore, to ensure national and global financial and economic stability, it is necessary to adopt measures at a global level to reduce the size and volatility of international capital flows, like the setting up of capital controls, tighter regulations of capital movements or the taxation of international capital transactions. However, this can only be made under the umbrella of a coordinated international strategy that encompasses the most significant developed and emerging economies.

It is important to note that the aforementioned problems are exacerbated in a context of open economies. Thus, in an environment of intense internationalisation and globalisation, an appropriate coordination of national economic policies becomes essential to guarantee a harmonious and sustained global economic growth. This coordination is even more necessary in those economies where interactions are so large that put limits to the effectiveness of economic policy measures unilaterally (domestically) implemented. In the case of monetary integration processes, like the Eurozone, for instance, this coordination is more necessary because, along with the freedom of capital and goods-services movements, joins the disappearance of the exchange rates and the existence of a single monetary policy for all the member states of the monetary union.

Furthermore, the experience of the Eurozone shows, first, that a monetary union alone does not lead to a real convergence process among member states and, second, that member economies can suffer asymmetric shocks, with the result that individual economies can be at the same time operating in different phases of the business cycle and/or that the intensity (depth and duration) of national shocks may significantly be different. This implies that in a monetary union there is no guarantee that the national business cycle is synchronised, a problem for the implementation of countercyclical macroeconomic policies, mainly in the case of the single monetary policy, which in this case can operate in a procyclical way in some countries. Obviously, if the freedom of capital movements within the monetary union can result in a lack of synchronisation of the national business cycle, the larger the capital flows, within the Eurozone, the weaker the effectiveness of monetary policy and the larger the domestic imbalances in the euro countries.

Indeed, financialisation alone (i.e., the complete liberalisation and deregulation of financial markets) has not produced a process of convergence or catching up of less developed economies. As Ferreiro et al. (2017) argue, since the creation of the European Monetary Union the existing real divergence among the euro countries has not declined; on the contrary, it has remained constant and even has increased in many parameters that show the macroeconomic performance of the euro economies. Actually, as has often been argued, the implementation of a single monetary policy, joined to the deregulation of financial markets and the liberalisation of capital movements among euro countries, has contributed to fuel domestic (inflation and assets bubbles) and external (current account imbalances) in those euro countries (mainly Southern countries) with the weakest macroeconomic fundamentals. Thus, the existence of these imbalances would be one of the main causes of the deepest impact of the GFC and the GR in these countries.

Therefore, the adjustment of domestic imbalances must be addressed in a coordinated way among all member states of a monetary union. This coordination implies that economic imbalances must be symmetrically defined, thus, leading to the adjustment of those imbalances in which the value of a variable is below the target value but also to the correction of those imbalances in which the value of the objective is above that target. Thus, for instance, countries with, for instance, high inflation rates (see footnote 4) of current account deficits must implement fiscal or wage policy measures to adjust these imbalances, at the same time that countries with lower inflation rates or current account surpluses must also adopt measures to correct them (e.g., implementing an expansionary fiscal policy or setting a wage growth guideline above productivity growth). In other words, in monetary unions there must be rules and norms that ensure a symmetric burden of the adjustment of macroeconomic imbalances.

Monetary unions and the European Monetary Union in particular are, therefore, an evident case that the free international movement of capital is an element that contributes to generate unsustainable growth strategies and to increase the size of economic imbalances. Furthermore, we cannot forget that the expansion of national financial systems is directly related to the existence of international capital movements that allow the acquisition of financial assets but also the higher indebtedness of national financial and non-financial agents. Lastly, we cannot forget that international capital flows are a powerful transmission mechanism of economic shocks, making the real and/or financial crisis episodes that arise in certain countries become systemic. This implies that monetary unions, in general, and the Eurozone, in particular, are not exempt from suffering contagion effects, like the Greek sovereign debt crisis proved. Nonetheless, the Greek crisis showed that the contagion effect was not generally affecting some economies, like Ireland, Portugal, Italy or Spain. It is important to note that these economies were not only those with the weakest macroeconomic imbalances (Carrasco and Ferreiro 2016), but also those that suffered before the crisis—with the highest increases in the size of their financial balance sheets and the deepest deterioration of the financial balance sheets of private agents, both financial and non-financial (households and corporations).

It is, therefore, important and necessary to ensure national and global financial and economic stability in a monetary union and also to adopt measures to reduce the size and volatility of international capital flows, like the setting up of capital controls, tighter regulations of capital movements or the taxation of international capital transactions. However, this can only be made under the umbrella of a coordinated international strategy that encompasses the most significant economies. This kind of measures can help to avoid an excessive size of financial balance sheets and the surge of internal and external financial balances that can generate real macroeconomic, domestic and external, imbalances.

6 Summary and Conclusions

The burst, first, of the GFC and, later, of the GR, has been a painful proof that the financialisation process, fuelled by an intense liberalisation and deregulation of financial markets, is a source of financial and real economic instability. In the current situation, it can be stated that the size of the financial markets, proxied by the size of the financial balance sheets of the total economy and those of the agents that form it, is excessive.

This excessive size of finance is not only generating a negative impact on economic growth but also leading to the appearance and rise of financial and real imbalances. It is only affecting negatively to the effectiveness of the traditional macroeconomic policies, such as fiscal and monetary policies. These problems are exacerbated in the case of highly integrated economies, mainly in the case of the countries belonging to a monetary union, as it is the case, for instance, of the European Monetary Union. Therefore, it is not an accident that the Eurozone, with the highest degree of economic and financial integration, but also with the highest size of the financial sector, has been the region of the planet that has suffered most deeply the negative consequences of the financial crisis that began in 2007.

It is, therefore, evident that in order to achieve a full recovery of the negative consequences of the GFC and the GR, it is necessary to implement a re-orientation of macroeconomic (fiscal and monetary) policies. Nonetheless, such a change in the general framework of macroeconomic policies would be useless unless it is accompanied by a re-regulation and rationalisation (downsizing) of the financial system.

Notes

1. The conclusions and arguments presented at this contribution are the result of the research carried out throughout the FESSUD research project (see www.fessud.eu). In particular, this contribution is based on Ferreiro (2016).
2. In a survey conducted to evaluate the foresights of a set of experts about the future of finance up to the year 2025, 88 per cent of experts estimated highly likely (i.e., with a probability above 50 per cent) the burst of a new financial crisis, whose origin will be at the non-banking financial sector (Ferreiro et al. 2016b).
3. It must be emphasised that in this scenario even the concept of full employment changes. The concept of full employment abandons its Keynesian meanings. That is, it is no longer defined as a low unemployment rate (say, three per cent of active population) or a situation in which any worker willing to work at the prevailing market wage has a job. In the New Consensus Macroeconomics terminology, full employment is identified as a non-accelerating inflation rate of unemployment (NAIRU) or non-accelerating wages rate of unemployment (NAWRU). That is, it is the unemployment rate compatible with the target of inflation rates, keeping stable such a rate. This implies that the labour market is at an equilibrium situation (a market-clearing equilibrium). Therefore, full employment, or the NAIRU, can exist with any unemployment rate, regardless how high it can look. If such a figure of unemployment is socially, politically or economically considered as excessive, then structural reforms in the labour market, making it more flexible, would have to be implemented.

References

Arcand, J. L., Berkes, E., & Panniza, U. (2015). Too much finance? *Journal of Economic Growth, 20*(2), 105–148.

Arestis, P. (2016). Financial liberalization, the finance-growth nexus, financial crises and policy implications. In P. Arestis & M. Sawyer (Eds.), *Financial liberalisation: Past, present and future. Annual edition of international papers in political economy*. Basingstoke: Palgrave Macmillan.

Bouis, R., Christensen, A. K., & Cournède, B. (2013). *Deleveraging: Challenges, progress and policies* (OECD Economics Department working papers, no. 1077). Paris: OECD.

Brown, A., Passarella, M. V., & Spencer, D. (2015). *The nature and variegation of financialisation: A cross-country comparison* (FESSUD working paper series, no. 127). Available online at www.fessud.eu

Brown, A., Spencer, D., & Pasarella, M. V. (2017). The extent and variegation of financialisation in Europe: A preliminary analysis. *Revista de Economía Mundial/Journal of World Economy, 46*, 49–50.

Carrasco, C. A., & Ferreiro, J. (2011). Inflation targeting and economic performance: The case of Mexico. *Panoeconomicus, 58*(5), 675–692.

Carrasco, C. A., & Ferreiro, J. (2013a). Inflation targeting and inflation expectations in Mexico. *Applied Economics, 45*(23), 3295–3304.

Carrasco, C. A., & Ferreiro, J. (2013b). Inflation targeting in Mexico. *Journal of Post Keynesian Economics, 35*(3), 341–372.

Carrasco, C. A., & Ferreiro, J. (2014). Latin American inflation differentials with the USA inflation: Does inflation targeting makes a difference? *Journal of Economic Policy Reform, 17*(1), 13–32.

Carrasco, C. A., & Ferreiro, J. (2016). *An analysis of the determinants of the impact of the great recession on the Eurozone countries* (FESSUD working paper series, no. 149). Available online at www.fessud.eu

Carrasco, C. A., & Peinado, P. (2015). On the origin of European imbalances in the context of European integration. *Panoeconomicus, 62*(2), 177–191.

Carrasco, C. A., & Serrano, F. (2014). *Global and European imbalances: A critical review* (FESSUD working paper series, no. 42). Available online at www.fessud.eu

Carrasco, C. A., Ferreiro, J., Gálvez, C., Gómez, C., & González, A. (2016). The impact of the financial and economic crises on European Union member states. In E. Hein, D. Dezter, & N. Dodig (Eds.), *Financialisation and the financial and economic crises. Country studies* (pp. 299–320). Cheltenham: Edward Elgar.

Cecchetti, S. G., & Kharroubi, E. (2015). *Why does financial sector growth crow out economic growth?* (Bank for International Settlements working papers, no. 490). Available online at https://www.bis.org/publ/work490.htm

Cournède, B., Denk, O., & Hoeller, P. (2015). *Finance and inclusive growth* (OECD economic policy paper, no. 14). Paris: OECD.

Creel, J., Hubert, P., & Labondance, F. (2014). *Financial stability and economic performance* (FESSUD working paper series, no. 35). Available online at www.fessud.eu

Creel, J., Hubert, P., & Labondance, F. (2015). Financial stability and economic performance in Europe. *Economic Modelling, 48*, 25–40.
Dabla-Norris, E., & Srivisal, N. (2013). *Revisiting the link between finance and macroeconomic volatility* (IMF working paper, WP/13/29). Available online at http://www.imf.org/en/Publications/WP/Issues/2016/12/31/Revisiting-the-Link-Between-Finance-andMacroeconomic-Volatility-40284
Detzer, D., & Herr, H. (2014). *Theories of financial crises* (FESSUD working paper series, no. 25). Available online at www.fessud.eu
Dodig, N., & Herr, H. H. (2015). Current account imbalances in the EMU: An assessment of official policy responses. *Panoeconomicus, 62*(2), 193–216.
Epstein, G. A. (Ed.). (2005). *Financialization in the world economy*. Cheltenham: Edward Elgar.
Esteban, M., Ferreiro, J., & Serrano, F. (2010). Financial liberalization, growth and financial crisis. In G. Fontana, J. S. L. McCombie, & M. Sawyer (Eds.), *Macroeconomics, finance and money. Essays in honour of Philip Arestis* (pp. 282–294). Basingstoke: Palgrave Macmillan.
Ferreiro, J. (2016, September 27–28). *Macroeconomic lessons for EU*. Speech given at the FESSUD final conference, Brussels.
Ferreiro, J., & Gómez, C. (2016). Financialisation and the financial balance sheets of economic sectors in the Eurozone. In P. Arestis & M. Sawyer (Eds.), *Financial liberalisation. Past, present and future* (pp. 87–131). Basingstoke: Palgrave Macmillan.
Ferreiro, J., & Serrano, F. (2011). The impact of the current crisis on emerging market and developing countries. In P. Arestis, R. Sobreira, & J. L. Oreiro (Eds.), *An assessment of the global impact of the financial crisis* (pp. 108–134). Basingstoke: Palgrave Macmillan.
Ferreiro, J., Garcia del Valle, T., Gómez, C., & Serrano, F. (2011). Is there a role for active fiscal policies? Supply-side and demand-side effects of fiscal policy. In P. Arestis (Ed.), *Microeconomics, macroeconomics and economic policy* (pp. 191–207). Basingstoke: Palgrave Macmillan.
Ferreiro, J., Gálvez, C., & González, A. (2015). Fiscal policies in the European Union during the crisis. *Panoeconomicus, 62*(2), 131–155.
Ferreiro, J., Gálvez, C., Gómez, C., & González, A. (2016a). Bank rescues and fiscal policy in the European Union during the Great Recession. *Panoeconomicus, 63*(2), 211–230.
Ferreiro, J., Gómez, C., & Landeta, J. (2016b). *The future expansion and proliferation of finance* (FESSUD working paper series, no. 160). Available online at www.fessud.eu

Ferreiro, J., Gálvez, C., Gómez, C., & González, A. (2017). Economic crisis and convergence in the Eurozone countries. *Panoeconomicus, 64*(2), 223–244.

Fine, B. (2013). *Towards a material culture of financialisation* (FESSUD working paper series, no. 15). Available online at www.fessud.eu

García-Arias, J., Horn, L., & Toporowski, J. (2017). Perspectives on financialisation and crises in Europe. *Revista de Economía Mundial/Journal of World Economy, 46*, 17–26.

Hein, E. (2015). *Causes and consequences of the financial crisis and the implications for a more resilient financial and economic system: Synthesis of FESSUD Work Package 3* (FESSUD working paper series, no. 128). Available online at www.fessud.eu

Hein, E., & Dodig, N. (2014). *Financialisation, distribution, growth and crises. Long-run tendencies*. (FESSUD working paper series, no. 23). Available online at www.fessud.eu

Hein, E., & Truger, A. (2014). Fiscal policy and rebalancing in the euro area: A critique of the German debt brake from a post-Keynesian perspective. *Panoeconomicus, 61*(1), 21–38.

Hein, E., & van Treeck, T. (2010). 'Financialisation' in post-Keynesian models of distribution and growth: A systematic review. In M. Setterfield (Ed.), *Handbook of alternative theories of economic growth* (pp. 277–292). Cheltenham: Edward Elgar.

Hein, E., Detzer, D., & Dodig, N. (Eds.). (2015). *The demise of finance-dominated capitalism. Explaining the financial and economic crises*. Cheltenham: Edward Elgar.

Hein, E., Detzer, D., & Dodig, N. (Eds.). (2016). *Financialisation and the financial and economic crises. Country studies*. Cheltenham: Edward Elgar.

Law, S. H., & Singh, N. (2014). Does too much finance harm economic growth? *Journal of Banking & Finance, 41*, 36–44.

Montanaro, E. (2016). *Central banks and financial supervision: New tendencies* (FESSUD working paper series, no. 134). Available online at www.fessud.eu

Saidi, H., Rachdi, H., & Mgadmi, N. (2017). Revisiting the role of governance and institutions in the impact of financial liberalization on economic growth using the PSTR model. *Panoeconomicus, 64*(3), 31–336.

Sawyer, M. (2011). Re-structuring the financial sector to reduce its burden on the economy. In P. Arestis, R. Sobreira, & J. L. Oreiro (Eds.), *The financial crisis. Origins and implications* (pp. 114–136). Basingstoke: Palgrave Macmillan.

Sawyer, M. (2015). *Financialisation, financial structures, economic performance and employment* (FESSUD working paper series, no. 93). Available online at www.fessud.eu

Sawyer, M. (2017). *Financialisation, economy, society and sustainable development: A overview* (FESSUD working paper series, no. 206). Available online at www.fessud.eu

Stockhammer, E. (2011). Neoliberalism, income distribution and the causes of the crisis. In P. Arestis, R. Sobreira, & J. L. Oreiro (Eds.), *The financial crisis. Origins and implications* (pp. 234–258). Basingstoke: Palgrave Macmillan.

Tyson, J., & McKinley, T. (2014). *Financialization and the developing world: Mapping the issues* (FESSUD working paper series, no. 38). Available online at www.fessud.eu

9

Financial Instability and Speculative Bubbles: Behavioural Insights and Policy Implications

Michelle Baddeley

1 Introduction

Common fallacies gained ground in the aftermath of the global financial crises of 2007/2008, including the fallacy that almost all economists were caught completely unawares, and that only a small handful foreshadowed the instability and crisis to come. Just one (and not the only) example of the ways in which this was an unwarranted indictment of academic economists as a whole, was a conference held in Gonville and Caius College, Cambridge in September 1999. At this conference a number of academic economists, including John McCombie and myself, presented papers around the broad theme of global economic and financial crises. For our contribution, John and I emphasised that speculative episodes are relatively common, and the financial crises that unfold in their wake are not anomalous. From the mid-twentieth century onwards, many economists, especially those from heterodox traditions, have built on the heritage of ideas from economists including John Maynard Keynes and Hyman

M. Baddeley (✉)
University of South Australia, Adelaide, SA, Australia

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_9

Minsky to unravel the economic, social and psychological foundations of speculation and financial instability.

Ours was not an original hypothesis. Economic historian Charles P Kindleberger, amongst others, emphasised this theme continually, and to general audiences, ever since the first edition of his book *Manias, Panics and Crashes: A History of Financial Crises*, published in 1978. He observed that speculative bubbles and financial crises, and not only in standard types of assets such as stocks, shares and currencies, were common from the seventeenth century onwards. Charles Mackay foreshadowed even Keynes in his histories of speculative and other frenzies in his 1841 book *Extraordinary Popular Delusions and the Madness of Crowds*. Our aim in 1999 was to show that speculative manias are not a new phenomenon, and also to address the revisionist interpretations from some economists, including Peter Garber, who argued that psychological explanations along Minsky-Kindleberger lines were not the simplest explanation possible. This was a development from previous empirical studies, for example, as outlined in Flood and Garber (1980), which had focussed on outlining empirical support for the hypothesis that speculative bubbles do not exist at all.

Various developments following the crisis, alongside the growth in the influence of behavioural finance, have meant that there is more consensus now than there used to be about the causes of speculative bubbles and the financial crises which often ensue. Developing the chapter that John McCombie and I wrote following the 1999 conference, in this chapter, I will outline some of the key ideas we explored at the time. Our aim then was to use the two historic episodes of Tulipmania and the South Sea Bubble to assess the interpretation that speculative bubbles can be explained as rational bubbles—where rationality is defined in the strict, mathematical sense associated with the rational expectations hypothesis. The years since 1999 have not only been marked by significant economic and financial instability, they are also the years marking the ascent of behavioural economics and finance. Our analysis did touch on some of the insights from Robert Shiller and others working in behavioural finance at the time. Robert Shiller's book, a general audience book, *Irrational Exuberance*, which outlines many of the key insights, was first published in 2000. Robert Shiller and others in behavioural finance have developed new, richer insights since then. This chapter will explore how new insights around behavioural finance can bring in a more substantial rejoinder to the rational expectations school hypothesis of rational speculative bubble. In the first part of this chapter, I will clarify the definition of a speculative bubble and then outline the key insights that John and I explored in our 2001 chapter. I will then re-assess our analysis in the light of recent developments in behavioural finance. I will conclude with a discussion of policy implications and lessons. Overall, this chapter will argue that the mismanagement of the financial crisis and the speculative episodes which preceded were not a reflection of all economists' ignorance or ineptitude. The problem was not that economists did not know or did not want the world to know some of the dangers of modern capitalism for international financial stability. The problem was, and possibly still is, that there was little consensus across the various 'tribes' that make up the population of academic economists and the 'tribes' that have the most political and commercial influence are not those who have explored the dangers.

2 What Is a Speculative Bubble?

Charles Kindleberger described speculative bubbles as fluctuations characterised by rapid price increases, followed by a more rapid collapse. Kindleberger (1978) defined these events, where an asset's price rises just because investors expect it to rise, as speculative bubbles. A more technical definition is when the price of the asset deviates from its fundamental value, where the fundamental value is defined in terms of the present value of the asset's earnings over its lifetime. A speculative bubble occurs when the asset's price follows any path that does not track the fundamental value of an asset; for example the fundamental value of a share will be driven by dividends growth, and a speculative share price bubble emerges when a share's price grows more rapidly than dividends growth.

There are a number of logistical problems with this definition, including problems around how to calculate the discount rate in present value calculations. But for speculative bubbles such as Tulipmania, there are more substantial logistical constraints because it is hard to see that a single tulip bulb, on its own, will be worth much at all—unless someone plants it in the ground and grows flowers from it and so on. There are too many uncertainties. Forecasting the value of a share over its lifetime is problematic but possible, at least in some circumstances. For the tulip bulbs of Tulipmania, it can be hard to even imagine what its fundamental value might be. Given more than one potential paths for asset price growth, assuming rational expectations and efficient markets, and only one tracking growth in fundamental value, every other path is consistent with a speculative bubble.

2.1 Rational Bubbles

In our 2001/2004 chapter we outlined different conceptions of bubbles from across the different economics disciplines (Baddeley and McCombie 2001/2004). The dominant paradigm then, and still to a lesser extent now, is the rational bubble paradigm, for example, as outlined by Blanchard and Watson (1982) who investigate the evolution of rational bubbles by exploring what is possible in theory, as opposed to what is likely in reality. In our chapter, John and I structured our theoretical analysis around three broad theories of speculation, popular at that time: the rational bubbles models, as championed by Garber (1989, 1990); the contagion bubbles that allowed in some social influences, though in a largely mathematical way; and the 'irrational' bubbles described in the heterodox literature. However, the word 'irrational' is too strong and too loaded a term to be of very much use. It does not follow logically that if something is not black, therefore it is white. Just because speculative activity is not rational in the very strict, confined ways in which some economists describe it, it does not follow that it is completely irrational either. Human behaviour is much more subtle and modern psychology allows that decisions and choices that might seem wrong-headed to a mathematical robot in fact reflect complex interactions between different thinking styles. This is true for our economic and financial decision-making too, as I will explore in a later section.

In our 2001/2004 chapter, we categorised these three theories of speculative bubbles according to different variants of assumptions about rationality, financial market efficiency and risk versus uncertainty. With risk, very broadly and simplistically, the emphasis is on a distinction between 'Knightian' quantifiable risk and 'Knightian' unquantifiable uncertainty. Divergent conceptions of risk and uncertainty underpin divergent explanations for bubbles which focus on starkly different ranges of factors as the catalysts to the genesis and subsequent collapse of speculative bubbles and manias. These different understandings of risk and uncertainty find their way through to the different models of bubbles. In the rational bubbles literature, investors are balancing risk that they can measure and match with their risk preferences.

2.2 Rational Expectations Bubbles

As noted above, our 2001/2004 chapter focussed on exploring the revisionist models of asset price fluctuations consistent with strong assumptions of rationality. Rational expectations theorists assume quantifiable risks, where subjective probability estimates coincide with an objective probability distribution. Two related assumptions are critical to this approach. In the revisionist literature the more nuanced hypothesis is that speculative bubbles do exist, but they are consistent with strong assumptions about rationality. The difference can be understood if we bring together rational expectations and efficient markets. The early analyses of Flood and Garber (1980) and others were focussed on the idea that financial markets efficiently process all new information and any differences between the observed and fundamental values of an asset which will be traded away by rational agents. These early models were essentially founded on the argument that speculative bubbles do not exist (e.g. see Flood and Garber 1980). This is a hard assertion to defend so in their revisionist explanations, Garber (1989, 1990) and others are allowing that financial markets are not informationally efficient but rational agents can nonetheless retain their strict form of rationality even though asset prices deviate from fundamental value. Early models assumed both strong form of rational expectations and information efficient financial markets; revisionist models relaxed the efficient markets hypothesis but retained the assumption of a strong form of rational expectations. Instead, Garber (1989, 1990) suggested that speculation is more easily explained as a rational response to changing risk and uncertainty.

In our 2001 chapter, we used the episodes of Tulipmania and the South Sea Bubble to test the power of rational bubbles models and theories, concluding that the rational bubbles argument was not consistent with the evidence. Part of our analysis was based on the foundations of theories of rational bubbles. Blanchard and Watson (1982) describe rational bubbles as just one of the possible outcomes in a stable, ergodic world in which decision-makers form rational expectations, where rationality is defined strictly in terms of optimising agents who are assumed to be able efficiently to process information and news, as if they are mathematical machines. Risk is assumed to be quantifiable and subjective estimates of probability are assumed to coincide with objective probability distributions and the data generating systems which govern reality. In the rational bubbles research, a speculative bubble is defined as a path in an asset price which diverges from the fundamental value of that asset. For a stock or share, its fundamental value will be the discounted value of expected future dividends and this fundamental value will follow a random walk in that all changes are unpredictable and that asset prices adjust quickly to all news. Expectations in a world of rational speculative bubbles parallel the rational choice decision-making associated with rational expectations macroeconomics. Asset markets are assumed to be populated with perfectly rational and identical agents, all adopting the same optimising decision-making rules. Risk is assumed to knowable and quantifiable and investors are assumed to be able to match these risk with their own stable risk preferences. These investors operate in a world that is ergodic, that is, it is immutable and changing only in response to exogenous shocks—paralleling the assumptions from real business cycle theory, the sister theory for business cycles. As noted above, they form subjective expectations, subjective in the sense that they are based on the information available to them at the time. These subjective expectations coincide with an objective probability distribution (Muth 1961). This world is ergodic, immutable, fixed and risk is assumed to be measurable. How can a speculative bubble emerge in such a world?

A number of rational bubble theorists, including Shleifer and Summers (1990) and Blanchard and Watson (1982), focus on the issue of timing. A speculative bubble persists because rational investors do not know for sure when the bubble will burst. Uncertainty plays a role not only reflecting uncertainty about the time path of the bubble but also because uncertainty about fundamentals is consistent with holding an asset that is not tracking fundamental value. Whilst the bubble continues to grow then speculators can rationally expect to make money if they are able to sell assets in a liquid market. Thus rational speculators can trade in speculative bubbles even whilst the asset's price deviates from its long-term fundamental value, *if* speculators have chances to sell assets quickly at a profit.

To capture this, Blanchard and Watson (op. cit.) assume that returns on assets will be driven to the point at which arbitrage will cease, assuming either finitely lived agents, successive generations of entrants or trading in a perpetuity. In these conditions, the return on the asset will be defined by the following condition:

$$R_t = \frac{p_{t+1} - p_t + x_t}{p_t} \tag{9.1}$$

where R is the return on the asset, p is the asset price and x is the dividend. For a given information set Ω commonly known by all investors, the expectations of return, conditional on the information set, will be:

$$E\left(R_t \middle| \Omega_t\right) = r \tag{9.2}$$

Taking the expectation of R_t and noting that p_t and x_t will be known at time t:

$$\frac{E\left(p_{t+1}\middle|\Omega_t\right)-p_t+x_t}{p_t}=r \tag{9.3}$$

It follows that:

$$E\left(p_{t+1}\middle|\Omega_t\right)-p_t+x_t=rp_t \tag{9.4}$$

Solving this gives p_t^* which is the present value of expected future dividends:

$$p_t^*=\sum_{\infty}^{i=0}\left[\frac{1}{1+r}\right]^{i+1}E(\cdot) \tag{9.5}$$

For this arbitrage condition, Blanchard and Watson (1982) show that rational bubbles are *a possibility* given that solutions to this condition take the general form:

$$p_t=p_t^*+c_t \tag{9.6}$$

where c captures deviations from fundamental value. It is the 'bubble' term in the evolution of the asset price. Note that this model allows that rational bubbles are mathematically possibile, not that they are likely. It also implies multiple equilibrium paths and all but one of these time paths will be a speculative bubble. The single time path when the asset price tracks the fundamental value is the non-bubble path. The model is also consistent with the idea that the c term will grow over time, and so bubbles will grow over time. This model suggests that the probability that the asset price will track the fundamental value is vanishingly small, which seems anomalous with the broader rational expectations approach.

Without further assumptions to narrow down the possible time paths, this model seems to imply that asset prices are almost always going to follow a speculative bubble time path. Whilst speculative bubbles are common, even heterodox economists would not claim that anything else is virtually impossible.

In terms of explaining why a bubble bursts, Blanchard and Watson (op. cit.) state that there is a probability π that the bubble will continue in any given period, and it follows that the probability of a crash is $(1-\pi)$. Depending on their risk preferences, some investors will continue to hold the asset whilst $\pi > 0$. The evolution of this probability will depend on how long the bubble has lasted and the extent of the deviation between the asset price and the fundamental value. So, if the asset price has risen far beyond fundamental value and/or the bubble has lasted a long time, then the asset price will have to rise even more to compensate new entrants to the market for the increasing risk that the bubble will collapse soon.

2.3 Bayesian Bubbles

An alternative model of rational bubbles, but one associated with weaker assumptions about rationality, connects with the herding and information cascade models of Banerjee (1992) and Bikhchandani et al. (1992, 1998), in which rational agents are not fully informed in the ways assumed by rational expectations theorists. Instead they are using Bayesian reasoning processes on the basis of limited information to infer probabilities via an application of Bayes's rule (see also Chamley 2003, for a survey of these models). The essential idea parallels some of Keynes's insights (as explored in more depth below) around the idea that people may sometimes assume that others around them know more than they do. When we observe other people's actions, we incorporate that information into our own information sets, and the balance of this social information about others' actions is balanced against private information we may have to update our prior probabilities. Bayesian herding and information cascade models found their way into behavioural finance, for example, in the work of Avery and Zemsky (1998) amongst many others (see, also, Devenow and Welch 1996; Drehmann et al. 2005). With information cascades and herding in financial models, speculators look at the price that other speculators are paying and use this information to update their own probabilistic expectations of what will happen to the asset price in the future. So speculators are still assumed to be strictly rational in the sense that they are aiming to maximise utility and they are using Bayes Rule to optimise, but this mathematical rule requires lower levels of cognitive power than are assumed for the rational expectation bubbles.

2.4 Contagion Bubbles

There are a number of limitations in the theory of rational bubbles, and alternatives were developed that allow a softer set of assumptions about rationality, uncertainty and financial market efficiency. In our 2001 chapter, we focussed on the contagion bubble model presented by Topol (1991), which was particularly interesting because it represented a compromise between the extremes of rational and irrational bubbles. Topol (op. cit.) develops an analysis that essentially builds an encompassing approach based around a general model of contagion bubbles in which the extent of rationality is determined by the influence of social influences via mimetic contagion. Topol's model starts from an approach consistent with some elements of rational expectations theory but with weaker assumptions about rationality and expectations, though still allowing that probability is quantifiable and the coincidence of subjective expectations and objective probability distributions, as consistent with the rational expectations models above. Whilst Topol's baseline model does embedding implicit assumptions of quantifiable risk and an ergodic world, if mimetic contagion dominates, then Topol's model has many features in common with heterodox and post-Keynesian models of a non-ergodic world.[1]

What is the essence of Topol's mimetic contagion model? He argues that our capacity for rational decision-making is constrained and so his model presents an alternative to models embedding extreme assumptions about rationality. With these weaker assumptions about rationality, if each individual speculator is unsure, then asset prices are partly driven by collective views. Speculators infer something about what other speculators think from the prices they are willing to pay for assets. This insight is consistent with Keynes's ideas about social influences in financial markets, as we will explore in more detail below.

The time path of bubbles is driven by mimetic contagion given a different set of assumptions to those in rational expectations theory. Topol assumes incomplete information and given this lack of information held by the individual, speculators will extend their information sets by looking to the prices that other speculators are willing to pay. Topol's model is different in that it does allow a less extreme view of rationality than is seen in rational expectations and Bayesian models, but still assumes that people are able to apply relatively sophisticated mathematical decision-making rules. Speculators adjust their price expectations as they see others adjusting their price expectations. Via a process stochastic aggregation, this generates an additive learning process via which speculators are weighting the different sources of information including other buyers' prices, other sellers' prices and the 'agent-efficient price', which reflects the fundamental value. This gives the following model:

$$p_t = w_b p_b + w_s p_s + w_F p_F \tag{9.7}$$

where p_b is the price paid by other buyers, p_s is the price paid by other sellers, p_F is the agent-efficient price, driven by fundamental value, and w is the weight assigned to each of these different price signals, with the weights summing to one:

$$w_b + w_s + w_F = 1 \tag{9.8}$$

Topol's model can be understood as an encompassing model because when $w_F = 1$ and $w_b = w_s = 0$, his model reverts to a rational expectations model, consistent with Blanchard and Watson's model if p_F corresponds to Eq. (9.6). On the other hand, when $w_F = 0$, the bubble path is driven entirely by perceptions about prices that other speculators are paying. To capture the uncertainty dimension, Topol looks at the variance of prices and shows that mimetic contagion generates excessive volatility and cross-correlations, but when mimetic contagion disappears (that is, when the weights on others prices revert to zero) excess volatility is eliminated and the bubble collapses. Given Topol's quantitative definition of uncertainty (with which many heterodox economists would take issue, see below), he shows that the mimetic contagion weights are driven by uncertainty: when the variance of p_F is small, the mimetic contagion weights, that is the weights on other buyers and seller prices, will be small. Stock price movements and the bubbles which emerge therefore reflect a process of social transmission and mimetic contagion of stock price movements.

To recap on its essence, Topol's model is an eclectic, encompassing model which, in its general form, can capture the extremes of assumptions about rationality—from the strong rationality assumptions associated with rational expectations bubbles through to bubbles explained using weaker versions of rationality assumptions such as those associated with some heterodox approaches. With strict assumptions about rationality, including assumptions about informationally efficient financial markets, all agents with access to the same information and able instantaneously to arbitrage away price differentials, then all agents will converge on the same 'true' model (assuming also that a 'true' model exists, i.e. assuming an ergodic world). With these assumptions, Topol's model becomes a rational expectations bubble model.

At the other extreme, with its emphasis on herding and mimetic contagion, it can also be made consistent with some (not all) interpretations of Keynes (1936, 1937) analyses of financial market instability and in particular his famous insight that when we do not know what to do, then it makes sense to rely on information we can infer from others' decisions. Topol and Keynes also converge in terms of their emphasis on the idea that herding and conventional choices are more likely to dominate when uncertainty is endemic; though Topol sets this out in a more mathematical form in terms of the weights assigned to others' prices as uncertainty is increasing.

Topol's (1991) ideas also parallel the insights from Keynes (1936, 1937) about uncertainty. Conventions are not needed when we are certain. Conventions are necessary when knowledge is shrouded by uncertainty. Then, to paraphrase Keynes, it is necessary to rely on judgement of the rest of the world, because perhaps others are better informed. Nonetheless Topol's contagion bubbles do still sit more easily with a rational bubbles view of the world, in which decision-making is driven by mathematical reasoning. Also, whilst Topol does not specifically address this theme in detail, his contagion bubbles do also assume an ergodic, certain world in which systems are stable and unchanging, except in response to exogenous shocks.

In our 2001 chapter, we identified a number of problems with Topol's model, including that the dynamics cannot be captured via reference to individual differences in preferences or behaviour (Baddeley and McCombie 2001/2004). Also, Topol's model does not easily allow for a non-ergodic unstable world though he does allow that with strong enough forces of mimetic contagion, ergodicity will no longer prevail and then the world will be more like the Keynesian-Minskian world, as described below. More generally, Topol's model only requires that ergodicity lasts for as long as it takes a speculator to infer the state of the world in which they are operating.

2.5 Keynesian Bubbles

Using the term 'irrational' loosely, the models on which heterodox economists draw have their roots in Keynes (1936), Minsky (1992), and Kindleberger (1978). These models often assume that the world is non-ergodic, that is, it is not a stable system. Reality is changeable and bubbles are created by endogenous instability within the system. From the heterodox traditions, and specifically the post-Keynesian traditions, this sort of information and stability is not achievable. The future is not only unknown it is also unknowable and immeasurable (Davidson 1996; Palley 1993). This theme connects Keynes's insights with the Minskian models. An essential insight from Keynes is that speculators are not preoccupied by fundamentals. Instead, speculators are focussing on predicting the average opinion of average opinion and speculative bubbles are the consequence. So Keynes's analysis is not consistent with a rational expectations approach, which makes strong assumptions not only about rationality but also about information. Keynes (1936) explicitly emphasises that asset valuations do not coincide with fundamentals when he observes: 'certain classes of investment are governed by the average expectation of those who deal on the Stock Exchange as revealed in the price of shares rather than by the genuine expectations of the professional entrepreneur' (p. 151).

Another set of insights from Keynes that are not captured in the rational expectations literature, and that link with subsequent insights from heterodox and post-Keynesian economics, are the relationships between speculative bubbles and financial instability in the macroeconomy driven by interactions between speculators' financial choices and the fixed asset investment decisions of entrepreneurs. Speculation will connect with entrepreneurial investments in fixed assets because, as Keynes (1936) observes, 'there is no sense in building up a new enterprise at a cost greater than that at which an existing one can be purchased' (p. 151). Financial markets provide liquid sources of finance. These connections between speculators and entrepreneurs drive macroeconomic fluctuations. Booms and busts in financial markets link to expansionary and contractionary phases in the macroeconomy.

Financial markets connect speculators and entrepreneurs because entrepreneurs look to financial markets not only for financing their fixed asset investment projects but also for signals about the likely future potential of these investments (e.g. as explored in q theories, including both Tobin's mainstream q theories and post-Keynesian interpretations). So, when speculative bubbles dominate, the real economy will be adversely affected. Fixed asset investment will fall, a reverse multiplier will kick in, and instability will spread from financial markets through the macroeconomy. The crisis will be compounded as general instability and uncertainty lead to an increase in the propensity to hoard money via increases in precautionary and speculative demands for money. The economy will not be self-equilibrating; 'uncontrolled' and 'disobedient' business psychology, collapses in the state of confidence, and collapses in the state of credit will make the economy resistant to the usual monetary therapies (Keynes 1936, p. 317).

In some interpretations, Keynes's analysis of speculation and financial instability is not inconsistent with softer assumptions about rationality. So the nature of Keynes's assumptions about rationality versus irrationality of speculators (and entrepreneurs) is unclear. Some of the more persuasive literature on Keynes's views around rationality analyses the evolution of Keynes's ideas from his early work on probabilistic decision-making, as outlined in *A Treatise on Probability* (1921) (TP). Keynes did explore ideas around Bayesian decision-making in TP, and these ideas connect with the Bayesian bubble models noted above. One set of interpretations of Keynes argue that there is a continuity of ideas from TP through to *The General Theory of Employment, Interest and Money* (1936) and his *Quarterly Journal of Economics* article 'The general theory of employment' (1937). According to these continuity interpretations, Keynes was not setting out a model of human decision-making in which behaviour is fundamentally irrational. Nonetheless, given unmeasurable risk or 'Knightian uncertainty', the strict rationality assumptions embedded within neo-classical models (as the precursors to the rational expectations models of the 1960s and 1970s onwards) are not plausible.

Keynes's ideas share something in common with Topol's model too, with his focus on limited information and uncertainty. As uncertainty increases, it is harder for people to assign precise numbers to their expectations and so they rely more on socio-psychological influences. Uncertainty will also link to what Keynes calls the 'state of confidence', and this confidence not of an individual in their own expectations but instead captures general confidence in the economy as a whole. Speculation in unstable stock markets can play a significant de-stabilising role under these conditions. As Keynes (1936) observes: 'Speculators may do no harm as bubble on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on the whirlpool of speculation. When the capital development of a country become a by-product of the activities of a casino, the job is likely to be ill-done' (p. 159).

Keynes's ideas set a foundation for Hyman Minksy's analyses, focussing on the idea that the seeds of financial crisis are planted during bubble phases when euphoria is at its height. In Minksy (1992) speculation is more obviously linked to what some might call ‘irrational’ but more accurately could be described as ‘psychological’ influences. These psychological influences connect speculation with financial crisis, with financial crisis having its roots in earlier euphoria, including the euphoria associated with speculative bubbles. Minsky (1982) draws on some insights from Keynes in linking financial markets and speculative bubbles with fluctuations in the macroeconomy more widely. He adds into his analysis the role played by endogenous forces in money and financial markets, and argues that these are crucial catalysts to bubbles and crises.

In common with Keynes, Minsky is showing how financial crises are the inevitable consequence of the fluctuations that characterise capital systems; and speculative bubbles are the catalyst for crises. Euphoria plants the seeds of crisis. Specifically, Minsky develops some of Keynes's insights around the interdependencies of financial markets and real economic activity. He focuses on the ‘deviation-amplifying complementarities’ that develop during economic expansions but planting the seeds of subsequent financial crisis which inevitably follow. This links to the idea developed by Keynes that fixed asset investment drives real macroeconomic activity. Fixed asset investment is also a key determinant of profits (an idea also developed in Kaleckian models).

Minsky's fundamental innovation to Keynes's model is in a much clearer analysis of different types of finance and their impacts on investment and macroeconomic performance more generally. He focuses on three sources of finance: hedge finance, speculative finance and Ponzi finance. When cash flow into a company exceeds the cash flow out, this is the safest form of finance, which Minsky called hedge finance. Speculative finance is sustainable if current conditions persist because it is characterised by near-term cash flow which at least matches immediate costs, mainly the interest cost of debt; and expectations of future cash flow at least match cash outflows in the future, namely, repayments of capital. Ponzi finance is the most unstable form because it is characterised by businesses taking on more debt in order to match their current financial obligations. It is unsustainable and is founded on expectations of bonanzas in the future. These expectations can sometimes be justified in expansionary phases but are not at all justifiable when an economy starts to go into reverse. One of the crises of modern capitalism emphasised by Minsky, paralleling Keynes (1936) idea that private investment will never be sufficient to ensure full employment, is that unsustainable Ponzi finance is a part of many long-term investment projects.

The impacts from this unsustainability emerge when economies start to change. In tranquil phases, holding cash is less lucrative so fixed asset investment increases and the price of capital assets increases concomitantly. In terms of the three forms of finance, there will be a portfolio shift towards speculative and Ponzi finance. This process is generated endogenously by the characteristics of the system and this marks the start of a speculative bubble phase as, increasingly, investments are funded by expectations of future bonanzas. The financial structure of the macroeconomy starts to show signs of instability. As this instability increases, reflecting the increasing dominance of Ponzi finance, the system becomes more and more fragile, and more sensitive to changes in interest rates. Any increases in short-term interest rates mean that tranches of hedge finance units become speculative finance units, and tranches of speculative finance units become Ponzi finance units. Endogenous processes kick in to raise interest rates still further. As increases in short-term interest rates become more rapid and sustained, this leads to rises in long-term interest rates; so expectations of long-term sustainability start to shift. Rising interest rates also erode the present value of future profits, leading to what Minsky defines as 'present value reversals'. Present value reversals have two sets of consequences. First, new investments will fall because profit expectations are falling. The price of capital assets will fall as a consequence and this will erode businesses' capacity to fulfil their financial commitments. Second, the justification for Ponzi finance starts to look more and more shaky. Short-term deficits in cash flow turn into long-term deficits in cash flow, thus increasing the chance of default. To avoid default, holders of Ponzi finance will try to sell their assets in order to meet their financial obligations. From a loanable funds perspective, these processes are also exacerbated by shifts in the balance of investment demand and demand for finance. As investment demand rises relative to the supply of loanable funds, demand for finance rises and so interest rates increase. Demand for finance becomes more inelastic, fuelling further rises in interest rates.

2.6 Behavioural Bubbles

In our 2001/2004 paper, John McCombie and I focussed on the difference between rational bubbles and contagion bubbles. We argued that contagion bubbles could be used as a general model and adapted to capture the rational versus irrational extremes. What are the new insights we could have included in our analysis if we had been writing our chapter now? Since we wrote our chapter in 2001, behavioural finance has burgeoned in influence, though elemental insights from behavioural finance were being published by economists such as Richard Thaler (2005), from the 1970s onwards.

We touched on themes relating to behavioural finance but without exploring the range of ideas in great depth. In our 2001 chapter we did reference some of the early work in behavioural finance, specifically Robert Shiller's (1981, 1990, 2000) analyses of speculative bubbles. For example, we cited Shiller's study in which he uses survey methodologies to explore how financial investors form their expectations of asset price fluctuations, and he discovered that most people herd by following 'popular models'. Essentially, speculators are trend-spotters and tend to buy only after prices have begun to rise (Shiller 1990). In this section, I will develop our 2001 analysis by exploring in much more depth the ways in which behavioural finance can be used to explain speculation, focussing in particular on the role of herding and social influences. Modern analyses of these influences have many parallels with early insights from Keynes (1936, 1937). Connecting socio-psychological influences from behavioural finance with Keynes's parallel insights is consistent with a post-Keynesian model of speculation and represents less of a compromise with the unrealistic models from the rational bubbles literature, and to a lesser extent the contagion bubbles literatures. Behavioural bubble models also link to the insights from Minsky in which, as explored above, speculative bubbles are generated during euphoric phases, which are often interpreted as reflecting irrational decision-making. This decision-making is not so much irrational as psychological.

Another theme that we connected with from a behavioural, psychological perspective was the links between Keynes and Minsky and behavioural or psychological explanations for speculation. As noted above, both Keynes and Minsky focussed on links between speculative episodes, financial crises and macroeconomic instability, with speculative bubbles acting as the main triggers for more widespread instability. Entrepreneurs will also be affected by the euphoria developing in financial markets. When financial markets are buoyant, entrepreneurs' animal spirits and urges to act will be buoyant too. As Keynes (1936) observes: 'Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as a result of animal spirits – of a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities' (p. 161).

These ideas link to another set of insights from behavioural economics that can be applied to the analysis of speculative bubbles: the dual process thinking models popularised in Kahneman (2011). The idea here is that we have different thinking styles—fast, intuitive, automatic thinking (System 1) versus slower, more deliberative, cognitive thinking (System 2). System 1 might be associated with what some economists refer to as 'irrationality'. These ideas are also explored by neuroeconomists who use neuroscientific techniques to capture how emotions feed into financial decisions; for an example see Lo et al. (2005). The social influences that distract people away from private information about fundamental value are associated with complex interactions of neural structures usually associated with emotional versus cognitive decision-making (see, e.g. Baddeley et al. 2007, 2010; Burke et al. 2010).

In the context of speculative bubbles and financial instability, the dual process approaches of Kahneman (2011), and others, connect economic psychology with Keynes's earlier insights about the different decision-making systems driving our decision-making, including speculators and entrepreneurs. It is important to note that Keynes (1936) himself emphasised that this is not necessarily a reflection of irrationality: 'We should not conclude from this that everything depends on waves of irrational psychology. On the contrary, the state of long-term expectation is often steady, and, even when it is not, the other factors exert their compensating effects' (p. 161). In financial markets similar interactions of dual thinking processes will determine our choices; and Keynes (1936) extends his insight beyond financial markets and even beyond economics to many facets of our decision-making: 'We are merely reminding ourselves that human decisions affecting the future, whether personal or political or economic, cannot depend on strict mathematical expectation … it is our innate urge to activity which makes the wheels go round, our rational selves choosing between the alternatives as best we are able, calculating where we can, but often falling back for our motive on whim or sentiment or chance' (pp. 162–163).

Keynes's insights about animal spirits also connect with insights from psychology about the role of optimism bias in human decision-making (Baddeley 2014, 2016, 2017). Keynes (1936) observed that instability from speculation is compounded from instability generated by psychological influences: 'there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than on mathematical expectation, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as a result of animal spirits – of a spontaneous urge to action rather than inaction' (p. 161).

Some neuroscientists argue that optimism bias is a trait that characterises healthy adults, perhaps acquired for evolutionary reasons, and the absence of optimism bias is often a feature of depression (see, e.g. Sharot 2011). How can this explain speculative bubbles? If we allow that optimism bias is mirrored by pessimism bias, then speculative bubbles and crashes are generated from expectations built on precarious foundations. This euphoria cannot last. Shifting psychological influences have traction when disillusioned entrepreneurs realise that their expectations of future profits cannot be justified. Optimism bias is replaced by pessimism bias and bubbles are followed by crashes, partly a reflection of these shifts in mood and sentiment. The impacts spread to the real economy as speculative bubbles create financial instability, dampening enterpreneurs' investment, output and employment decisions.

Another key theme connects Keynes and economic psychology is the influence of social influences, including conventions and herding. When economies are very fragile, the state of confidence will be at low ebb. Without knowledge, people will be forced to rely on the rest of the world for information. Then herding and crowd psychology will overwhelm private information. Speculators realise that their perceptions about asset prices and profits from investments are built on precarious foundations. With no firm anchor for expectations, herding will generate instability and volatility. When these social influences are overwhelming, in a world of uncertainty any assumptions about individuals' capacity to be rational versus irrational are irrelevant, because the crowd takes on a nature of its own. This takes us back to Victorian conceptions of the crowd, whether crowds of speculators or political protestors, for example as explored by Gustave le Bon (1895) and Charles Mackay (1841). In crowds, individual's identity is lost and the crowd develops a nature of its own. As le Bon (1895) observed:

> … however like or unlike be [the individual's] mode of life, their occupations, their character, or their intelligence, the fact they have been transformed into a crowd puts them in possession of a sort of collective mind which makes them feel, think, and act in a manner quite different from that in which each individual of them would feel, think and act were he in a state of isolation ... the intellectual aptitudes of the individuals, and in consequence their individuality, are weakened …. (pp. 11–12)

Aside from economic psychology, behavioural finance also introduces some new perspectives on speculation. Richard Thaler and others have identified a range of behavioural anomalies that can illuminate the problem of speculative bubbles. Richard Thaler's insights are also informed by psychologists Daniel Kahneman and Amos Tversky, particularly their insights about the role of heuristics and biases in decision-making and the 'prospect theory' alternative to expected utility theory of risk (Tversky and Kahneman 1974; Kahneman and Tversky 1979). Kahneman and Tversky (op. cit.) focus on the idea that we make decisions with respect to reference points. We respond asymmetrically to losses versus gains. This links to the idea of loss aversion: we are far more upset about losing something than we are pleased when we gain the equivalent amount. Benartzi and Thaler (1995) develop these insights in their analysis of myopic loss aversion in financial markets, combining behavioural theories of time inconsistency from David Laibson and others (see, e.g. Laibson 1997; and also Frederick et al. 2002, for a survey). Anomalous financial decisions, for example, in favouring bonds over stocks even whilst there are persistent differentials in returns favouring stocks over bonds, reflect an interaction of present bias with loss aversion. Speculators are disproportionately focussed on the short-term fluctuations in share prices, over very short time horizons. This myopia combines with their loss aversion, so they avoid stocks because they are avoiding potential losses over short periods of time (Benartzi and Thaler 1995). Financial investors would do better if they shifted their time horizons towards the long-term and/or had more symmetric responses to losses versus gains.

3 Policy Implications and Conclusions

Policy implications depend on the model of speculation that policy-makers hold in their minds. If Keynes and Minksy are right, then conventional styles of monetary policy will not work in averting financial crises. The financial crises of 2007/2008 illustrate the point: the fashion for inflation targeting was at the very least an irrelevance in their aftermath. The jury is still out on what were called 'unconventional' monetary policies viz. quantitative easing. Whether or not quantitative easing was successful is still a question that needs much more empirical investigation. Many commentators argue that it was not enough and that austerity policies were exactly the wrong thing to do. True to Keynes, following financial crises, expansionary fiscal policy is required because monetary policy can only do so much in a world of profound pessimism, uncertainty and liquidity traps.

What can behavioural finance add to these policy prescriptions? First, behavioural finance allows that there are significant constraints on speculators' ability to judge fundamental values of assets. These constraints reflect limits on information, the presence of uncertainty—none of which preclude rationality in themselves. But these combine with some behavioural biases. Whilst these biases do not imply that everyone is always making mistakes, in the context of the profound uncertainty and substantial social influences that take hold during speculative episodes, the constraints on rational decision-making are likely to be overwhelming. Also, given the ideas outlined above about how crowds often take on an identity and mission that is completely different to those of the individuals within it, then financial policies focussed on assuming that each speculator is autonomous are likely to be misguided. Financial policy and financial regulation should be designed to allow that individual speculators may lose their individual autonomy when joining a crowd of other speculators, particularly when those other speculators are spread across large and complex international networks. Uncertainty and poor information magnify unstable speculation. Given that an individual speculator's capacity for rationality is likely to be severely constrained in complex financial markets given an uncertain world and poor information. Policy-makers could also do what they can to slow decision-making down so that the fast thinking does not dominate slower, more reflective and deliberative styles of decision-making. This gives a behavioural justification for the 'throwing sand in the wheels' arguments used by advocates of Tobin taxes, developing from Eichengreen et al.'s (1995) prescriptions originally devised in the context of international financial instability.

Finally, in looking back over the chapter that John McCombie and I wrote in 2001, what conclusions did we come to then? We argued that financial crisis and deflationary influences are signs on inefficient market processes, driven by endogenous fluctuations. In the face of these perverse and unstable market processes, capitalist economies need strong support from robust financial policies to limit the evolution of speculative bubbles. In this government spending will play a role in ensuring that fixed asset investment is not overly sensitive to the influences of speculation. We also argued for financial regulation because, without it, private speculation would drive unsustainable rises in asset prices. We argued that, in these circumstances, central banks should be willing to take on the role of lender of last resort and to float-off untenable debt structures when financial crises emerge (Baddeley and McCombie 2001). Our words seem prescient now, but in truth we were not the only economists espousing this view. Our analysis mirrored a substantial consensus from across the heterodox and post-Keynesian communities. More financial regulation, not less, was needed. Financial instability is an inherent feature of capitalist systems, and speculative bubbles are part of it, an insight well-explored in Keynesian, post Keynesian and heterodox literature ever since Keynes first explored related insights in the 1920s and 1930s. So the conventional wisdom about economists' ignorance is not well-founded. Many economists had predicted, and could explain, the 2007/8 financial crises. The problem was that these economists were not the economists with influence over the financial services sector and financial regulators.

Notes

1. Although Topol does allow that if mimetic contagion is powerful enough, the world may no longer be ergodic.

References

Avery, C., & Zemsky, P. (1998). Multidimensional uncertainty and herd behavior in financial markets. *American Economic Review, 88*(4), 724–748.

Baddeley, M. (2014). Rethinking microfoundations of macroeconomics: Insights from behavioural economics. *European Journal of Economics & Economic Policies: Intervention, 11*(1), 99–112.

Baddeley, M. (2016). Behavioural macroeconomics: Time, optimism and animal spirits. In R. Frantz & S. Mousavi (Eds.), *Routledge handbook of behavioural economics – Routledge international handbooks*. New York: Routledge.

Baddeley, M. (2017). Keynes's psychology & behavioural macroeconomics: Theory and policy. *Economic and Labour Relations Review, 28*(2), 177–196.

Baddeley, M., & McCombie, J. S. L. (2001/2004). An historical perspective on speculative bubbles and financial crises: Tulipmania and the South sea bubble. In P. Arestis, M. Baddeley, & J. S. L. McCombie (Eds.), *What global economic crisis?* Basingstoke: Palgrave Macmillan.

Baddeley, M, Christopoulos, G., Pillas, D., Schultz, W., & Tobler, P. (2007). *Herding and social pressure in trading tasks: A behavioural analysis* (Cambridge working papers in economics, no. 730). Cambridge: Faculty of Economics, University of Cambridge.

Baddeley, M., Burke, C., Tobler, P., & Schultz, W. (2010). *Impacts of personality on herding in financial decision-making* (Cambridge working papers in economics, no. 1006). Cambridge: Faculty of Economics, University of Cambridge.

Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics, 107*(3), 797–817.

Benartzi, S., & Thaler, R. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics, 110*, 73–92.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*(5), 992–1026.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives, 12*(3), 151–170.

Blanchard, O. J., & Watson, M. W. (1982). Bubbles, rational expectations and financial markets. In P. Wachtel (Ed.), *Crises in the economic and financial structure*. DC Heath: Lexington.

Burke, C., Baddeley, M., Tobler, P., & Schultz, W. (2010). Striatal BOLD response reflects the impact of herd information on financial decisions. *Frontiers – Human Neuroscience, 4*, article 48.

Chamley, C. P. (2003). *Rational herds: Economic models of social learning*. Cambridge: Cambridge University Press.

Davidson, P. (1996). Reality and economic theory. *Journal of Post Keynesian Economics, 18*(4), 479–508.

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review, 40*, 603–615.

Drehmann, M., Oechssler, J., & Roider, A. (2005). Herding and contrarian behavior in financial markets: An internet experiment. *American Economic Review, 95*(5), 1403–1426.

Eichengreen, B., Tobin, J., & Wyplosz, C. (1995). Two cases for sand in the wheels of international finance. *The Economic Journal, 105*, 162–172.

Flood, R. P., & Garber, R. J. (1980). *Speculative bubbles, speculative attacks and policy switching*. Cambridge, MA: MIT Press.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting: A critical review. *Journal of Economic Literature, 40*(2), 351–401.

Garber, P. M. (1989). Tulipmania. *Journal of Political Economy, 97*(3), 535–560.

Garber, P. M. (1990). Famous first bubbles. *Journal of Economic Perspectives, 4*(2), 35–54.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane/Penguin.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan/Royal Economic Society.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. London: Macmillan/Royal Economic Society.

Keynes, J. M. (1937). The general theory of employment. *Quarterly Journal of Economics, 51*(2), 209–223.

Kindleberger, C. P. (1978). *Manias, panics and crashes*. New York: Wiley.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112*, 443–478.
Le Bon, G. (1895). *The crowd: A study of the popular mind*. Lexington: Maestro Reprints.
Lo, A. W., Dmitry, V., Repin, B. N., & Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day traders. *American Economic Review, 95*(2), 352–359.
Mackay, C. (1841). *Extraordinary popular delusions and the madness of crowds* Lexington KY: Maestro Reprints.
Minsky, H. P. (1982). *Inflation, recession and economic policy*. Brighton: Wheatsheaf.
Minsky, H. P. (1992). *The financial instability hypothesis* (Levy Institute working paper WP74). New York: Levy Institute of Bard College.
Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica, 29*(3), 315–335.
Palley, T. I. (1993). Uncertainty, expectations and the future: If we don't know the answers, what are the questions? *Journal of Post Keynesian Economics, 16*(1), 3–18.
Sharot, T. (2011). *The optimism bias: Why we're wired to look on the bright side*. New York: Pantheon Books.
Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent movements in dividends. *American Economic Review, 71*(3), 421–436.
Shiller, R. J. (1990). Speculative prices and popular models. *Journal of Economic Perspectives, 4*(2), 55–65.
Shiller, R. J. (2000). *Irrational exuberance*. Princeton: Princeton University Press.
Shleifer, A., & Summers, L. H. (1990). The noise trade approach to finance. *Journal of Economic Perspectives, 4*(2), 19–33.
Thaler, R. (2005). *Advances in behavioural finance*. New York/Princeton: Russell Sage Foundation/Princeton University Press.
Topol, R. (1991). Bubbles and volatility of stock prices: Effect of mimetic contagion. *Economic Journal, 101*, 786–800.
Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

10

Sophistication, Productivity and Trade: A Sectoral Investigation

João P. Romero and Gustavo Britto

1 Introduction[1]

In the Kaldorian approach to economic growth, income elasticities of exports and imports are the crucial parameters determining the long-term growth rate. In this tradition, the requirement of balance-of-payments equilibrium represents the main constraint on the growth of domestic aggregate demand. If relative prices have little impact on trade flows, as the evidence suggests is the case, and balance-of-payments deficits cannot be financed indefinitely, income elasticities of exports and imports become the crucial parameters determining the long-term growth rate (Thirlwall 1979).

Consequently, it is crucial to understand what determines the magnitude of the income elasticities of trade. As McCombie and Thirlwall (1994) argued, income elasticities capture the non-price competitiveness of each country's production. Yet, very few contributions have sought to

J. P. Romero (✉) • G. Britto
Department of Economics (FACE) and Center for Development and Regional Planning (Cedeplar), Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_10

test what variables impact these elasticities. Empirical evidence provided by Ang et al. (2015) for Asian countries suggests that introducing technological competitiveness into a standard export demand function leads to a reduction in the income elasticity. Similar results were found by Romero and McCombie (2017). Focusing on a sample of developed economies and employing relative productivity to measure non-price competitiveness instead of technological competitiveness, they found that introducing changes in relative productivity reduces the magnitude of the income elasticities of demand for exports and imports. They argue that such changes are consistent with omitted variable bias, so that it is possible to argue that these results confirm that income elasticities are explained by non-price competitiveness.

If differences in non-price competitiveness provide the basic explanation for differences in income elasticities, sectoral differences represent a fundamental complement to this approach. It is well known that income elasticities change between sectors, so that different trade baskets are associated with different aggregate income elasticities (e.g. Araújo and Lima 2007; Gouvêa and Lima 2010; Romero and McCombie 2016). More specifically, evidence suggests that high-tech goods present higher income elasticities than low-tech goods. Hence, taking into account sectoral trade differences is paramount to understand the variability of aggregate income elasticities.

According to the structuralist approach to economic growth, economic development is a process necessarily associated with changes in the sectoral composition of production (Lewis 1955; Kuznets 1966; Kaldor 1966; Hirschman 1958; Prebisch 1962; Furtado 1964). Development and growth depend on moving the economy's structure towards high-tech, high value-added, sectors that produce goods that are complex. More recently, Hausmann et al. (2007) explored the richness of disaggregate trade data to provide compelling evidence that initial economic sophistication exerts a positive and significant impact on future GDP per capita growth rate, even when controlling for human capital, institutions and initial GDP per capita. Subsequently, more elaborate measures of product and economic sophistication were developed by Hidalgo and Hausmann (2009). These authors used Balassa's (1965) index of revealed comparative advantage (RCA) to derive measures of diversification and ubiquity of each country's exports, which were then combined to arrive at more accurate indexes of product and economic sophistication. Evidence suggests that these indexes are powerful predictors of subsequent GDP per capita growth (e.g. Felipe et al. 2010, 2013).

Taking into account the scarcity of evidence regarding the determinants of income elasticities of trade, especially when trade is disaggregated into different sectors, this chapter aims to investigate whether modern measures of productive sophistication can explain the magnitudes of income elasticities and export and import growth. More specifically, this chapter's contribution is twofold. First, using a measure of industry sophistication, the impact of initial industry sophistication on subsequent total factor productivity growth is tested using industry-level data. This analysis is carried out dividing the sample of industries into low- and high-tech, in order to assess if these sectors present different dynamics. In other words, the chapter examines whether Hausmann et al.'s (2007) hypothesis is valid for different technological sectors. Second, the chapter investigates if changes in industry sophistication impact exports and imports in low- and high-tech sectors. In this case, the index of industry sophistication is used as a proxy for the quality of the exports of a given industry. Special attention is paid to the impact exerted by the introduction of sophistication on the magnitudes of the income elasticities of demand.

The empirical exercises use product-level trade data from UN Comtrade which is combined with price data from Feenstra and Romalis (2014) and productivity data from EU KLEMS. The final dataset used comprises data for 13 industries, classified into low- or high-tech, in seven countries, over the period 1984–2006. Combining these databases restricts the time span but allows estimating the impact of industry sophistication on the total factor productivity growth of each industry in each sector. Moreover, this also allows for the estimation of export and import functions introducing changes in relative industry sophistication as an additional explanatory variable to assess its impact on the income elasticities of trade in low- and high-tech sectors.

The remainder of this chapter is organized as follows. Section 2 discusses the balance-of-payments-constrained growth theory, and the measures of product, economic and industry sophistication. Section 3 presents the empirical investigation of the impact of industry sophistication on total factor productivity in low- and high-tech sectors as well as on export and import growth. Section 4 reports the concluding remarks of the chapter.

2 Theoretical Framework

2.1 The Balance-of-Payments Constraint to Growth and Non-price Competitiveness

The Kaldorian tradition has a long track of theoretical and empirical studies investigating why economic growth is so uneven amongst countries. It emphasizes the role of demand growth as the ultimate determinant of a country's economic growth rate. In this framework, the balance of payments becomes the fundamental limit to the growth of an open economy. Exports, in turn, play the dual role of stimulating demand and of providing the foreign currency that allows the other elements of autonomous demand to grow, particularly investment. From a dynamic viewpoint, the stimulus to demand can trigger a virtuous growth cycle that tends to increase the global productivity of the economy, due to the migration of factors to more productive sectors (manufacturing) and to the greater learning-by-doing these sectors will display (Kaldor 1966).

This is the underlying argument of the balance-of-payments-constrained growth models. Thirlwall (1979) demonstrates that long-term growth is directly related to the income elasticities of demand for exports and for imports. The model is composed of three equations:

$$x_t = \eta\left(p_{dt} - p_{ft} - e_t\right) + \varepsilon z_t \tag{10.1}$$

$$m_t = \psi\left(p_{ft} + e_t - p_{dt}\right) + \pi y_t \tag{10.2}$$

$$m_t + p_{ft} + e_t = p_{dt} + x_t \tag{10.3}$$

Equations (10.1) and (10.2) represent the functions of demand for exports and for imports, respectively, both expressed in growth rates. The variable x stands for the growth rate of exports, m for the growth rate of imports, p_d and p_f are the rate of change of domestic and foreign prices, e is the rate of change of the nominal exchange rate, z is the growth of the income of the rest of the world, y is the growth of real output, η (<0) is the price elasticity of demand for exports, ψ (<0) is the price elasticity of demand for imports, ε is the income elasticity of demand for exports, and π is the income elasticity of demand for imports. Equation (10.3) is the balance-of-payments equilibrium condition.

Solving the system of Eqs. (10.1), (10.2) and (10.3), we arrive at the balance-of-payments equilibrium growth rate (the time subscripts have been dropped for expositional convenience):

$$y_{B1} = \frac{(1+\eta+\psi)(p_d - p_f - e) + \varepsilon z}{\pi} \tag{10.4}$$

This equation leads to many results: (i) a domestic inflation higher than the inflation of the rest of the world reduces the balance-of-payments equilibrium growth rate, if $|\psi + \eta| > 1$ (Marshall-Lerner condition); (ii) a currency devaluation ($e > 0$) tends to increase the balance-of-payments equilibrium growth rate, if $|\psi + \eta| > 1$; (iii) a faster growth of world income increases the balance-of-payments equilibrium growth rate; (iv) the higher is the income elasticity of demand for imports (π), the lower will be the balance-of-payments equilibrium growth rate. However, by assuming the stylized fact that terms of trade are constant in the long run ($p_{dt} - p_{ft} - e_t = 0$), the equation can be reduced to the ratio represented by Eq. (10.5), known as Thirlwall's Law:

$$y_{B2} = \frac{\varepsilon}{\pi} z \tag{10.5}$$

or

$$y_{B3} = \frac{x}{\pi} \tag{10.6}$$

This last equation represents the highest growth rate compatible with balance-of-payments equilibrium. A faster growth rate would be achieved via policies that stimulate increases in the income elasticity of demand for exports and reductions in the income elasticity of demand for exports. It is worth mentioning that Eq. (10.5) is also valid if the Marshall-Lerner conditions are just met (i.e. $\eta + \psi = -1$), even if there are substantial variations in relative prices.[2]

Although the output growth rate is determined by the growth of demand, balance-of-payments-constrained growth models also consider supply side factors. Nevertheless, these factors do not refer only to the increase of the stock of factor inputs, but also, to qualitative aspects, related to what has come to be called non-price competitiveness. In fact, a major part of the industrial output has been characterized by an oligopolistic competitive environment, in which aggressive price competition is not to be found. The predominant form of competition is, rather, non-price competition (McCombie and Thirlwall 1994).

Authors from different theoretical backgrounds have carried out empirical tests to assess the impacts of non-price competitiveness on foreign trade. Several types of proxies were used, including, amongst others, the number of patents and R&D expenditures. Some of these studies are based on the theory of the technological gap (Posner 1961; Hufbauer 1970; Greenhalgh 1990; Schott and Pick 1984; Fagerberg 1988; Wakelin 1998), while others are based on the product life cycle theory (Vernon 1966, 1970; Wells 1972) or even on the hypothesis of product differentiation and the preference for variety (Linder 1961; Davies 1976; Barker 1977). As a rule, the studies verify the importance of non-price competitiveness for the expansion of exports and, hence, for the growth of income.

The focus on non-price competitiveness, however, goes against the neoclassical assumption that similar goods are homogeneous and would, therefore, follow the 'law of one price'. Price differences, according to the neoclassical approach, would reflect a differentiation of the compared products. This procedure entirely voids the law of any empirical basis (McCombie and Thirlwall 1994). The growth of non-price competitiveness, therefore, indicates the degree of product differentiation and increases of the quality of national output. In this context, therefore, manufacturing would be more liable to be subject to such competitive gains, for primary goods tend to be more homogeneous. This is exactly what Kravis and Lipsey (1971) found, demonstrating that basic goods are more prone to price competition than manufactured goods.

The conclusion of this debate is that non-price competitiveness is an important factor explaining exports, given the preference for a variety that grows with income—even if, for the same reason, it does not lead to a reduction of imports. Theoretically, however, gains from non-price competitiveness can be obtained in any kind of products. Freeman (1979) tests the impact of different non-price competitive strategies on a set of sectors. The results show that, for the production of capital goods, competition focuses on the development of new, more technological, products. In the production of consumption goods, on the other hand, design and marketing play a more important role, while for basic materials most innovations focus on reducing inputs. Hence, sectors with higher technological intensity are more susceptible to non-price competitiveness and their elasticity of demand is, therefore, higher.

Exploring the idea that income elasticities capture non-price competitiveness and are different between sectors, Araújo and Lima (2007) introduced the Multi-Sector Thirlwall's Law (MSTL). By considering that each sector of the economy is subject to a different income elasticity of demand for its production, the model implies that shifts in sectoral shares affect the growth rate of the economy as a whole. Hence, a country's growth rate can increase even if the rest of the world continues to grow at the same pace, as long as the composition of exports and imports is favourably altered (Gouvêa and Lima 2010; Romero and McCombie 2016). In sum, the long-term growth rate depends on the sectoral structure of the economy.

A number of works have estimated income elasticities for different sectors within countries (e.g. Gouvea and Lima 2010, 2013; Romero et al. 2011; Romero and McCombie 2016). They have found that technology-intensive sectors present greater income elasticities. These studies also conclude that both the original Thirlwall's Law and its multi-sector version adequately represent the economy's real growth rate. Hence, the tests confirm the importance of increasing the share of high-tech sectors in order to accelerate growth.

Despite the recent evidence indicating that income elasticities vary considering between sectors, there have been very few attempts to investigate the determinants of the magnitudes of sectoral income elasticities. This important gap in the existing literature is partially explained by the fact that in most works that employ innovation-based measures of non-price competitiveness, income growth is not introduced as an explanatory variable of export performance.

Recent studies, however, have sought to analyse the significance of measures of non-price competitiveness when introduced in traditional export and import demand functions. Ang et al. (2015) introduced a measure of innovation stocks relative to the competitors into export demand functions. The authors have tested the effect of this measure of technological (or non-price) competitiveness on export growth for a sample of six Asian countries over the period 1953–2010, and have found robust evidence that non-price competitiveness exerts a positive and significant impact of export growth.

Romero and McCombie (2017) used total factor productivity (a measure of economic efficiency) as a proxy for product quality in different export and import industries. This proxy for non-price competitiveness is based on McCombie and Roberts (2002) and Setterfield (2011), who argue that productivity growth might determine the magnitude of the income elasticities, given that the former might result from quality improvements. Romero and McCombie (2017) tested the impact of total factor productivity relative to the frontier country for a sample of seven European countries over the period 1984–2006, dividing the sample into low- and high-tech industries. The authors found that changes in relative non-price competitiveness have a positive and significant impact on the growth rates of exports and imports of both low- and high-tech sectors. Nonetheless, the effect is greater in the high-tech sector. Most importantly, Romero and McCombie (2017) call attention to the fact that income elasticities vary considerably when relative productivity is introduced, which is consistent with omitted (quality) variable bias. Moreover, they also highlight that similar movements are observed in Ang et al.'s (2015) tests.

2.2 Product and Economic Sophistication

Seeking to investigate the importance of the composition of a country's production for economic growth, Hausmann et al. (2007) proposed two measures of product and economic sophistication.

The product sophistication index, called PRODY, is represented by the income level associated with each product, and is calculated as the weighted average of the income per capita of the countries that export the given product. Formally:

$$\mathrm{PRODY}_k = \sum_j \left[\frac{\left(x_{jk} / \sum_k x_{jk} \right)}{\sum_j \left(x_{jk} / \sum_k x_{jk} \right)} \right] \left(Y / L \right)_j \tag{10.7}$$

where x denotes the exports of good k by country j and Y/L is income per capita.

The PRODY index, therefore, ranks commodities based on the exports and each country's income levels. Hence, this index does not capture differences in product sophistication between countries. In other words, the index is an outcome-based measure of sophistication that is based on the assumption that, if a given product is largely produced by rich countries, then the product is regarded as 'sophisticated'.

The economic (or country) sophistication index, called EXPY, in turn, represents the productivity level associated with a county's export basket, and is calculated as the weighted average of the sophistication of the products exported by the country. Formally:

$$\mathrm{EXPY}_{jt} = \sum_{k} \left(\frac{x_{jkt}}{\sum_{k} x_{jkt}} \right) \mathrm{PRODY}_{k} \tag{10.8}$$

This index, therefore, is a weighted average of the PRODY indexes of each product k for a particular country j at time t, where the weights are the value shares of each product in the country's total exports.

Using this approach, Hausmann et al. (2007) provide evidence that current export sophistication is a good predictor of the future growth rate of income per capita. In other words, this approach suggests that fast-growing countries have EXPY indexes higher than their actual per capita incomes (such as China and India), which indicates they are producing goods associated with higher income.

Nonetheless, the authors show that producing sophisticated goods leads to high growth rates; the authors' investigation provided only an initial approximation to the determinants of EXPY. Their empirical investigation only indicates that EXPY is positively correlated with population size and land area, and not correlated with human capital and institution quality.

Hidalgo et al. (2007) addressed this limitation by investigating whether the productive structure of a country influences the path, the costs and the speed of change towards the production of sophisticated goods. As the authors stress, the production of different types of goods requires different capabilities. Consequently, the capabilities possessed by a country determine the goods the country can produce and how difficult it is for the country to start producing goods that require different (or additional) capabilities.

However, directly measuring capabilities is a complex task. As an alternative, therefore, the authors proposed using conditional probabilities to establish how close products are in terms of the capabilities required for their production. This method is based on the assumption that the probability of producing two products that require similar capabilities is higher than the probability of producing two goods that require different capabilities. Thus, the exercise used disaggregated trade to calculate the probability of a country exporting product *i* given that it exports product *k*. The authors called *proximity* this conditional probability. Finally, adopting a threshold value for proximity, the authors established linkages between products, creating a network that they called *product space*.

Using product space, Hidalgo et al. (2007) reached three interesting conclusions: (i) different countries face different opportunities for increasing their economic growth; (ii) structural change and economic growth are highly path dependent, given that each country's initial productive structure reflects a different set of capabilities; and (iii) moving towards sophisticated goods takes time, since this process requires learning new capabilities.

Another limitation of the measures proposed by Hausmann et al. (2007) is that the proposed measures do not explain what makes the products exported by rich countries important for economic growth. Indeed, the PRODY index is simply based on the assumption that sophisticated (high-productivity) goods are the goods exported by high-income countries. As Felipe et al. (2012) stresses, this makes the approach circular. Moreover, this creates some counter-intuitively high measures of product sophistication. To illustrate this problem, Reis and Farole (2012) point out that the PRODY of bacon and ham is higher than the PRODY of internal combustion engines.

Hidalgo and Hausmann (2009) address this shortcoming by developing alternative measures of product and economic complexity. The authors defined the degree of product diversification of a country as the number of products that a country exports with RCA, and the degree of ubiquity of a product as the number of countries that export a product with RCA. Formally:

$$RCA_{jkt} = \left(\frac{x_{jkt} / \sum_{k} x_{jkt}}{\sum_{j} x_{jkt} / \sum_{j} \sum_{k} x_{jkt}} \right) \tag{10.9}$$

$$D_{jt} = \sum_k N_{jkt} \tag{10.10}$$

$$U_{kt} = \sum_j N_{jkt} \tag{10.11}$$

where D denotes diversification, U denotes ubiquity and $N = 1$ if country j exports product k with RCA at time t, and $N = 0$ otherwise. The index of RCA developed by Balassa (1965) has a straightforward interpretation. If the index is higher than 1, then the country has high competitiveness in the production of the given good. The opposite holds if the index is lower than 1. Thus, the higher the diversification of a country's exports is, the higher this country's sophistication is. In contrast, the lower the ubiquity of a good is, the higher its sophistication is.

Using these indexes, Hidalgo and Hausmann (2009) and Felipe et al. (2012) show that economic growth is strongly correlated with the production of a diversified basket of goods that are not exported by many other countries. Indeed, the latter finds that the measures of economic and product sophistication proposed by Hidalgo and Hausmann (2009) are highly correlated with measures of technological capabilities used in Schumpeterian works (e.g. Archibugi and Coco 2005). Consequently, this approach shows that not only diversification and ubiquity are negatively correlated, which means diversified countries tend to produce more complex (less ubiquitous) goods, but diversification is positively correlated with income level.

However, as Hidalgo and Hausmann (2009) and Hausmann et al. (2011) stress, diversification and ubiquity are crude approximations of economic (or country) and product sophistication. They argue that ubiquity and diversity can be combined to obtain better measures of economic and product sophistication. A country with low diversification but that produces goods with high ubiquity can be considered more sophisticated than a country that has similarly low diversification but produces goods will low ubiquity. Analogously, a good with high ubiquity but produced by countries that have low diversification can be considered less sophisticated than goods with similarly high ubiquity but produced by countries that have high diversification. As Hausmann et al. (2011)

argues, this process can be repeated to progressively increase the information captured by the measures, which will converge after a few iterations. These are the product sophistication (PS) and economic sophistication (ES) indexes used in this chapter. Formally:

$$\mathrm{PS}_{kt,n} = \left(\frac{1}{U_{kt}}\right)\sum_{j} N_{jkt}\mathrm{ES}_{jt,n-1} \tag{10.12}$$

$$\mathrm{ES}_{jt,n} = \left(\frac{1}{\sum_{k} N_{jkt}\mathrm{PS}_{kt,n-1}}\right) D_{jt} \tag{10.13}$$

where *n* denotes the number of iterations.

The measures developed by Hausmann et al. (2007) and Hidalgo and Hausmann (2009) have been employed by a number of works to analyse the development trajectories of different countries, taking into account the transformations in their productive structures. Felipe et al. (2010), for instance, has shown that Pakistan was not able to move towards the production of more sophisticated goods, which resulted in recurrent balance-of-payments problems, curtailing the country's growth. Felipe et al. (2013), in turn, showed that the successful development trajectory of China was associated with progressive increases in the RCA of products with high sophistication (especially machinery and electronics).

In addition, recent works have been extrapolating these measures and using them in econometric investigations. Boschma et al. (2013), for example, applied the approach to the analysis of technological proximity and technological change in US cities. Using patent data from the United States Patent and Trademark Office (USPTO) disaggregated by International Product Categories (IPC), the authors calculated an index of Revealed Technological Advantages (RTA) analogous to Balassa's (1965) RCA and used it to construct a *technology space* analogous to Hidalgo et al.'s (2007) product space. Using the technological proximity between different patent classes, the authors showed that different technological capabilities influenced different trajectories of technological specialization between cities. Bahar et al. (2014), in turn, used RCAs and an export similarity index to show that geographic proximity influenced the productive specialization of neighbouring countries. In other words, countries that are geographically close tend to present RCAs in similar products. The authors attribute this result to technological diffusion.

2.3 Industry Sophistication

In this chapter, EXPY is transposed to the industry level to measure the sophistication of the production of a given industry in each country. Calculating this index for each of the industries in the EU KLEMS database allows analysing the relationship between sophistication and productivity at the industry level. Moreover, using the same level of aggregation allows to assess the results found by Romero and McCombie (2017), investigating the impact of industry sophistication on trade performance.

The industry sophistication index, IEXPY, is calculated as the weighted average of the *PRODY* of the *n* products that integrate each industry *i*, for each country *j*, at time *t*:

$$\mathrm{IEXPY}_{ijt} = \sum_{n} \left(\frac{x_{jkt}}{\sum_{k} x_{jkt}} \right) \mathrm{PRODY}_{k} \qquad 10.14$$

Furthermore, an additional measure of industry sophistication is proposed in this chapter. Following the methodology proposed by Hausmann et al. (2007), the IEXPY index measures industry sophistication as the weighted average of PRODY for each product *n* in industry *i* and country *j*. The alternative measure proposed here, IEXPS, replaces PRODY with the product sophistication index *PS* based on Hidalgo and Hausmann's (2009) approach:

$$\mathrm{IEXPS}_{ijt} = \sum_{n} \left(\frac{x_{jkt}}{\sum_{n} x_{jkt}} \right) \mathrm{PS}_{k} \qquad (10.15)$$

3 Empirical Investigation

3.1 Data Description

The trade data used to calculate the industry sophistication indexes discussed in the previous section are from the UN Comtrade database, classified according to the Standard International Trade Classification (SITC) (Revision 2, 4 digits), and the data on GDP per capita (2011 PPP$) are from the World Development Indicators. The indexes were calculated for the period 1984–2006, given that price data from Feenstra and Romalis (2014) are available between 1984 and 2011, and the EU KLEMS data required to calculate productivity for each industry are available from 1976 to 2006. The final sample, therefore, comprises 13 goods-producing industries in seven countries, Austria, Denmark, Finland, Germany, the Netherlands, Spain and the United Kingdom.

Table 10.1 reports the products with highest and lowest values of the PRODY and PS indexes. PS and ES indexes were calculated using the first iteration betwen diversity and ubiquity. This table illustrates the problem with the PRODY index stressed by Reis and Farole (2012), given that some primary- and resource-based products figure amongst the most sophisticated products. This table shows also that the PS index partially solves this problem, given that only one of the five most sophisticated products is a resource-based product. The other products are all medium- or high-tech products. On the other end, both indexes indicate that the products with lowest sophistication are all primary, resource-based or low-tech products.

In spite of the differences between the indexes, however, the Spearman rank correlation between the two is still considerably high (0.75). Using the average PRODY index as reference for product sophistication and adopting Leamer's (1984) classification, the most sophisticated products are machinery (PRODY of $17,696) and chemicals (PRODY of $16,770). Capital-intensive products (PRODY of $12,657) appear after forest products (PRODY of $13,954) and petroleum (PRODY of $12,669). The goods with the lowest sophistication are labour-intensive products, raw materials, animal products, cereals and tropical agriculture products, respectively. A similar picture emerges if the average PS index is used as reference for product sophistication. Chemicals (PS of 190.8) are the most sophisticated products, followed by machinery (PS of 189.4),

Table 10.1 Products with higher and lower values of the PRODY and PS indexes

SITC	Product description	PRODY	PRODY rank	PS	PS rank	Lall's (2000) technological class
Top 5 PRODY						
3413	Petroleum gases and other gaseous hydrocarbons, nes, liquefied	32462.06	1	93.60573	752	Primary products
5147	Amide-function compounds; excluding urea	29774.54	2	199.2528	167	Resource-based
7412	Furnace burners; mechanical stokers, etc., and parts thereof, nes	29135.82	3	230.5944	10	Medium-tech
5415	Hormones, natural or reproduce by synthesis, in bulk	28903.62	4	216.3814	60	High-tech
7268	Bookbinding machinery; parts thereof, nes	27274.98	5	230.9483	9	Medium-tech
Bottom 5 PRODY						
1212	Tobacco, wholly or partly stripped	2115.613	753	128.7319	678	Primary products
2631	Raw cotton, excluding linters, not carded or combed	2098.813	754	98.88403	748	Primary products
741	Tea	2003.272	755	105.7965	745	Primary products
2771	Industrial diamonds	1981.667	756	142.4812	598	Primary products
2634	Cotton, carded or combed	1974.082	757	125.4862	688	Primary products
Top 5 PS						
6880	Uranium depleted in U235, thorium and alloys, nes; waste and scrap	25010.42	14	243.9294	1	Resource-based
5827	Silicones	24326.98	28	237.6393	2	Medium-tech
7753	Domestic dishwashing machines	24585.14	22	237.3484	3	Medium-tech

(*continued*)

Table 10.1 (continued)

SITC	Product description	PRODY	PRODY rank	PS	PS rank	Lall's (2000) technological class
7187	Nuclear reactors, and parts thereof, nes	20562.21	94	233.9071	4	High-tech
5836	Acrylic and methacrylic polymers; acrylo-methacrylic copolymers	22150.27	59	232.5783	5	Medium-tech
Bottom 5 PS						
2320	Natural rubber latex; natural rubber and gums	5746.701	675	92.38287	753	Primary products
2232	Palm nuts and kernels	4119.705	724	91.9226	754	Primary products
711	Coffee green, roasted; coffee substitutes containing coffee	2325.538	751	91.16449	755	Primary products
2655	Manila hemp, raw or processed but not spun, its tow and waste	5093.396	701	89.62418	756	Resource-based
611	Sugars, beet and cane, raw, solid	5946.703	669	88.33054	757	Resource-based

Source: Authors elaboration

capital-intensive (PS of 176.1), and labour-intensive goods (PS of 168.9). Interestingly, petroleum now figures the second least sophisticated product (PS of 140), which highlights the superiority of this index in measuring product sophistication.

3.2 Descriptive Analysis

In order to assess the relationships between sophistication, exports, imports and productivity for different groups of industries, the proposed indexes of industry sophistication (IEXPY and IEXPS) were calculated for the 13 goods-producing EU KLEMS industries, for which high quality data on sectoral productivity is available.

Figure 10.1 shows the evolution of the diversification of the economies under investigation, dividing the productive structure into core products (machinery, chemicals and capital-intensive) and peripheral products (labour-intensive, forest products, raw materials, animal products, cereals, petroleum and tropical agriculture products). This figure shows that Germany is the country with the highest number of core products with RCA during the whole period. Nonetheless, this number has been falling since the 1980s. The same occurs with the United Kingdom, while Spain and Finland have been increasing the number of core products in which the country has RCA. The remaining countries present relatively stable figures during the period. The scenario is similar for peripheral products, with three important distinctions: (i) Germany has a much lower number of products with RCA, although this number is relatively stable; (ii) the Netherlands is the country with the highest

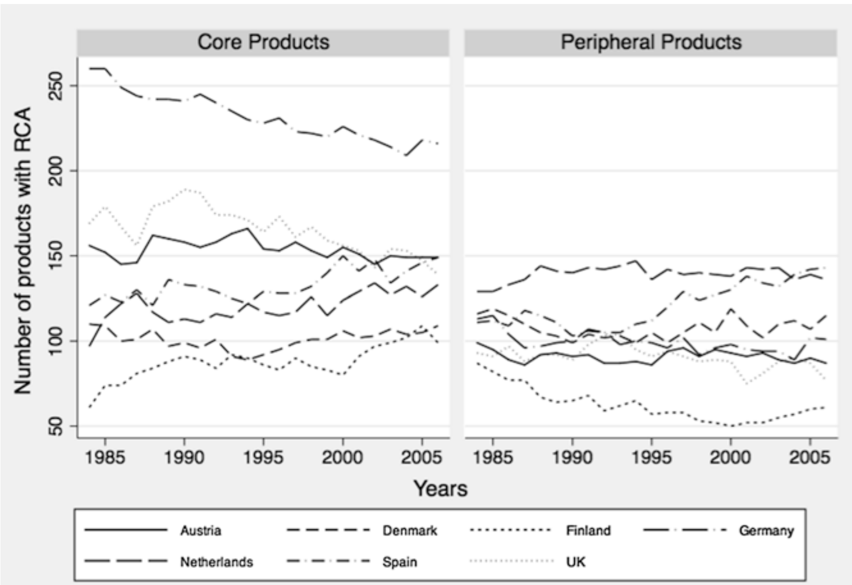


Fig. 10.1 Productive diversification of selected European countries (1984–2006) (Source: Author's own elaboration)

number of peripheral products with RCA; and (iii) Finland presents a falling number of peripheral products with RCA, which indicates the country has been moving from the production of this type of goods to the production of core goods.

Comparing the numbers shown in Fig. 10.1 with the average index of economic sophistication (ES) of the countries under analysis indicates that there is a close correlation between productive diversification (especially in core products) and economic sophistication, as found in previous works (e.g. Felipe et al. 2012). According to the ES index, the most sophisticated of these countries is Germany (ES of 0.064), followed by the United Kingdom (ES of 0.061) and Finland (0.056). The Netherlands and Austria have almost the same sophistication (ES of 0.054). Spain has the second lowest sophistication (ES of 0.051), only ahead of Denmark (ES of 0.050). Moreover, a similar ranking emerges when the average EXPY index is used, although in this case Finland figures as the most sophisticated economy. Nonetheless, when the evolution of the two indexes is compared a striking difference emerges: while according to the ES index the sophistication of all the economies has been falling during the period, the opposite holds for the EXPY index. This difference stems from the fact that the former is based on the diversification of each economy (given the ubiquity of the products produced), which tends to decrease through time, while the latter is based on the export shares of sophisticated goods, which tends to increase through time. Hence, this result indicates an important limitation of ES in comparison with EXPY, which shows that the latter is a superior index of economic sophistication.

After analysing the diversification of the economies under investigation, Table 10.2 turns to the analysis of the shares of core and peripheral products in total exports. This table shows that, although the number of core products with RCA has been falling in Germany and the United Kingdom, the share of these products has increased over the last couple of decades, and these countries possess the highest shares in this type of good. Indeed, the share of core products in total exports has been increased in all the countries. Nonetheless, Finland and the Netherlands have presented the highest increases. This suggests that increasing this share seems to be more important than diversifying the country's productive structure after a certain level of diversification is reached.

Table 10.2 Value shares in total exports: core and peripheral products

Period	1986–1989	1990–1994	1995–1999	2000–2006
Peripheral products				
Austria	0.35	0.33	0.33	0.31
Denmark	0.56	0.53	0.50	0.47
Finland	0.58	0.52	0.44	0.34
Germany	0.22	0.22	0.20	0.19
The Netherlands	0.52	0.49	0.42	0.34
Spain	0.43	0.34	0.35	0.34
United Kingdom	0.37	0.31	0.28	0.28
Core products				
Austria	0.65	0.67	0.67	0.69
Denmark	0.44	0.47	0.50	0.53
Finland	0.42	0.48	0.56	0.66
Germany	0.78	0.78	0.80	0.81
Netherlands	0.48	0.51	0.58	0.66
Spain	0.57	0.66	0.65	0.66
United Kingdom	0.63	0.69	0.72	0.72

Source: Author's own elaboration

Figure 10.2 shows the relationships between IEXPS, exports and productivity for the low-tech and the high-tech sectors. This index can be considered the preferred industry sophistication index, given that PS is a superior product sophistication index than PRODY, as argued in the previous section, while EXPY is a superior economic sophistication index than ES. Consequently, this chapter's discussion focuses on the IEXPS index, although the results found using the IEXPY index are similar.

As expected, Fig. 10.2 shows that industry sophistication is positively correlated with industry exports and productivity. Indeed, as Hausmann et al. (2007) have constructed EXPY to serve as a proxy for productivity, this is not an unexpected result. Interestingly, however, the relationship between industry sophistication and productivity is much stronger for high-tech industries than for low-tech industries. This preliminary finding shows that although productivity is positively correlated with quality in high-tech industries, this correlation seems to be less important in low-tech industries. A possible explanation is that low-tech industries rely more heavily on cost-competitiveness. Thus, as cost-competitiveness is often associated with specialization, it is not surprising that productivity

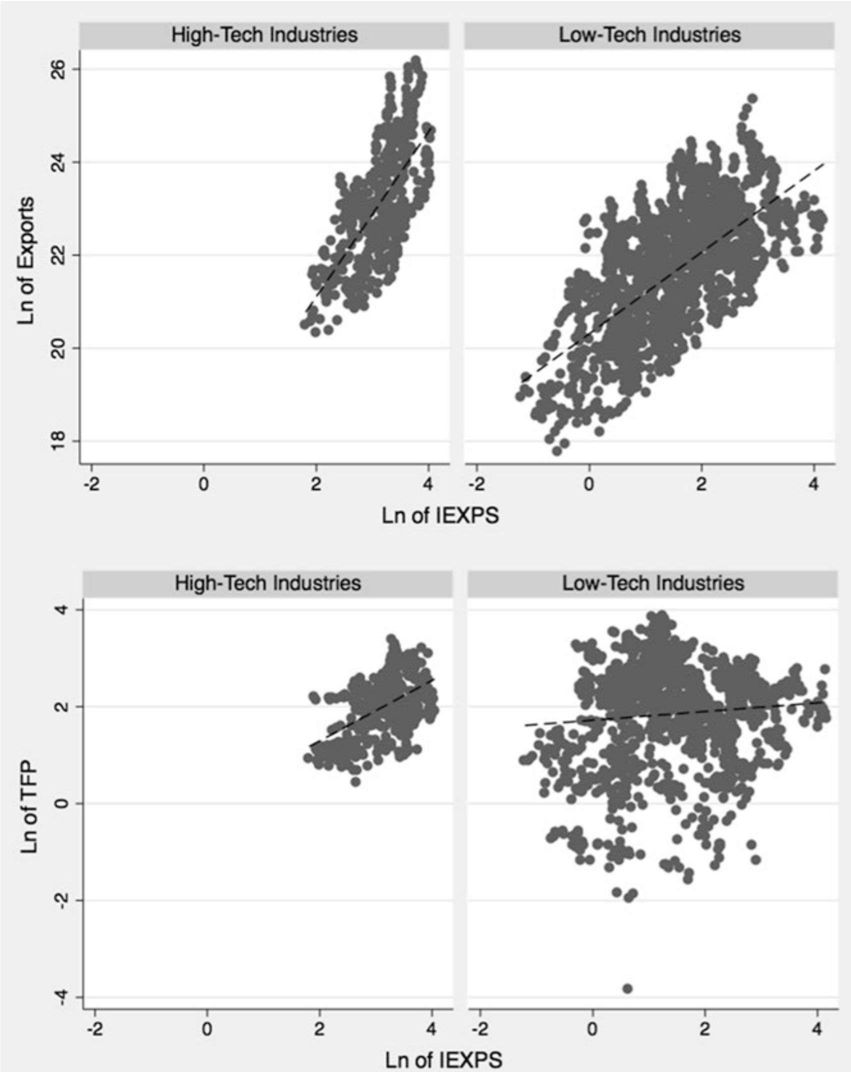


Fig. 10.2 Sophistication, exports and productivity at the industry level (Source: Author's own elaboration)

in low-tech industries is only weakly correlated with measures of sophistication, given that the latter are associated with diversification.

3.3 Estimation Method

The relationship between productivity and sophistication was tested following Hausmann et al. (2007), who estimated the impact of initial industry sophistication on subsequent productivity growth. In the tests reported in this chapter, product quality is proxied by the industry sophistication index IEXPS. Taking into account the impact of demand growth on productivity growth via Verdoorn's Law (e.g. Romero and Britto 2017), output growth was introduced as a determinant of productivity growth in each industry. Thus, the estimated equation is an expanded version of Kaldor-Verdoorn's Law:

$$T\hat{F}P_{ijt} = \beta_0 - \beta_1 \ln G_{ijt-1} + \beta_2 \ln N_{ijt-1} + \beta_3 \hat{Y}_{ijt} + u_{ijt} \qquad (10.16)$$

where TFP is total factor productivity, N is industry sophistication (i.e. IEXPS), Y is value added and G is the technology gap. The circumflexes over the variables denote growth rates. TFP growth rates ($T\hat{F}P \equiv \hat{Y} - \alpha K + (1-\alpha)\hat{L}$, where α is the share of capital in value added) were calculated using the log-level index number approach, which is more commonly used in the literature, while capital stocks were divided into information and communication technology (ICT) assets and non-ICT assets. The technology gap was calculated as the difference between the logarithms of domestic and foreign TFPs.[3] Data on real value added and capital stocks in 1995 US dollars, labour shares and number of hours worked by persons engaged in production were used to calculate TFP growth rates. Variables in constant 1995 prices were transformed from national currencies to 1995 US dollars using industry-specific PPPs from the Groningen Growth and Development Centre (GGDC) Productivity Level Database (Inklaar and Timmer 2008).

The 13 industries were split into two samples following the OECD technological classification (OECD 2003). The first sample, henceforth called low-tech industries, comprises five low-tech industries (Food, Textiles, Wood, Paper and Other Manufactures) plus three medium-low-tech industries (Plastics, Minerals and Metals). The second sample, henceforth called high-tech industries, comprises three medium-high industries (chemicals, machinery and transport) plus the high-tech industry (Electrical). The export and import demand functions estimated in this chapter follow the specifications proposed by Romero and McCombie (2017), which incorporate relative non-price competitiveness into standard export and import demand functions:[4]

$$\begin{aligned} \ln \hat{X}_{ijt} = \beta_0 - \beta_1 \ln \hat{P}_{ijt} + \beta_1 \ln \hat{P}_{fijt} + \beta_2 \ln \hat{N}_{ijt} \\ - \beta_2 \ln \hat{N}_{fijt} + \beta_3 \ln \hat{Z}_{jt} + u_{ijt} \end{aligned} \tag{10.17}$$

$$\begin{aligned} \ln \hat{M}_{ijt} = \beta_5 - \beta_6 \ln \hat{P}_{fijt} + \beta_6 \ln \hat{P}_{ijt} + \beta_7 \ln \hat{N}_{fijt} \\ - \beta_7 \ln \hat{N}_{ijt} + \beta_8 \ln \hat{Y}_{jt} + u_{ijt} \end{aligned} \tag{10.18}$$

where X is exports, M is imports, P is prices, Z is foreign income and N is quality (i.e. the product sophistication index IEXPS). Moreover, f denotes variables for the foreign economy, and i are industries in j countries at time t. Quality-adjusted price indexes calculated by Feenstra and Romalis (2014) for each SITC category were used to deflate the respective export and import values. Then, trade data was transformed from SITC (Rev. 2) 4-digits to ISIC (Rev. 2) 3-digits using the correspondence table developed by Muendler (2009), which is based on the OECD correspondence between SITC and ISIC. This data was then transformed into EU KLEMS industries. Import prices were used as proxies for foreign prices for each country and industry. Export and import prices in the EU KLEMS industries were calculated as weighted averages of the quality-adjusted price indexes of each product within each EU KLEMS industry.

The System Generalized Method of Moments (GMM) estimator was employed to control for fixed effects and simultaneity in the regressions reported in this chapter (see Blundell and Bond 2000; Roodman 2009).

3.4 Estimation Results

Table 10.3 reports estimates of the relationship between changes in sophistication and productivity growth. To assess the measures of sophistication calculated in this chapter, Hausmann et al.'s (2007) test of the relationship between initial EXPY and subsequent productivity growth

Table 10.3 Industry sophistication and productivity growth

Dependent variable	Growth rate of GDP per capita	Growth rate of TFP	Growth rate of TFP	Growth rate of TFP	Growth rate of TFP
Method	OLS	SYS-GMM	SYS-GMM	SYS-GMM	SYS-GMM
	Countries	Low-tech industries	Low-tech industries	High-tech industries	High-tech industries
Sample	(i)	(ii)	(iii)	(iv)	(v)
Ln of initial technology gap	−0.0106* (0.00436)				
Ln of initial EXPY	0.0378** (0.0117)				
Lagged technology gap		0.100 (0.206)	0.00498 (0.0815)	−0.0575 (0.0415)	−0.0638** (0.0212)
Lagged Ln of IEXPS		−0.0318 (0.0656)	−0.00314 (0.0269)	0.0473++ (0.0256)	0.0471* (0.0173)
Growth rate of value added			0.557* (0.273)		0.583* (0.240)
Constant	−0.239** (0.0768)	0.142 (0.257)	0.0202 (0.108)	−0.165 (0.112)	−0.187* (0.0666)
N. Observations	102	350	350	105	105
No. Groups		70	70	21	21
No. Instruments/ Lags		6/2–5	10/2–4	4/3	6/4
Arellano-Bond AR Test		0.753	0.662	0.653	–
Hansen's J Test		0.372	0.351	0.534	0.489

Note: The values reported for the tests are p-values. The p-value reported for the Arellano-Bond AR Test refers to the first lag used as instrument in the regression. The sample 'All Industries' comprises 13 goods-producing industries, excluding the Fuel and Chemical industries. Significance: ***=0.1%; **=1%; *=5%; ++=10%; +=15%

Source: Authors' elaboration

(proxied by GDP per capita) was replicated using cross-country OLS. The test was regressed using a sample of 102 countries for which data is available for all years of the period 1996–2006. The test employed the average of each variable during the period investigated. The estimated regression is reported in column (i). The significance and magnitude of the estimated coefficients are very similar to the results of Hausmann et al. (2007).

In columns (ii) to (v) of Table 10.3, industry-level data is used to estimate the impact of sophistication on productivity growth. Hence, the growth rate of GDP per capita is replaced by the growth rate of industry TFP, and EXPY is replaced by IEXPS. Moreover, System GMM is now utilized. Arellano and Bond's (1991) AR Test and Hansen's J Test indicate that the instruments are valid at a 5% significance level in all these regressions. Columns (ii) and (iv) of Table 10.3 replicate the specification tested in column (i) using samples of low- and high-tech industries, respectively. For low-tech industries, none of the variables is significant and sophistication has a negative sign. For high-tech industries, however, initial sophistication is positive and significant, so that the results are similar to the estimates of Hausmann et al. (2007).

Finally, in columns (iii) and (v) the growth rate of value added is introduced, and an expanded Kaldor-Verdoorn's Law is estimated. The results of these regressions are similar to the estimates of Romero and Britto (2017), suggesting that returns to scale are slightly higher in high-tech industries. Nonetheless, while sophistication is positive and significant for high-tech industries, the opposite holds for low-tech industries. These results indicate once again that sophistication is more important for productivity growth in high-tech industries, while it seems to be less relevant for low-tech industries. Although not significant, the fact that sophistication has a negative sign for the latter sample might be due to the fact that this variable is calculated based on the importance of diversification. In low-tech industries, however, where cost-competitiveness seems to be more important, specialization is likely to be more relevant than diversification. Furthermore, the measure of sophistication used here is not free from problems. Hence, these results should be taken with caution.

Table 10.4 reports estimates of export demand functions by technological sectors. Arellano and Bond's (1991) AR Test and Hansen's J Test indicate that the instruments are valid at a 5% significance level in all regressions but the one reported in column (iii). Nonetheless, given that foreign and domestic sophistication are highly correlated (0.71) in this sample, the regressions that include only domestic sophistication present the most relevant results.

The elasticities of demand and of domestic sophistication are both positive and significant, except for sophistication in column (iv). Focusing on the regressions that only include domestic sophistication, it is possible to observe that the income and the sophistication elasticities of demand are slightly higher for high-tech industries. Most importantly, comparing the estimates reported in Table 10.4 with the estimates of simple export demand functions presented in columns (i) and (iv), one observes that the income elasticities of demand change when sophistication is introduced. This result is consistent with omitted variable bias, as discussed by Romero and McCombie (2017). As expected, for high-tech industries, the elasticity reduces when domestic sophistication is introduced and its effect is removed from the income elasticity, and then increases when foreign sophistication is added. These results are similar to the ones found by Romero and McCombie (2017). For low-tech industries, however, the elasticity increases with the introduction of domestic sophistication instead of decreasing.

Table 10.5 reports estimates of import demand functions by technological sectors. Arellano and Bond's (1991) AR Test and Hansen's J Test indicate that the instruments are valid at a 5% significance level only in the regressions reported in columns (i), (ii) and (iv). Hence, these results must be considered with caution. Once again, given that foreign and domestic sophistication are highly correlated (0.71) in this sample, the regressions that include only domestic sophistication present the most relevant results.

For the import demand functions, the income and the foreign sophistication elasticities of demand are again both positive and significant, while domestic sophistication is negative and significant in column (ii), as expected, but is positive in column (v). Moreover, the changes in the income elasticities observed when measures of sophistication are introduced are not the expected movements. Hence, the results found for the import demand functions are not as consistent as the results found for the export demand functions.

Table 10.4 Export demand functions by sector

Dependent variable	Ln of exports	Ln of exports	Ln of exports	Ln of exports	Ln of exports	Ln of exports
	Low-tech industries	Low-tech industries	Low-tech industries	High-tech industries	High-tech industries	High-tech industries
Sample	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Ln of foreign income	1.776***	2.076***	2.215***	2.609***	2.169***	2.495**
	(0.147)	(0.189)	(0.212)	(0.398)	(0.363)	(0.739)
Ln of domestic prices	−0.842	−0.481	−2.529	2.174	0.548	1.803
	(2.389)	(0.583)	(1.928)	(1.394)	(1.121)	(3.376)
Ln of foreign prices	0.882	0.469	2.579	−0.0449	0.361	−1.035
	(2.340)	(0.597)	(1.884)	(1.613)	(1.151)	(3.882)
Ln of domestic IEXPS		0.635++	0.812*		0.850*	0.283
		(0.357)	(0.375)		(0.403)	(0.715)
Ln of foreign IEXPS			0.0252			−0.0107
			(0.346)			(0.568)
Constant	−33.18***	−43.49***	−47.95***	−54.75***	−45.44***	−53.89*
	(4.461)	(6.275)	(6.885)	(13.11)	(10.35)	(21.60)
N. Observations	420	420	420	126	126	126
No. Groups	70	70	70	21	21	21
No. Instruments/ Lags	6/2–3	11/2–4	12/3–5	6/2–3	9/3–4	10/2–3
Arellano-Bond AR Test	0.802	0.431	0.477	0.700	0.155	0.696
Hansen's J Test	0.654	0.258	0.000	0.455	0.839	0.417

Note: The estimation method used is SYS-GMM. The values reported for the tests are p-values. The p-value reported for the Arellano-Bond AR Test refers to the first lag used as instrument in the regression. The sample 'All Industries' comprises 13 goods-producing industries, excluding the Fuel and Chemical industries. Significance: ***=0.1%; **=1%; *=5%; ++=10%; +=15%

Source: Authors' elaboration

Table 10.5 Import demand functions by sector

Dependent variable	Ln of imports	Ln of imports	Ln of imports	Ln of imports	Ln of imports	Ln of imports
	Low-tech industries	Low-tech industries	Low-tech industries	High-tech industries	High-tech industries	High-tech industries
Sample	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Ln of domestic income	2.307*** (0.385)	1.001* (0.496)	1.066*** (0.125)	2.599*** (0.629)	0.647* (0.280)	0.951** (0.270)
Ln of domestic prices	−0.778 (3.442)	−1.105 (2.571)	9.511** (2.801)	−0.551 (2.147)	−1.621 (1.582)	0.917 (4.257)
Ln of foreign prices	1.097 (3.618)	1.152 (2.512)	−9.654** (3.141)	−0.275 (2.844)	−0.737 (3.021)	−1.044 (6.046)
Ln of domestic IEXPS		−1.250+ (0.783)	−1.420 (1.156)		2.475++ (1.244)	−0.0545 (1.176)
Ln of foreign IESPY			1.513++ (0.906)			1.251+ (0.777)
Constant	−39.35*** (10.14)	−2.781 (14.01)	−8.833** (3.122)	−46.94* (18.09)	−4.726 (7.183)	−6.553 (8.208)
N. Observations	420	420	420	126	126	126
No. Groups	70	70	70	21	21	21
No. Instruments/Lags	8/2–3	11/2–4	10/3–4	8/2–3	11/2–4	8/4
Arellano-Bond AR Test	0.139	0.571	0.768	0.384	0.138	0.003
Hansen's J Test	0.356	0.114	0.002	0.940	0.046	0.012

Note: The estimation method used is SYS-GMM. The values reported for the tests are p-values. The p-value reported for the Arellano-Bond AR Test refers to the first lag used as instrument in the regression. The sample 'All Industries' comprises 13 goods-producing industries, excluding the Fuel and Chemical industries. Significance: ***=0.1%; **=1%; *=5%; ++=10%; +=15%

Source: Authors' elaboration

A possible explanation for these minor inconsistencies is the fact that the measures of sophistication are constructed based on the stringent assumption that the sophistication of a given product does not vary between countries. If this assumption does not hold, then the measures of industry sophistication will not be good measures of industry quality. Moreover, the fact that the countries in the sample are all developed countries reduced the variance of the measures of industry sophistication, given that they depend on the shares of each exported product within each industry. This reduces the explanatory power of the variable as well.

4 Summary and Conclusions

The investigation presented in this chapter indicates that changes in product sophistication influence productivity, export and import growth. This chapter's tests suggest that productivity growth is associated with improvements in industry sophistication. Hence, the findings of the present chapter corroborate the findings of Romero and McCombie (2017). Nonetheless, the positive impact of industry sophistication on productivity growth is only significant in high-tech industries. This provides evidence that productivity growth in low-tech industries is to a higher extent associated with cost reductions (efficiency) and to a lesser extent associated with quality improvements, while the opposite holds for high-tech industries. However, given the limitations of the sophistication indexes employed, the impact of quality improvements for productivity growth in low-tech industries should not be dismissed without further investigation on the topic. In spite of this, the impact of sophistication on exports is positive and significant for both groups of industries. Most importantly, the impact of sophistication on exports is higher for high-tech industries. As for imports the tests provided some evidence that sophistication has a significant impact on imports as well.

Finally, considering the Kaldorian theoretical background of this chapter, the results further strengthen the longstanding notion that the long-term path to sustained growth is one of faster growth of exports to sustain increases in imports, in which the manufacturing sector plays a central role. However, in order to be a sufficient condition, diversification of production and exports towards progressively higher-tech, more sophisticated goods, is necessary. This is the important contemporary lesson with widespread policy implications for developed and developing countries alike.

Notes

1. Financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES), from the Brazilian National Research Council (CNPq) and from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) is gratefully acknowledged. The usual disclaimer applies.
2. The original model was also extended to account for capital flows. See Thirlwall and Hussain (1982), Barbosa-Filho (2001), Moreno-Brid (2003).
3. See Romero and Britto (2017) for more detailed discussion on the data treatment.
4. Capacity constraints are not considered in this chapter's tests.

References

Ang, J. B., Madsen, J. B., & Robertson, P. E. (2015). Export performance of the Asian miracle economies: The role of innovation and product variety. *Canadian Journal of Economics, 48*(1), 273–309.

Araujo, R. A., & Lima, G. T. (2007). A structural economic dynamics approach to balance-of-payments-constrained growth. *Cambridge Journal of Economics, 31*(5), 755–774.

Archibugi, D., & Coco, A. (2005). Measuring technological capabilities at the country level: A survey and a menu of choice. *Research Policy, 34*, 175–194.

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.

Bahar, D., Hausmann, R., & Hidalgo, C. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics, 92*, 111–123.

Balassa, B. (1965). Trade liberalization and revealed comparative advantage. *Manchester School of Economics and Social Studies, 33*, 99–123.

Barker, T. (1977). International trade and economic growth: an alternative to the neoclassical approach. *Cambridge Journal of Economics, 1*(2), 153–172.

Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews, 19*(3), 321–340.

Boschma, R., Balland, P.-A., & Kogler, D. F. (2013). Relatedness and technological change in cities: The rise and fall of technological knowledge in U.S. metropolitan areas from 1981 to 2010. *Papers in Evolutionary Economic Geography*, *13*(16). Utrecht University.

Davies, R. (1976). *On the relation between product differentiation and international trade flows*. Discussion paper. University of Bath.

Fagerberg, J. (1988). International competitiveness. *Economic Journal, 98*(391), 355–374.

Feenstra, R., & Romalis, C. (2014). International prices and endogenous quality. *Quarterly Journal of Economics, 129*(2), 477–527.

Felipe, J., McCombie, J. S. L., & Naqvi, K. (2010). Is Pakistan's growth rate balance-of-payments constrained? Policies and implications for development and growth. *Oxford Development Studies, 38*(4), 477–496.

Felipe, J., Kumar, U., Abdon, A., & Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics, 23*, 36–68.

Felipe, J., Kumar, U., Usui, N., & Abdon, A. M. (2013). Why has China succeeded? And why it will continue to do so. *Cambridge Journal of Economics, 37*, 791–818.

Freeman, C. (1979). Technological innovation and British trade performance. In F. Blackaby (Ed.), *De-industrialization, national institute of economic and social research*, *Economic policy papers 2*. London: Heinemann.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economy*. Cambridge, MA: MIT Press.

Furtado, C. (1964). *Development and underdevelopment*. Berkley: University of California Press.

Gouvêa, R. R., & Lima, G. T. (2010). Structural change, balance-of-payments constraint and economic growth: Evidence from the multi-sectoral Thirlwall's law. *Journal of Post Keynesian Economics, 33*, 171–206.

Greenhalgh, C. (1988). *Innovation and the structure of UK trade 1951–81: An exploration* (Applied economics discussion paper, n. 63). Institute of Economics and Statistics, University of Oxford.

Greenhalgh, C. (1990). Innovation and trade performances in the United Kingdom. *Economic Journal, 100*(400), 105–118.

Hausmann, R., Hwang, J., & Rodrik, D. (2007). What you export matters. *Journal of Economic Growth, 12*(1), 1–25.
Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simões, A., & Yildirim, M. A. (2011). *The atlas of economics complexity – Mapping paths to prosperity*. Hollis: Puritan Press.
Hidalgo, C., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10570–10575.
Hidalgo, C., Klinger, B., Barabasi, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*, 482–487.
Hirschman, A. (1958). *The strategy of economic development*. New Haven: Yale University Press.
Hufbauer, G. C. (1970). The impact of national characteristics & technology on the commodity composition of trade in manufactured goods. In R. Vernon (Ed.), *The technology factor in international trade, NBER Books*. NBER: New York.
Kaldor, N. (1966). *Causes of the slow rate of economic growth of the United Kingdom*. Cambridge: Cambridge University Press.
Kravis, I. B., & Lipsey, R. E. (1971). *Price competitiveness in world trade*. New York: National Bureau of Economics Research.
Krugman, P. (1991). *Geography and trade*. Cambridge, MA: MIT.
Kuznets, S. (1966). *Modern economic growth*. New Haven: Yale University Press.
Leamer, E. (1984). *Sources of international comparative advantage: Theory and evidence*. Cambridge, MA: MIT Press.
Lewis, A. (1955). *The theory of economic growth*. Homewood: Irwin.
Linder, S. B. (1961). *An essay on trade and transformation*. New York: Wiley.
Mayes, D., Buxton, T., & Murfin, A. (1988). *R&D innovation and trade performance*. Unpublished. NEDO.
McCombie, J. S. L., & Thirlwall, A. P. (1994). *Economic growth and the balance-of-payments constraint*. London: Macmillan Press.
Muendler, M. A. (2009). *Converter from SITC to ISIC*. Unpublished mimeo. University of California-San Diego.
OECD. (2003). *OECD science, technology and industry scoreboard*. Paris: OECD Publishing.
Pavitt, K., & Soete, L. (1980). *Technical innovation and British economic performance*. London: Macmillan.
Posner, M. V. (1961). International trade and technical change. *Oxford Economic Papers, 13*(3), 323–341.

Prebisch, R. (1962). The economic development of Latin America and its principal problems. *Economic Bulletin for Latin America, 7*, 1–22. United Nations.

Reis, J. G., & Farole, T. (2012). *Trade competitiveness diagnostic toolkit*. Washington, DC: The World Bank.

Romero, J. P., & Britto, G. (2017). Increasing returns to scale, technological catch-up and research intensity: An industry-level investigation combining EU KLEMS productivity data with patent data. *Cambridge Journal of Economics, 41*, 391–412.

Romero, J. P., & McCombie, J. S. L. (2016). Differences in increasing returns between technological sectors. *Journal of Economic Studies, 43*, 863–878.

Romero, J. P., & McCombie, J. S. L. (2017). Thirlwall's law and the specification of export and import demand functions. *Metroeconomica,* 00:1–30. https://doi.org/10.1111/meca.12185

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal, 9*(1), 86–136.

Schott, K., & Pick, K. (1984). *The effects of price and non-price factors on U.K. export performance and import penetration* (Discussion paper, 84-01). Department of Political Economy, University of London.

Setterfield, M. (2011). The remarkable durability of Thirlwall's law. *PSL Quarterly Review, 64*(259), 393–427.

Thirlwall, A. P. (1979). The balance of payments constraint as an explanation of international growth rates differences. *Banca Nazionale del Lavoro Quarterly Review, 128*, 45–53.

Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics, 80*(2), 190–207.

Wells, J. T. (1969). Test of a product cycle model of international trade: U.S exports of consumer durables. *Quarterly Journal of Economics, 83*(1), 152–162.

Wells, J. T. (1972). International trade: The product life cycle approach. In *The product life cycle and international trade*. Cambridge, MA: Harvard University Press.

11

Urban Growth in South Asia: A View from Outer Space

Mark Roberts

1 Introduction[1]

Since the turn of the century, South Asia has added an estimated 130 million people to its towns and cities, a number equivalent to the entire population of Japan.[2] In the process, the share of its population that lives in officially defined urban areas has grown from 27.3 percent in 2000 to 30.9 percent in 2011, implying a pace of urbanization that has been on a par with that in sub-Saharan Africa.[3] However, although South Asia's relatively rapid pace of urbanization over the last decade is well-known, less understanding exists of patterns of urban growth both across and within the region's countries. This includes both patterns of physical expansion of urban areas—that is, patterns of growth in the spatial footprints or urban extents of towns and cities—and patterns of urban economic growth.

To illuminate detailed spatial patterns of urban physical expansion and economic growth, this chapter makes use of data on night-time lights for

M. Roberts (✉)
The World Bank, Washington, DC, USA

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_11

the period 1999–2010 which has been remotely collected by satellites orbiting the earth. In so doing, the chapter follows Zhou et al. (2015), Ellis and Roberts (2016) and Tewari et al. (2017) by building on insights from two different but related literatures. First, it builds on literature from the remote sensing community which shows that night-time lights (NTL) data provides a suitable basis for monitoring dynamics of urban physical expansion at a national or regional level (e.g. Zhang and Seto 2011).[4] Second, it builds on recent work from the economics literature by Henderson et al. (2011, 2012) which demonstrates the existence of a strong positive correlation between the growth in intensity of a country's NTL and its growth rate of real GDP. Based on this correlation, Henderson et al. (op. cit.) argue that the lights data can be used to proxy, or otherwise improve the measurement of, real GDP growth in settings where GDP data is either of poor quality or entirely unavailable. Notably, for our purposes, this includes for towns and cities, for which national statistical agencies (even in developed countries) do not typically compile and publish GDP data.

In addition to describing patterns of urban growth, this chapter also uses the NTL data to explore the link between urbanization and extreme poverty, where we define extreme poverty as the share of the local population that lives on less than $1.25 per day, based on 2005 Purchasing Power Parity (PPP) exchange rates.[5] Specifically, we examine whether (i) sub-national districts, which have more brightly lit urban centers, have lower poverty rates in general, and (ii) poverty rates tend to be lower in districts which have better access to brightly lit cities, not just in their own district but also in surrounding districts.

The main findings reported in the chapter are as follows:

- For South Asia, overall, there has been a rapid expansion of urban lit area, which reflects rapid spatial expansion of towns and cities, over the period 1999–2010. Urban lit area has, thus, grown approximately twice as fast as the region's overall urban population, implying that urban areas became, on average, less dense.
- There is, nevertheless, considerable heterogeneity across countries in the pace at which urban physical expansion occurred. While the pace of expansion was extremely rapid in Afghanistan and Bhutan and fast

in India and Sri Lanka, it was much more modest in the other South Asian countries.

- South Asia is characterized by a number of agglomerations containing multiple cities, each of which has a population in excess of 100,000, set within continuously lit belts of urbanization. The number of these 'multi-city agglomerations' increased from 37 in 1999 to 45 in 2010, while the overall average number of cities per agglomeration increased by almost 25 percent. The growth in the number and sizes of agglomerations points to an increasingly connected network of cities.
- Again, however, there is considerable heterogeneity across countries with respect to the process of the formation of multi-city agglomerations. Thus, while the process is well under way in India, Pakistan and Sri Lanka, Nepal and Bangladesh are at a more incipient stage of this process. The process, meanwhile, has yet to start at all in Afghanistan.
- The most notable, and by far the largest, example of a multi-city agglomeration is the Delhi-Lahore mega-agglomeration which straddles the international border between India and Pakistan. This agglomeration has emerged from the fusing together of the Delhi and Lahore agglomerations, which existed independently in 1999. The existence of this agglomeration despite the well-known lack of permeability of the Indian-Pakistan border suggests more effective regional integration could yield potentially large gains from the exploitation of agglomeration economies.
- In general, the fastest rates of urban economic growth during the period 1999–2010 were concentrated on the peripheries of the region's major agglomerations, including on the peripheries of the Bangalore, Delhi, Hyderabad (India) and Lahore agglomerations.
- Intensive and extensive growth of cities tended to go together—cities whose economic activity increased fastest at the core also expanded their footprints fastest. This is consistent with the idea that vibrancy at the core also pushes growth in cities outwards through increasing land and housing prices.
- In some cities, the brightness of NTL at the core has actually diminished. Furthermore, in a sub-set of these cities with dimming cores, there has been some apparent shrinkage of the overall urban lit area. Such dimming cities are particularly prominent in Bangladesh, Nepal

and Pakistan, which indicates that increased load shedding is likely to be an important part of the dimming story. Where this is the case, it is suggestive of failing urban infrastructure, which is likely to be correlated with slow urban economic growth.

- For South Asia as a whole, sub-national districts with more brightly lit urban centers also experienced lower overall poverty rates. Likewise, districts with better access to brightly lit urban areas in neighboring districts tended to have lower overall poverty rates. This is suggestive of potential spillovers from urban areas both across districts and between urban and rural areas within districts. Such spillover effects are strongest in India.

The structure of the remainder of this chapter is as follows. Section 2 introduces the NTL data that underpins the analysis and discusses the methods used to delineate the spatial footprints of urban areas. It also examines the correlations between growth rates of GDP and NTL both across countries and Indian districts, as a means of helping to justify the use of the data to examine patterns of urban economic growth. Section 3 presents our results on patterns of urban growth, including patterns of both physical urban expansion and urban economic growth. Section 4 explores the link between urbanization and extreme poverty. Section 5 summarizes and concludes.

2 Data and Methods

2.1 Night-Time Lights Data

The NTL data product that provides the foundation for the analysis in this chapter is derived from the processing of the 'raw' lights data collected by the United States' Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) sensors. As the name suggests, the satellites which collect or 'sense' the data are, in actual fact, weather satellites. Thus, the sensors onboard the satellites were originally designed to collect data on clouds illuminated by the moon for purposes of making short-term cloud cover forecasts. However, as a fortuitous side-benefit, it came to be realized that the sensors could also detect, primarily artificial, sources of light emanating from the earth's surface, including, most notably, lights from urban areas (Croft 1978; Doll 2008).

The National Centers for Environmental Information Earth Observation Group at the US Government's National Oceanic and Atmospheric Administration (NOAA) has produced several NTL DMSP-OLS-based products.[6] The most commonly used product, and that which Henderson et al. (2011, 2012) made use of in their seminal work, provides data on the intensity of NTL for pixels which have a resolution of 30 arc seconds, which is approximately 0.86 square kilometers at the equator. For a given year, the intensity of NTL for a pixel is measured as a simple average over all cloud-free nights after the filtering out of both sunlit and moonlit data, glare and lighting features from the aurora.[7] Intensity of NTL is typically recorded between 8:30 pm and 9:30 pm each night and is measured on an integer Digital Number (DN) scale that ranges from 0–63 with larger values being associated with greater levels of average luminosity. One problem with this scale is that it implies that the NTL data is top-coded such that levels of NTL intensity beyond a certain threshold are all reported as 63. This top-coding is a consequence of the saturation of the satellite sensors under their normal operational settings. While top-coding does not represent a problem with respect to the delineation of urban areas, it does with respect to the use of NTL data to examine patterns of urban economic growth. This is because it prevents the detection of variations in the luminosity of bright urban cores of major cities such as Karachi, leading to potential downward bias in the estimation of urban economic growth rates, at least for some urban areas, using NTL data.

To overcome the top-coding problem, instead of the standard NTL product, we make use of the Global Radiance Calibrated Nighttime Lights, or, for brevity, rad-cal, product, for which the DN-scale is not bounded from above.[8] This product is based on experimental alterations that were made to the DMSP-OLS satellite sensors to avoid their saturation by bright urban lights. These alterations represent a deviation from the normal operational settings of the satellite sensors, which are based on their core function of detecting moonlit clouds.[9] By using the rad-cal product we can more reliably measure variations in NTL growth. Our preference for the rad-cal product is not, however, entirely without cost. Whereas the standard NTL product is available for the period 1992–2013, the rad-cal product is only available for select years: namely, 1996, 1999, 2000, 2002, 2004, 2005 and 2010. The sample period on which we focus is 1999–2010. Apart from the removal of the top-coding problem, the rad-cal data is identical to the standard NTL data. Much of our analysis focuses on the 685 South Asian cities which had a circa 2010 population greater than 100,000.[10,11]

2.2 Measuring Urban Expansion

In this chapter, we use NTL data for two inter-related purposes: (1) to examine patterns of expansion of the spatial footprints of urban areas; and (2) to analyze patterns of urban economic growth. We are able to do this by virtue of the fact that, in the data, we observe not just the absence or presence of light, but also its intensity. In essence, defining urban areas and examining patterns of expansion in urban area involves asking a different question of the data than that required to analyze patterns of urban economic growth. Thus, in the former case, we are simply concerned with whether we observe levels of NTL intensity consistent with the presence of built-up urban areas and, beyond that, we are not interested in the level of NTL intensity. Meanwhile, in the latter case, we are asking how NTL intensity has changed over time.

To delineate urban from rural areas using the NTL data, and, therefore, examine patterns of urban expansion, we follow several other recent studies, including Zhou et al. (2015), Ellis and Roberts (2016), Harari (2016) and Tewari et al. (2017), by making use of a *DN* threshold ($\overline{DN}$) such that pixels with $DN > \overline{DN}$ are classified as urban and all other pixels are classified as rural. As with Zhou et al. (2015) and Ellis and Roberts (2016), we select $\overline{DN} = 13$ as the dividing line between urban and rural areas. Justification for this choice is provided by Fig. 11.1, which shows, for South Asia, histograms of observed *DN* values in areas classified as urban (red) and agricultural (blue) in a conventional land-use map for the region.[12] As can be seen, *DN* values greater than 13 tend only to be observed in urban areas.[13]

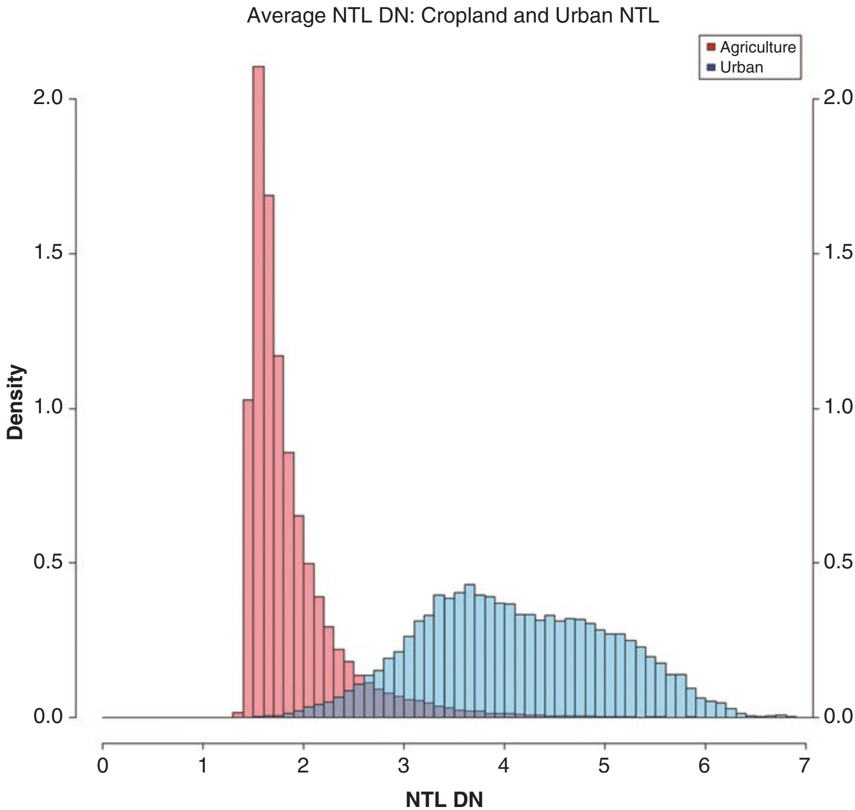


Fig. 11.1 *DN* values greater than 13 only tend to be observed in urban areas (Source: Own construction. Note: *DN* values are measured on a natural log scale. On this scale, a *DN* value of 13 corresponds to 2.57. *DN* values greater (less) than this number are more (less) commonly observed in urban than in agricultural areas)

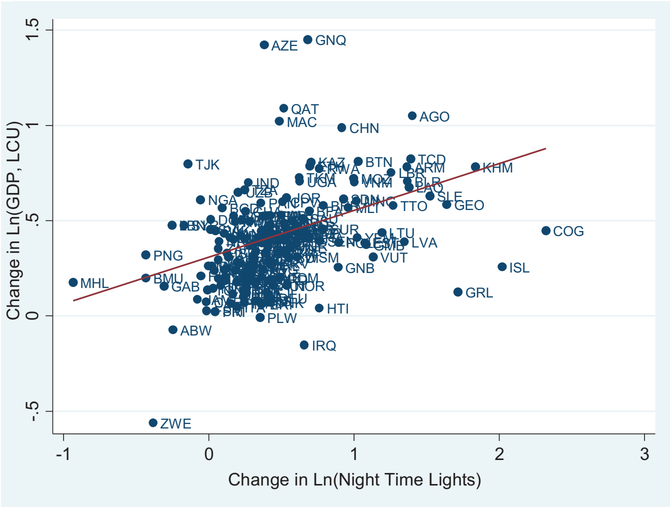
The use of a constant *DN* threshold also implies that we are imposing a consistent definition of urban areas across all South Asian countries. This allows us to overcome the measurement problems associated with the use of official national definitions of urban areas. These definitions vary across South Asian countries and, more generally, across countries globally. Hence, Ellis and Roberts (2016) identify seven criteria that usually feature in the official definition of 'urban' areas in South Asian countries: local government, population, population density, area of settlement, access to services, structure of the local economy and literacy rates. However, the countries vary both in the combination of the seven criteria they use and in their definitions of them. For example, while Nepal uses just one of the criteria (a simple population threshold of 9000 people), Bangladesh uses five of them. Moreover, within any given country, officially defined urban areas often fail to accurately capture the on-the-ground reality of urbanization. Hence, there is evidence of considerable 'hidden urbanization' in South Asia, that is, of the existence of a large number of places that possess urban characteristics but which are, nevertheless, classified as rural in official statistics (Ellis and Roberts 2016).

2.3 Measuring Urban Economic Growth[14]

Turning to the use of the NTL data to measure urban economic growth, this is based on the work of Henderson et al. (2011, 2012). In particular, it is based on their suggestion that the strong positive linear relationship between the growth rate in the intensity of a country's NTL and its rate of growth of real GDP can be used to predict, or otherwise improve the measurement of, GDP growth in settings—such as the urban level in South Asia—where GDP data is either of poor quality or simply unavailable.

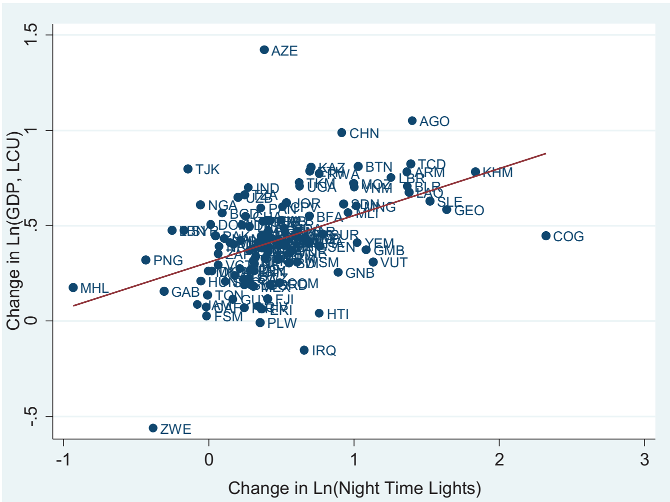
As noted above, we are using a slightly different NTL data product (i.e. the rad-cal data product) to that used by Henderson et al. (2011, 2012). Furthermore, our sample period of 1999–2010 also differs from Henderson et al.'s (op. cit.) sample period of 1992–2005. This being the case, it is important to verify that the growth of NTL intensity can be taken as a reasonable proxy measure of GDP growth for our data set. To this end, Fig. 11.2 shows, for our sample period, the correlation between the (natural) log change in a country's level of real GDP and the (natural) log change in the intensity of its NTL for both countries of all income levels (Fig. 11.2a) and just low- and middle-income countries (Fig. 11.2b). In both cases, we find positive linear relationships that are statistically significant at the one percent level. As with Henderson et al. (2011,

a



Slope coefficient = 0.2284, S.E. = 0.0384
R^2 = 0.162

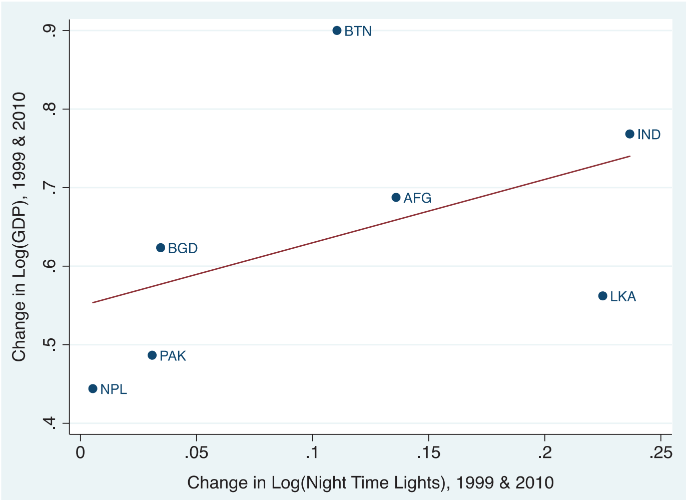
b



Slope coefficient = 0.2452, S.E. = 0.0427
R^2 = 0.210

Fig. 11.2 GDP growth is strongly predicted by the growth of NTL intensity for both: (a) a global sample of countries at all income levels and (b) middle- and low-income countries (Source: Own construction. Notes: Following Henderson et al., GDP is measured in local currency units at constant prices. GDP data is from the World Bank's Development Data Platform (DDP). Middle- and low-income countries are as identified in DDP. S.E. denotes robust standard error)

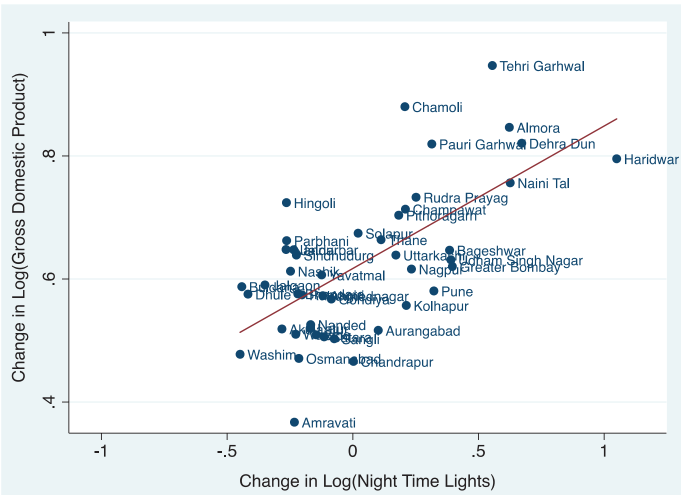
a



Slope coefficient = 0.8055, S.E. = 0.5378

R^2 = 0.222

b



Slope coefficient = 0.2316, S.E. = 0.0390

R^2 = 0.426

Fig. 11.3 GDP growth is positively correlated with the growth in the brightness of lights for both: (a) SAR countries and (b) Indian districts (excluding outliers) (Source: Own construction. Notes: Fig. 11.3a follows Fig. 11.2 in measuring GDP in local currency units at constant prices. GDP in Fig. 11.3b, which is also measured in constant prices, is from the Planning Commission, Government of India. Owing to the lack of district GDP data for 1999, the growth rate of GDP is calculated as the natural log change between 2000 and 2010. S.E. denotes robust standard error)

2012) we also find that, even after controlling for growth of electricity consumption, the growth in the intensity of a country's NTL retains some independent ability to predict its GDP growth rate.

If we look at just the South Asian countries which provide our focus of analysis, then we find that the positive linear relationship between the growth of GDP and NTL intensity continues to hold (Fig. 11.3a). The relationship is not, however, statistically significant at conventional levels. This should come as no surprise given the extremely small sample size involved. Once we exclude the outliers of Gadchiroli and Raigard, we do, however, find a statistically significant relationship—at the one percent level—between GDP growth and NTL intensity growth for (the limited sample of) Indian districts for which GDP data is available for both 1999 and 2010 (Fig. 11.3b). It is also worth mentioning that when we look at the relationship between GDP and lights in (natural log) *levels*, pooling together data for both 1999 and 2010, then we observe this to be very strongly statistically significant, both for the SAR countries (slope coefficient = 1.0904; S.E. = 0.0763) and Indian districts (slope coefficient = 0.8397; S.E. = 0.0293).[15]

Based on the above, we can conclude that the use of the rad-cal NTL data product to proxy for rates of economic growth is reasonable for our sample period of 1999–2010. Our use of NTL data in this way not only follows the suggestion of Henderson et al. (2011, 2012) but also a rapidly expanding number of other studies, including Alder (2015), Baum-Snow and Turner (2012), Pinkovskiy (2013), Zhou et al. (2015), Ellis and Roberts (2016) and Storeygard (2016).

3 Results

3.1 Rapid Expansion of Urban Lit Area[16]

Based on the methods described in Sect. 2.2, Fig. 11.4 illustrates a rapid growth of urban lit area by showing the expansion of urban footprints that occurred between 1999 and 2010, which we treat as being reflective of the physical expansion of urban areas. From Table 11.1, urban lit area for the region overall grew at a rate of just over five percent per annum over the period, increasing from around 173,000 km^2 in 1999 to almost

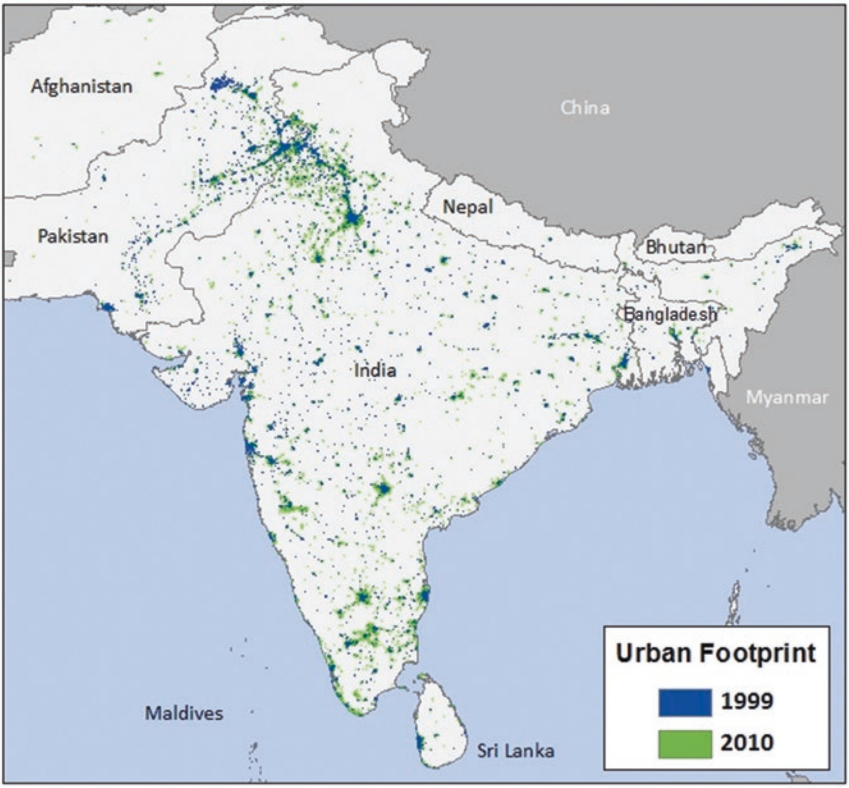


Fig. 11.4 The expansion of urban footprints over the period 1999–2010 (Source: Own construction)

297,300 km^2. This is equivalent to urban areas in the region having expanded, in aggregate, by an area equivalent to more than 100 times the area of New York City. By contrast, urban population growth for the region over the same period was a little less than 2.5 percent per annum.[17] Cities and towns, therefore, grew in area about twice as fast as they grew in population. This suggests an overall tendency toward declining average urban population densities, which is as we would expect as societies become richer and their populations demand more land per person for housing. The result that, in South Asia, urban area expanded at roughly twice the speed as urban population is consistent with earlier findings,

Table 11.1 Rapid growth of urban land area for the region overall, but with considerable heterogeneity across countries

		Urban area (km^2)			Urban growth rate (% pa)		
Country	Total area (km^2)	1999	2010	Absolute change	Area	Population	Ratio
Afghanistan	1,148,830	481	1969	1488	13.67	4.08	3.35
Bangladesh	188,049	3960	4865	905	1.89	2.81	0.67
Bhutan	58,895	38	148	110	13.16	5.19	2.53
India	4,210,083	126,525	236,924	110,399	5.87	2.41	2.44
Nepal	222,527	724	742	18	0.22	3.90	0.06
Pakistan	1,390,494	38,649	47,956	9307	1.98	2.42	0.82
Sri Lanka	66,358	2996	4695	1699	4.17	0.58	7.19
South Asia	**7,285,236**	**173,373**	**297,299**	**123,926**	**5.02**	**2.47**	**2.03**

Source: Own construction

Notes: Urban population growth rates are based on *World Urbanization Prospects: 2011 Revision* data (http://esa.un.org/unup/) and are for the period 2000–2010. Both growth rates of urban area and population are calculated as compound annual growth rates. Table excludes Maldives which possesses only one significant urban area – Malé – whose area corresponds to that of the island of Malé

covering the period 1990–2000, by Angel et al. (2011) based on a stratified global sample of 120 cities.

Unsurprisingly, the above overall regional picture largely reflects trends in India, whose share of total urban lit area in South Asia in 2010 was 57.8 percent. From Table 11.1, however, the most rapid rates of expansion in urban area were to be seen in Afghanistan and Bhutan, both of which recorded compound annual growth rates which exceeded 13 percent. In both cases, the rate of physical expansion also occurred at a faster pace relative to urban population than for the region overall. The country which, however, witnessed the fastest expansion of urban lit area relative to urban population was Sri Lanka, where the ratio of the growth rate of urban area to the growth rate of urban population exceeded seven. This occurred because, while Sri Lanka's total urban lit area grew at a rate close to that for the region overall, its urban population growth rate was much slower than for the region overall. The rapid expansion of urban area relative to urban population reflects the sprawl and ribbon development that are characteristic of Sri Lanka's urban development (World Bank 2012; Ellis and Roberts 2016).

In contrast, Pakistan, Bangladesh and Nepal, all experienced much more sluggish rates of growth of urban lit area, both in absolute terms and relative to the growth of their urban populations. For all three, urban population grew faster than overall urban lit area, implying that, in these countries, urban areas became more densely populated on average. Nepal represents a somewhat extreme case insofar as its overall urban lit area is estimated to have remained largely unchanged between 1999 and 2010 despite its urban population growth rate of 3.90 percent per annum.

3.2 Existence, Emergence and Growth of Multi-city Agglomerations

Another notable feature of the results is the existence, emergence and growth of multi-city agglomerations, where, following CIESIN (2013), Zhou et al. (2015) and Ellis and Roberts (2016), a multi-city agglomeration (or, for brevity, 'agglomeration') is defined as a continuously lit belt of urbanization consisting of two or more cities with a circa 2010 population greater than 100,000. In 1999, Table 11.2 shows that there existed 37 such agglomerations. Out of these 37 agglomerations, 35 were in either India or Pakistan with the remaining two agglomerations of Kathmandu and Colombo being in Nepal and Sri Lanka, respectively. By contrast, agglomerations were noticeably absent from the urban landscapes of the

Table 11.2 Existence and emergence of multi-city agglomerations, 1999–2010

	No. of agglomerations		No. of cities in agglomeration				Area (km^2)		
			1999		2010				Growth
Countries	1999	2010	Mean	Max.	Mean	Max.	1999	2010	(% pa)
India	23	30	4.09	17	4.73	38	22,240	75,499	11.75
Pakistan	12	10	4.00	10	6.50	29	1536	2558	4.75
Bangladesh	–	2	–	–	2.00	2	–	1340	–
Sri Lanka	1	2	5.00	5	3.50	5	182	205	1.09
Nepal	1	1	2.00	2	2.00	2	12,969	12,495	–0.34
South Asia	**37**	**45**	**3.92**	**17**	**4.89**	**38**	**36,927**	**92,097**	**8.66**

Source: Own construction

Notes: Afghanistan, Bhutan and Maldives are not included in the table because they had no agglomerations in either 1999 or 2010 remaining South Asian countries, including, perhaps surprisingly given the country's large overall urban population and high population density, from Bangladesh's urban landscape. The average number of cities per agglomeration was 3.92 with the largest agglomeration being Kolkata (India, 17 cities) followed by Delhi (India, 15 cities).[18]

The rapid expansion of urban lit area over the period 1999–2010 brought with it not only a net increase in the overall number of agglomerations to 45 but also in the average number of cities per agglomeration to 4.89.[19] India added seven agglomerations over the period, while Sri Lanka witnessed the emergence of the Galle agglomeration to add to the pre-existing agglomeration of Colombo. Bangladesh also experienced the birth of its first two agglomerations; the Dhaka agglomeration, which consists of Dhaka and Rupganj, and the Brahmanbaria agglomeration, which includes Bhairab in addition to Brahmanbaria. Pakistan, on the other hand, witnessed a net decline in the number of agglomerations from 12 to 10. This occurred as a consequence of the birth of new agglomerations being outpaced by the merging together of pre-existing agglomerations. Most notably, the Lahore agglomeration expanded to absorb the agglomerations of Chiniot, Gujranwala, Gujrat, Lalamusa and Sialkot. This merging together of agglomerations helps to explain the large increase in the average number of cities per agglomeration in Pakistan from four in 1999 to 6.50 in 2010, thereby making Pakistani agglomerations, on average, the largest in the region, at least on this criterion.

An example of the birth of a new agglomeration is provided by the Coimbatore agglomeration. In 1999, Coimbatore existed as a 'single city', that is, as a city with its own separately identifiable urban lit footprint. However, by 2010, urban expansion meant that Coimbatore's footprint had come to be indistinguishable from those of the nearby cities of Bhavani, Erode, Salem and Tiruppur, thereby making for one large multi-city agglomeration consisting of five cities with a combined population of 8.8 million (Fig. 11.5).[20]

Meanwhile, the most striking example of two or more agglomerations merging is the fusing together of the Delhi and Lahore agglomerations to form one enormous continuously lit belt of urbanization containing an estimated population of 73.4 million (Fig. 11.6). This mega-agglomeration stretches from Palwal, which is located to the south of Delhi, all the way

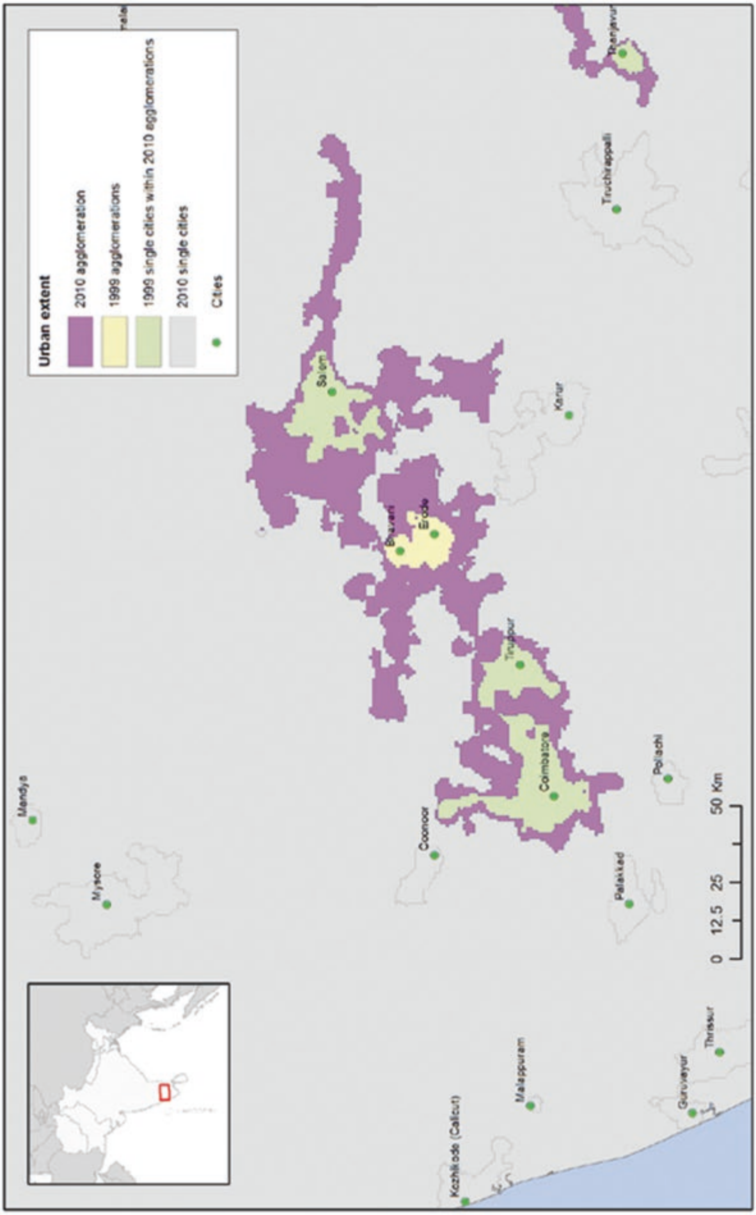


Fig. 11.5 The birth of the Coimbatore agglomeration, India (Source: CIESIN (2013))

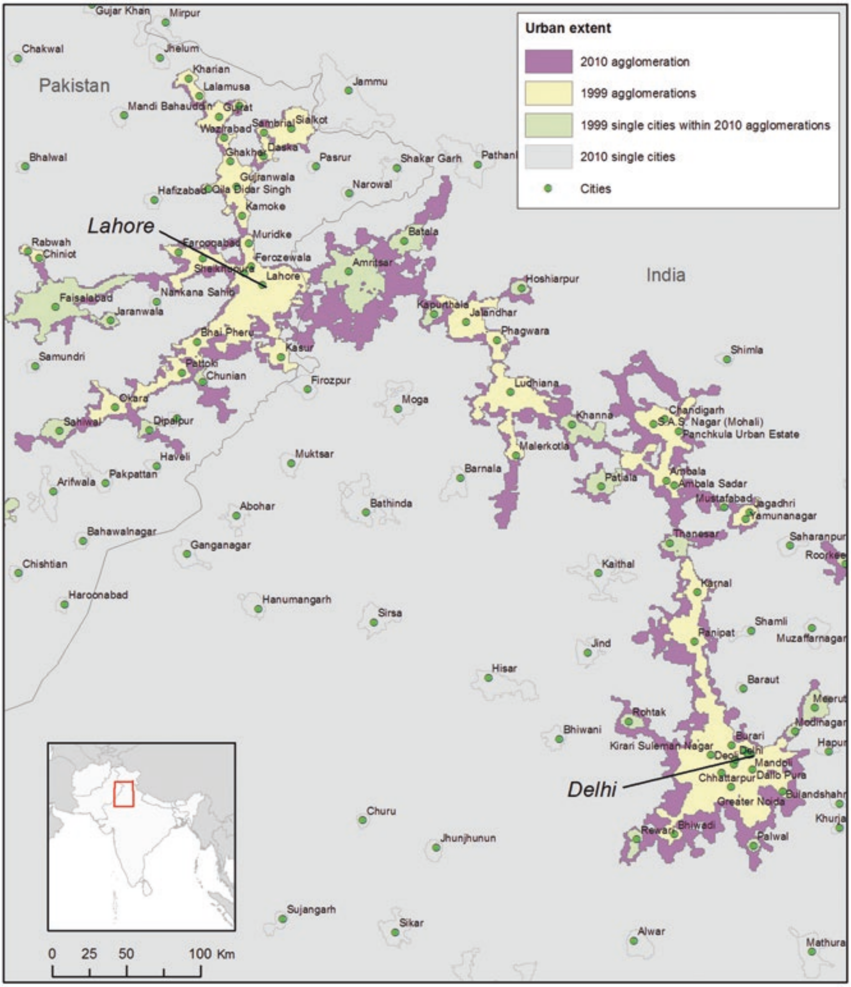


Fig. 11.6 The fusing of the Delhi-Lahore agglomerations across India and Pakistan (Source: CIESIN (2013))

to Kharian in the Pakistan province of Punjab. The formation of the Delhi-Lahore agglomeration resembles the emergence of cross-border mega-agglomerations of economic activity in more developed regions of the world—for example, the formation of Europe's 'Hot Banana' (so-called because the area it covers has the rough shape of a banana) which reaches from London in the UK to Milan in Italy.[21] What is notable, however, is that while the 'Hot Banana' is located in arguably the most highly integrated region in the world, the Delhi-Lahore mega-agglomeration has formed across one of the world's least permeable borders.

The birth of new agglomerations such as Coimbatore and the formation of the Delhi-Lahore mega-agglomeration points to an increasingly connected network of South Asian cities. In this sense, South Asia's urban landscape is following the same development path as that earlier traversed by the now developed regions of the world. The existence, emergence and growth of multi-city agglomerations are also notable because they represent expanded metropolitan areas in which the provision of infrastructure and basic urban services is often the responsibility of many different local governments. As such, compared to smaller towns or cities, these areas face more complicated governance challenges resulting from the difficulties of coordinating infrastructure and service delivery across local government jurisdictions. In the absence of appropriate institutions, these difficulties can contribute to inefficient outcomes in terms of infrastructure and basic urban service delivery with potentially negative effects for productivity (Ahrend et al. 2014).

3.3 Patterns of Urban Economic Growth

Turning from patterns of urban physical expansion to patterns of urban economic growth, we can reasonably treat the rate of growth in an urban area's total NTL intensity as a proxy measure of its rate of GDP growth. This follows from the strong linear relationships between the growth of NTL intensity and GDP at higher levels of spatial aggregation (i.e. at both the cross-country level and the level of Indian districts) that we documented in Sect. 2.3. These relationships suggest that an urban area's growth rate of GDP will be directly proportional to the growth in its NTL intensity.

Figure 11.7 shows the annualized NTL growth rates within the footprints of all South Asian cities and agglomerations whose physical expansion we have analyzed in the previous two sub-sections.[22] Visual inspection of this figure suggests that the fastest rates of urban economic growth over

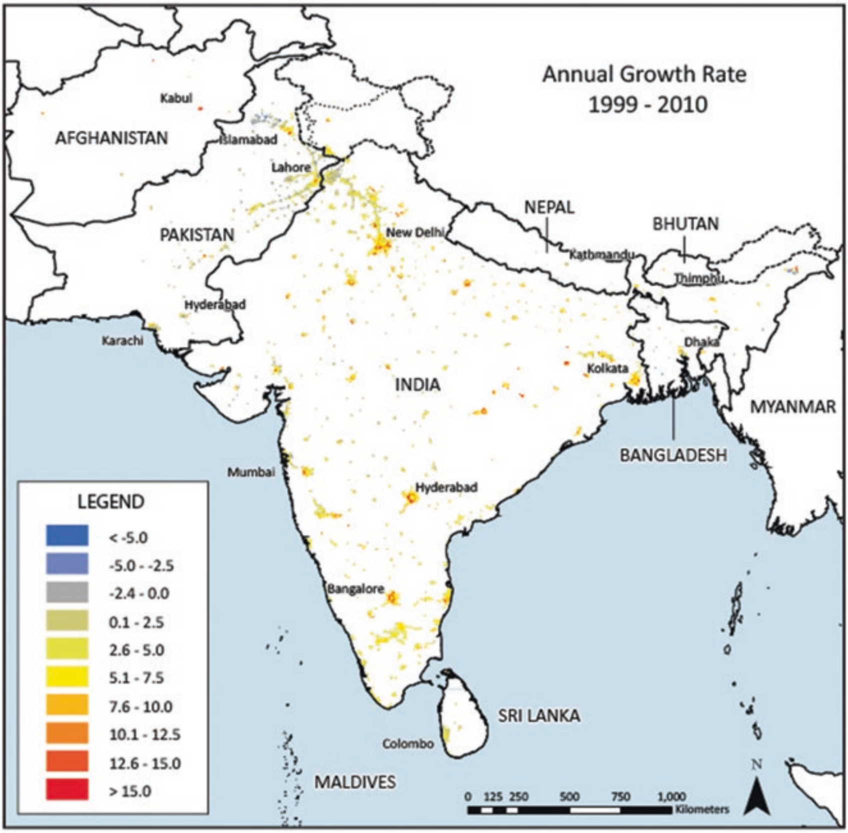


Fig. 11.7 Urban economic growth has been most rapid on the peripheries of major agglomerations (Source: Own construction)

the period 1999–2010 tended to be concentrated on the peripheries of the region's major agglomerations. Hence, we see growth rates of NTL intensity greater than 15 percent in, for example, selected parts of the peripheries of the Bangalore, Delhi and Hyderabad agglomerations in India, and the Lahore agglomeration in Pakistan. We can see this more clearly in Fig. 11.8. Thus, for the Bangalore agglomeration, the fastest rates of growth are clearly observed in a circular area surrounding the core city of Bangalore that includes the cities of Dasarahalli and Mahadevapura. A similar pattern exists for the Hyderabad agglomeration, while, for the

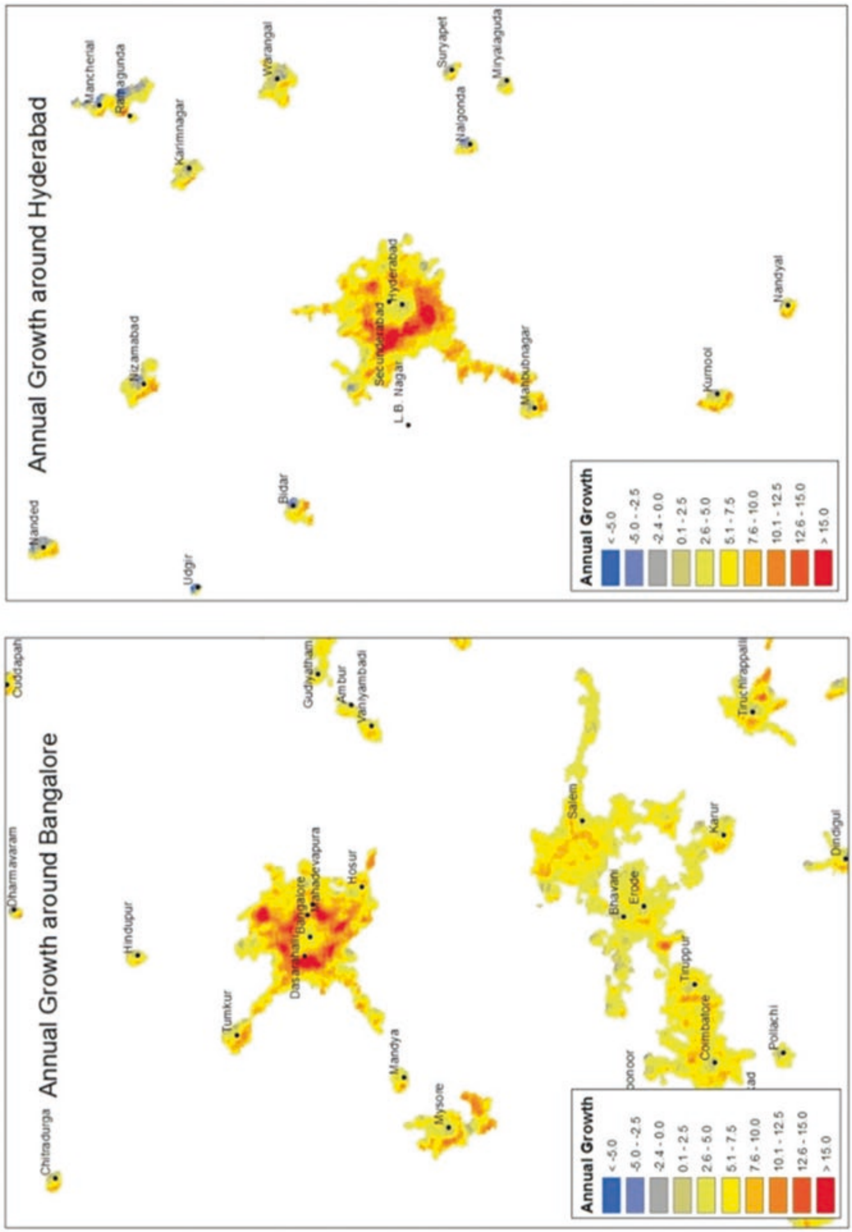


Fig. 11.8 Economic growth within the 2010 footprints of the agglomerations of Bangalore, Hyderabad (India), New Delhi and Lahore (Source: Own construction)

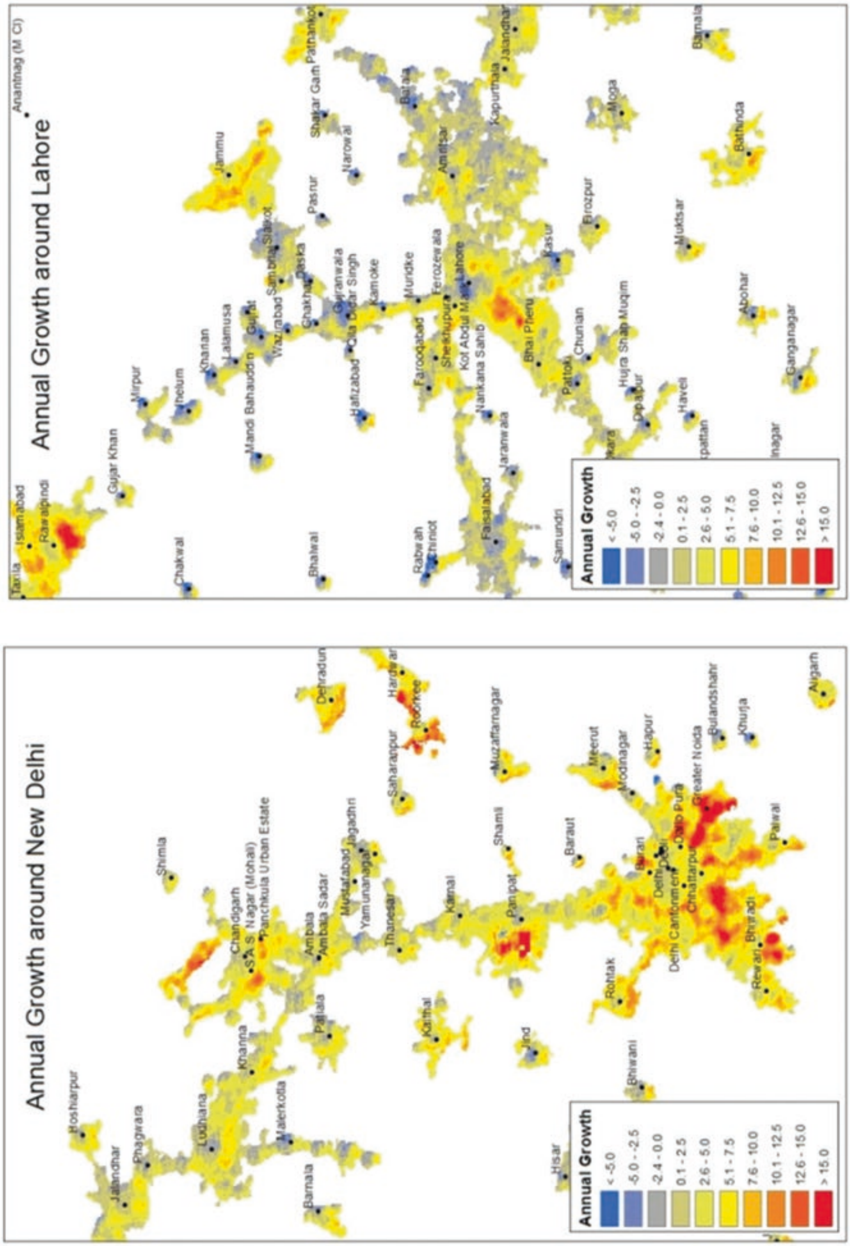


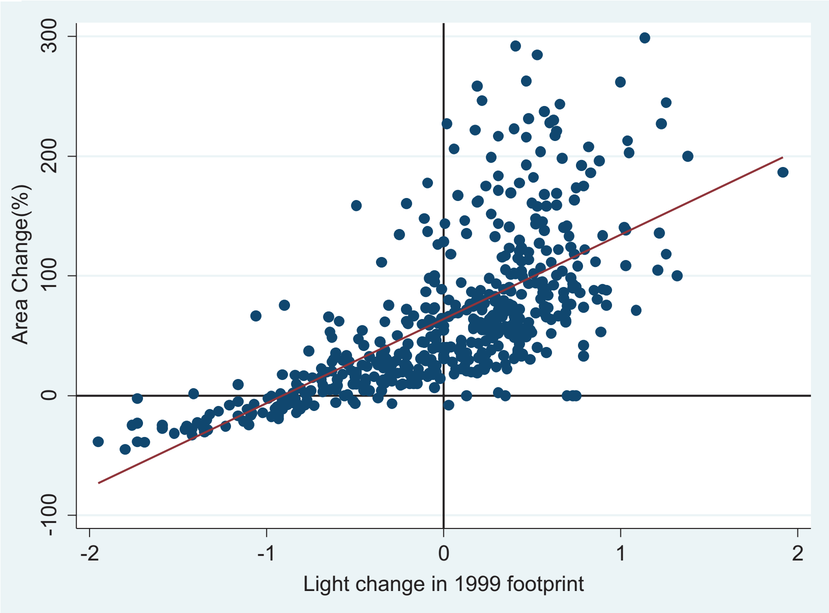
Fig. 11.8 (continued)

Delhi agglomeration, the fastest growth rates are generally to be seen to the southeast of Delhi in and around Greater Noida and to the north and south of Bhiwadi. Finally, for the Lahore agglomeration, we observe a hotspot of very fast growth to the southwest between Lahore and Bhai Pheru.[23]

3.4 Intensive Versus Extensive Growth of Cities and Agglomerations

From Figs. 11.7 and 11.8 it is clear that both agglomerations and cities across South Asia have not only differed in their rates of economic growth over the period 1999–2010 but also in terms of their patterns of growth. An interesting contrast in this respect is that between, for example, the Bangalore and Lahore agglomerations. For the Bangalore agglomeration, growth of NTL intensity in the periphery was accompanied by, albeit slower, growth of NTL intensity in the center. By contrast, for the Lahore agglomeration, the general growth in NTL intensity around the city of Lahore went together with a clear dimming of lights in the immediate vicinity of Lahore.[24]

Figure 11.9 examines this issue of different patterns of growth at a more general level for ‘single cities’, that is, for cities that were not part of an agglomeration in either 1999 or 2010. In doing so, it makes a distinction between two types of growth that a city can experience: (a) *extensive growth*, which is defined as growth in a city’s urban lit footprint between 1999 and 2010, and (b) *intensive growth*, defined as growth in the intensity of a city’s lights within its 1999 footprint over the period 1999–2010. The first thing that is evident from Fig. 11.9 is that there is a clear positive, and strongly statistically significant, relationship between intensive growth and extensive growth. The cities which experienced the fastest growth in NTL intensity within their 1999 footprints also tended to be the ones that increased most rapidly in area. This is consistent with the idea that fast economic growth within a city generates congestion pressures in, for example, land and property markets, which, by bidding up prices, then serves to also encourage growth in the periphery by creating incentives for firms and households to relocate.



Slope coefficient = 70.3875, S.E. = 3.4115
R^2 = 0.460

Fig. 11.9 Scatterplot of extensive growth versus intensive growth, single cities (Source: Own construction. Notes: Cities are excluded as outliers if they have absolute DFFITS values that exceed the critical value of 0.087 where the critical value is given as $2\sqrt{k/n}$ where k, the number of regression coefficients, equals 2 and n, the number of observations, equals 524)

A second notable point to arise from Fig. 11.9 is that there are three different types of single city in evidence in terms of patterns of intensive versus extensive lights growth. The first type of city is *thriving cities*; these are cities which fall in the top right-hand quadrant of Fig. 11.9. Such cities have experienced both positive intensive and extensive growth of night-time lights. 55 percent of the cities in Fig. 11.9 fall into this category. Meanwhile, the second type is what we may think of as *'donut' cities*; these are cities which, while they have experienced positive extensive growth, have also seen a dimming of lights within their 1999 footprints (31.5 percent of cities). As such, these cities, which appear in the top left-hand quadrant of Fig. 11.9, appear to be experiencing a relative hollowing out of economic activity at their centers. Finally, we have *dimming cities*, which have not only experienced a dimming of lights within their original footprints but also shrinkage of their footprints (11.8 percent of cities). These cities fall in the bottom left-hand quadrant of Fig. 11.9.[25]

The breakdown of types of single city for the different countries is shown in Table 11.3. We see that thriving cities are dominant in Bhutan, Sri Lanka, Afghanistan and, to a lesser extent, India. Afghanistan, India and Sri Lanka also have some donut cities. These are not as evident as in Bangladesh and, especially, Pakistan though. All countries, with the exceptions of Afghanistan and Sri Lanka, also have some dimming cities. However, such cities are a particularly notable feature of the urban landscapes of Pakistan, Bangladesh and Nepal.

The shrinkage of footprints witnessed for dimming cities is a result of NTL intensity falling below the $DN = 13$ threshold which we use to delineate urban areas. While it is hard to believe that cities have actually shrunk in terms of their physical size, and in this sense the appearance of shrinkage should be considered an artifact of the methods used, the general dimming of lights that has occurred in these cities is suggestive of slower relative economic growth compared to both thriving and 'donut' cities.[26] Given that dimming cities are most evident in Bangladesh, Pakistan and Nepal—together, these countries account for 43 out of the

Table 11.3 Breakdown of types of single city growth across countries

Country	Thriving	'Donut'	Dimming
Afghanistan	75.00	25.00	0.00
Bangladesh	7.14	46.43	46.43
Bhutan	100.00	0.00	0.00
India	69.11	23.58	5.69
Nepal	0.00	20.00	80.00
Pakistan	10.47	61.63	27.91
Sri Lanka	85.71	14.29	0.00
South Asia	**54.98**	**31.47**	**11.81**

Source: Own construction

Notes: Maldives has been excluded from the table because the only city from Maldives in the sample, Malé, does not fall into any of the three categories owing to the absence of any growth in its urban lit footprint between 1999 and 2010. Percentages for both India and South Asia overall do not add up to 100 because of the existence of one city that does not belong to any of the three 'types' (see footnote 25 on this point)

65 dimming cities shown in Fig. 11.9;[27] it seems likely that increased load shedding, resulting in power blackouts, is an important part of the story behind why we observe such cities. In cities where dimming is being driven by increased load shedding, this is suggestive of crumbling infrastructure, which, in turn, can be expected to contribute to slower relative economic growth. Failing infrastructure is, furthermore, likely to be correlated with other factors, such as poor urban governance and management, which contribute to slow growth.[28]

4 Urban Lights and Extreme Poverty

We finish our analysis by exploiting the NTL data to provide insights into the relationship between urbanization and extreme poverty across a comprehensive sample of South Asian districts.[29] This sample covers all South Asian countries with the exceptions of Afghanistan and the Maldives, for which we lack sub-national poverty data. We define extreme poverty as the proportion of a locality's population which lives on less than $1.25 per day (based on 2005 PPP exchange rates). The importance of this analysis is underscored by the fact that while South Asia has made tremendous progress in reducing the overall share of the population that lives in extreme poverty since 1999, around one-in-three people continued to live below the $1.25 per day global poverty line as of 2010 (Ellis and Roberts 2016).

Table 11.4 shows, first of all, that, for the region as a whole, there is a strong negative relationship between the overall intensity of urban NTL within a district and the proportion of its population that lives in extreme poverty. In other words, districts that, holding everything else constant, either have a greater proportion of their land dedicated to urban uses or have more vibrant urban centers also tend to have lower rates of extreme poverty. Although this is just a correlation, it is consistent with the idea that urban growth brings with it general benefits of poverty reduction. There are several mechanisms through which this may occur, including positive spillover effects from urban centers—which may, for example, be stimulated through trade or rural-urban migration flows—to surrounding rural regions within a district, implying that urban growth

Table 11.4 Rates of extreme poverty are lower in districts with more brightly lit urban areas and which have better access to other districts with brightly lit urban areas

Dependent variable	SAR	SAR	SAR
Poverty rate	(1)	(2)	(3)
ln(DN/Area)	−0.0206***		−0.0183***
	(0.0029)		(0.0030)
ln(Gravity)		−0.0428***	−0.0212*
		(0.0114)	(0.0116)
Constant	−0.0199	−0.5468**	−0.2895
	(0.1500)	(0.2122)	(0.2109)
n	692	692	692
R^2	0.546	0.520	0.546
Adj. R^2	0.515	0.491	0.517
F	17.6848	17.6330	19.0475
Prob>F	0.0000	0.0000	0.0000

Source: Own construction

Notes: *** significant at one percent level; ** significant at five percent level; * significant at ten percent level; F is the test statistic for an F-test of the joint significance of all explanatory variables with Prob > F being the corresponding p-value. In all columns, admin. level-one fixed effect as well as country fixed effect have been controlled

helps to generate rural poverty reduction within a district (Cali and Menon 2012).

Meanwhile, Table 11.4 also suggests that there exists a significant, although not quite as strong, negative relationship between a gravity measure of urban NTL intensity and a district's rate of extreme poverty. For any given district i, the gravity measure of urban NTL is given by $\sum_{j=1}^{N}\left(\mathrm{NTL}_j / d_{ij}^2\right)$, where NTL_j denotes the overall intensity of NTL in areas within district j that are classified as urban (i.e. in areas where $DN > 13$) and d_{ij} the distance between the centroids of districts i and j. This basically measures a district's proximity to brightly lit urban centers, including urban centers outside of the district. Hence, a district will tend to score highly on this gravity measure if: ***(a)*** the district itself is home to brightly lit urban centers; and/or ***(b)*** it neighbors districts with brightly lit urban centers. The fact that districts which perform well on the gravity measure tend to enjoy lower rates of extreme poverty is consistent with the idea that vibrant cities generate poverty reduction benefits which spillover not only to rural areas within the district but also to neighboring districts.[30] It is also consistent with the idea that improving connectivity and, hence, effectively, reducing distance between cities across districts through investments in transport infrastructure can have important poverty alleviating impacts. Crucially, the gravity measure of urban lights retains its significant negative relationship with extreme poverty even after controlling for the overall NTL intensity of a district's own urban centers, as well as vice versa (Column (3) in Table 11.4)).

Table 11.5 shows that the above regional results on the relationship between urbanization and extreme poverty are primarily driven by India which dominates the sample of districts (589 of the 692 sample of districts are in India). For Nepal and Pakistan, we also tend to find that district poverty rates are negatively correlated with both the overall intensity of a district's own NTL and the gravity measure of urban lights with some evidence of statistical significance.[31] By contrast, for Bangladesh, Bhutan and Sri Lanka, these variables appear to be of little importance in

Table 11.5 Heterogeneity across countries in the relationship between urbanization and extreme poverty

Dependent variable	BGD	BTN	IND	NPL	NPL	PAK	LKA
Poverty rate	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln(DN/Area)	0.0044	−0.0058	**−0.0212*****	**−0.0145****	–	**−0.0148***	−0.0007
	(0.0094)	(0.0050)	(0.0036)	(0.0066)		(0.0086)	(0.0034)
ln(Gravity)	0.0384	−0.0021	**−0.0374****	–	**−0.0543***	–	−0.0076
	(0.0399)	(0.0056)	(0.0160)		(0.0277)		(0.0066)
Constant	0.9032*	−0.0272	−0.5025*	0.1854***	−0.5283	0.0609	−0.0390
	(0.4804)	(0.0817)	(0.2587)	(0.0425)	(0.4043)	(0.0654)	(0.0742)
n	23	16	589	14	14	28	22
R^2	0.052	0.128	0.462	0.290	0.243	0.365	0.084
Adj. R^2	−0.042	−0.006	0.427	0.231	0.180	0.221	−0.012
F	0.5531	0.9516	13.1716	4.8942	3.8551	2.5324	0.8739
Prob>F	0. 5837	0.4114	0.0000	0.0471	0.0732	0.0590	0.4334

Source*:* Own construction

Notes: *** significant at one percent level; ** significant at five percent level; * significant at ten percent level; F is the test statistic for an F-test of the joint significance of all explanatory variables with Prob > F being the corresponding p-value. Admin level-one fixed effect have been controlled for India and Pakistan explaining variations in extreme poverty across districts. In the cases of Bhutan and Sri Lanka this may be because, as documented by the World Bank (2015) and Ellis and Roberts (2016), rates of extreme poverty are much more spatially uniform across districts.

5 Summary and Conclusions

This chapter has made use of night-time lights data that has been remotely collected by satellites orbiting the earth to examine both detailed spatial patterns of urban physical expansion and economic growth across the South Asia region over the period 1999–2010. In doing so, it has shown that the overall pace of expansion of cities, in terms of their spatial footprints or extents, has been extremely rapid, occurring at roughly twice the rate of urban population growth. Accompanying this, there has been the increasing emergence of multi-city agglomerations within continuously lit belts of urbanization, including the emergence of a Delhi-Lahore mega-agglomeration that straddles the Indian-Pakistan border. Rates of urban economic growth in the region, furthermore, appear to have been fastest on the peripheries of the most important of these agglomerations. Greater Noida, for example, stands out as having experienced very fast economic growth. More generally, while intensive growth of cities is clearly positively correlated with extensive growth, meaning that the cities which are growing fastest at their cores are also the ones that are tending to expand outward most rapidly, there is also evidence of different patterns of intensive and extensive growth across cities.

However, while the above describes the overall regional picture, there are also clearly important differences in dynamics across countries. The rapid rates of expansion of overall urban lit area witnessed in Afghanistan, Bhutan and, to a lesser extent, India and Sri Lanka can, therefore, be contrasted with the relatively much more sluggish rates seen in Bangladesh, Pakistan and, especially, Nepal. Likewise, while the process of formation of multi-city agglomerations is well under way in India, Pakistan and (taking account of its much smaller population size) Sri Lanka, both Bangladesh and Nepal are at a much more incipient stage of this process. The process, meanwhile, has yet to start at all in Afghanistan.

The above results have important implications for policy. For example, the increasing emergence of large multi-city agglomerations presents opportunities for the exploitation of static and dynamic agglomeration economies which can help to drive growth in productivity and, therefore, economic development. At the same time, however, realizing these opportunities will require a need for improved metropolitan governance and mechanisms for horizontal coordination between different municipalities.

Notes

1. This chapter is based on research that was originally conducted for the World Bank's Flagship Report *Leveraging Urbanization in South Asia: Managing Spatial Transformation for Prosperity and Livability* (Ellis and Roberts 2016). The author thanks both Peter Ellis and Ming Zhang for their feedback in performing this research. He also thanks both Benjamin Stewart and Katie McWilliams for the invaluable GIS support that they provided and also Jane Park and Eshrat Waris for excellent research assistance. The findings, interpretations and conclusions expressed in this chapter are entirely those of the author. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.
2. The urban population and share figures referred to in this paragraph are based on *World Urbanization Prospects: 2011 Revision* data (https://esa.un.org/unpd/wup/). In this chapter, we follow the World Bank as defining South Asia as consisting of the following eight countries: Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka.
3. Official definitions of urban areas differ across South Asian countries and, indeed, across countries globally. This means that comparisons of the pace of urban growth or urbanization across countries and regions based on official data need to be treated with caution (World Bank 2008; Ellis and Roberts 2016). One advantage of the data and methods that we use in this chapter is that they allow for the consistent definition of urban areas across countries.

4. More accurate results on patterns of urban physical expansion can be derived using higher-resolution satellite imagery (e.g. MODIS or LandSat imagery). However, the increased accuracy comes at a greater cost in terms of processing time, and these sources of imagery cannot be used to analyze patterns of urban economic growth.
5. $1.25 per day was the World Bank's global poverty line until the year 2015. The global poverty line is currently defined by a consumption threshold at $1.90, using the 2011 PPP exchange rates.
6. NOAA also provides NTL products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) instrument that was launched onboard the Suomi National Polar-Orbiting Partnership satellite in 2011. While this data is of higher resolution than the NTL data that we use in this chapter, it only has a limited temporal coverage and is, therefore, not suitable for the study of urban growth dynamics over the medium to long run.
7. See https://ngdc.noaa.gov/eog/gcv4_readme.txt.
8. The latest version of this product is available for download from http://ngdc.noaa.gov/eog/dmsp/download _radcal.html. The analysis reported in this chapter is based on an earlier (pre-general release) version of this product that was supplied by NOAA's National Centers for Environmental Information Earth Observation Group.
9. For a short technical discussion, see http://ngdc.noaa.gov/eog/dmsp/download_radcal.html.
10. Some additional cities, most notably, the secondary Sri Lankan cities of Anuradhapura, Galle, Batticaloa, Matara and Kurunegala, were also included in the sample despite not meeting this population criterion on the basis that they were considered to be important to understanding Sri Lanka's urbanization process.
11. These cities had a combined 2010 population of approximately 270.6 million, which equated to 54 percent of South Asia's overall urban population. This, however, probably considerably underestimates the share of South Asia's total urban population covered by our analysis. This is because the urban footprints that we define for the cities probably also include smaller urban settlements. A full list of cities in the sample, along with their circa 2010 populations, is available on request.
12. The land-use map used is the European Space Agency's GlobCover 2009 map, which can be downloaded from http://due.esrin.esa.int/globcover/. Zhou et al. (2015) report similar results based instead on the comparison of NTL data with a MODIS land-cover map.

13. Small et al. (2011) find that a threshold around $DN = 13$ results in a plausible estimate of the exponent of the power law for rank-size distributions of cities.
14. The full regression results discussed in this sub-section are available upon request from the chapter author.
15. These pooled regressions take the form $\ln(GDP_{i,t}) = \alpha + \beta D_{t=2010} + \gamma \ln(DN_{i,t}) + \varepsilon_{i,t}$ where the subscripts i and t denote the country or district and year respectively. $D_{t=2010}$ is a dummy variable that takes on the value 1 if $t = 2010$ and zero if $t = 1999$. The standard errors (SE) reported are robust standard errors. Regressions were also estimated that included an interaction term between $D_{t=2010}$ and $\ln(DN_{i,t})$. For both the samples of South Asian countries and Indian districts, the estimated coefficient on this interaction term was found to be small and not statistically significant.
16. The results reported in this sub-section partly draw on CIESIN (2013).
17. One caveat to be kept in mind when considering this result is that the growth rate for urban population is calculated using *World Urbanization Prospects, 2011 Revision*, data, which is based on official national definitions of urban areas. To the extent that these definitions tend to underestimate the sizes of urban areas, they will also tend to underestimate urban population. This will lead to biased estimates of urban population growth rates if the degree of underestimation has changed over time.
18. We name an agglomeration after the most populated city which falls within its area.
19. To be noted, however, is that, in both 1999 and 2010, the distribution of the number of cities per agglomeration is heavily right skewed. Thus, 17 out of 37 agglomerations in 1999 consisted of just two cities. In 2010, 22 out of 45 consisted of just two cities.
20. This estimate of the Coimbatore agglomeration's population was arrived at by using GIS techniques to layer Coimbatore's urban footprint with gridded population data for 2011 taken from *LandScan* (http://web.ornl.gov/sci/landscan). The same methods were used to arrive at the estimate of the population of the Delhi-Lahore agglomeration that is reported in the next paragraph.
21. The geographical area of the 'Hot Banana' covers South East England, the Netherlands, Belgium, the Ile de France, the Ruhr area, South East France, Southern Germany and Northern Italy.
22. More precisely, the figure shows the (natural) log change in DN values at the pixel level over the period 1999–2010 normalized by the length of the sample period.

23. The figure for Lahore also highlights the fast growth that has been occurring to the south of Rawalpindi.
24. There is also some evidence of dimming lights on the extreme periphery of the Lahore agglomeration.
25. Figure 11.9 also implies the potential existence of a fourth type of single city: namely, the seemingly paradoxical 'shrinking, but thriving city' that has experienced negative extensive growth, but positive intensive growth. We would expect such cities to fall into the bottom right-hand quadrant of Fig. 11.9. As can be seen, there is only one city that (just) falls into this category. However, given that both its rates of extensive and intensive growth are very close to zero, it is perhaps better characterized as a 'stagnant' city.
26. An important point to note is that dimming lights do not necessarily imply absolute economic decline. This follows from Figs. 11.2 and 11.3. In particular, in these figures, we see that the estimated cross-country relationships between the growth of GDP and the growth of lights all have positive intercept values. The corollary of this is that, even when the growth of lights is negative, the rate of GDP growth is likely to be positive. For middle- and low-income countries, for example, negative GDP growth will only be observed if $\Delta \ln(DN) < 1.26$.
27. India accounts for the remaining 22 dimming cities.
28. In addition to their heavy presence in Bangladesh, Nepal and Pakistan, dimming cities were also, on average, both smaller and less bright to begin with than both thriving and donut cities. Hence, in 1999, the mean size of a dimming city's urban footprint was 66.33 km^2 compared to 83.89 km^2 for donut cities and 111.03 km^2 for thriving cities. Meanwhile, mean 1999 NTL intensity values (i.e. DN values in natural logs) were 4.19, 4.24 and 4.25 for dimming, donut and thriving cities, respectively.
29. We define districts as the second tier of sub-national administrative units across the sample countries.
30. This is consistent with so-called 'New Economic Geography' theory, which implies the existence of a positive sub-national relationship between wages and levels of firm access to both final and intermediate goods markets (Fujita et al. 1999).
31. For both Nepal and Pakistan, both ln(DN/Area) and ln(Gravity) are statistically insignificant when included in regressions together. This is likely because the small number of observations for these two countries, 14 for Nepal and 28 for Pakistan, prevents the data from disentangling

the influence of these two variables. For Bangladesh, Bhutan and Sri Lanka, ln(DN/Area) and ln(Gravity) are always insignificant (with estimated coefficients close to zero) irrespective if they are included in the regressions jointly or separately.

References

Ahrend, R., Farchy, E., Kaplanis, I., & Lembcke, A. (2014). *What makes cities more productive? Evidence on the role of urban governance from five OECD countries, OECD regional development working papers, 2014/05*. Paris: OECD Publishing.

Alder, S. (2015). *Chinese roads in India: The effect of transport infrastructure on economic development* (Working paper). University of North Carolina at Chapel Hill.

Angel, S., Parent, J., Civco, D. L., & Blei, A. M. (2011). *Making room for a planet of cities*. Cambridge, MA: Lincoln Institute of Land Policy.

Baum-Snow, N., & Turner, M. A. (2012). *Transportation and the decentralization of Chinese cities*. Unpublished. Brown University, Providence.

Cali, M., & Menon, C. (2012). Does urbanization affect rural poverty? Evidence from Indian districts. *World Bank Economic Review, 27*(2), 171–201.

Center for International Earth Science Information Network (CIESIN). (2013). Report for phase I: Mapping, quantification and analysis of evolution of patterns of urban physical extent and morphology in South Asian cities, 1999–2010. Background paper for the World Bank's Flagship Report *Leveraging Urbanization in South Asia: Managing Spatial Transformation for Prosperity and Livability*, World Bank, Washington, DC.

Croft, T. A. (1978). Night-time images of the earth from space. *Scientific American, 239*, 86–98.

Doll, C. N. H. (2008). CIESIN thematic guide to night-time light remote sensing and its applications. *Center for International Earth Science Information Network (CIESIN).* Columbia University, Palisades.

Ellis, P., & Roberts, M. (2016). *Leveraging urbanization in South Asia: Managing spatial transformation for prosperity and livability*. Washington, DC: World Bank.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economy: Cities, regions and international trade*. Cambridge, MA: MIT Press.

Harari, M. (2016). *Cities in bad shape: Urban geometry in India* (Working paper). The Wharton School, University of Pennsylvania.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2011). A bright idea for measuring economic growth. *American Economic Review, 101*(3), 194–199.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review, 102*(2), 994–1028.

Pinkovskiy, M. L. (2013). *Economic discontinuities at borders: Evidence from satellite data on lights at night*. Unpublished. Massachusetts Institute of Technology, Cambridge, MA.

Small, C., Elvidge, C., Balk, D., & Montgomery, M. (2011). Spatial scaling of stable night lights. *Remote Sensing of Environment, 115*, 269–280.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth. *Review of Economic Studies, 83*(3), 1263–1295.

Tewari, D., Alder, S., & Roberts, M. (2017). *India's urban and spatial development in the post-reform period: An empirical analysis* (Working paper). New Climate Economy.

World Bank. (2008). *World development report 2009: Reshaping economic geography*. Washington, DC: World Bank.

World Bank. (2012). *Turning Sri Lanka's urban vision into policy and action*. Colombo: World Bank.

World Bank. (2015). *Sri Lanka: Ending poverty and promoting shared prosperity – A systematic country diagnostic*. Washington, DC: World Bank.

Zhang, Q., & Seto, K. C. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment, 115*, 2320–2329.

Zhou, N., Hubacek, K., & Roberts, M. (2015). Analysis of spatial patterns of urban growth across South Asia using DMSP-OLS nighttime lights data. *Applied Geography, 63*, 292–303.

12

Production Functions, the Kaldor-Verdoorn Law and Methodology

Marc Lavoie

1 Introduction

I first met John McCombie in December 1983. I was then visiting my father who was the general consul of Canada in Melbourne. I had written to Robert Dixon about a paper of his in *the Journal of Post Keynesian Economics*, and had told him that I would be in Melbourne around Christmas. Dixon invited me to meet him and to have lunch at the University of Melbourne. John McCombie then had a position there, and so he joined us at lunch, and so did Geoffrey Harcourt who happened to be in Melbourne at the time. The only thing I remember is a brief after-lunch discussion with John McCombie about whether the Kaldor-Verdoorn law could also be subjected to the critique of the neoclassical production function that Anwar Shaikh (1974) had

M. Lavoie (✉)
University Sorbonne Paris Cité, Paris, France

University of Paris 13 (CEPN), Paris, France

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_12

made, on the basis of its analogy with the national accounts. McCombie at the time had just published a few papers on the Kaldor-Verdoorn law, and I was using Shaikh's paper in my class on growth theory, so I was curious to know his feeling about this. I do not recollect what his reply was, but McCombie does not seem to recall this conversation either because he has written that he discovered the article of Shaikh by pure luck, 'almost by serendipity' (Felipe and McCombie 2013, p. 11).

I have met John McCombie many times since, in particular when he invited me to participate in a conference on open-economy macroeconomics which was held in 2002 at Emmanuel College, at the University of Cambridge, where I presented a stock-flow consistent model of the Eurozone (Lavoie 2003). One of our recent encounters was at the 2013 Berlin post-Keynesian summer school, where McCombie was asked to do a lecture on open-economy macroeconomics and the balance-of-payment constraint, and to present the main features of his new book on the critique of the neoclassical aggregate production function (Felipe and McCombie 2013). There Eckhard Hein and I had the pleasure of interviewing him for the *European Journal of Economics and Economic Policies* (Hein and Lavoie 2015). And even more recently, John was kind enough to do a long review of my latest book (McCombie 2015).

McCombie is mostly known for his articles on the balance-of-payment constraint, his critiques of the neoclassical aggregate production function and his work on the Kaldor-Verdoorn law. He has also done some incursions into methodology. In what follows, I will skip the first topic and start with a discussion of the neoclassical aggregate production function. The next section will be devoted to the question that I had put to McCombie in December 1983, that is, whether the critique of the neoclassical production function has any relevance for a possible critique of the Kaldor-Verdoorn law. In the final section I will deal with some methodological issues, mostly associated with the difficulties of making changes in macroeconomic theory.

2 Neoclassical Aggregate Production Functions

2.1 Anwar Shaikh's HUMBUG Data

McCombie (1987) published his first paper critiquing the empirics of the neoclassical production function in 1987, going over the critiques made by Shaikh and Herbert Simon. It is also in 1987 that I published my first book, in French, on the capital controversies and growth theory (Lavoie 1987). Early on in my career, I was fascinated by Anwar Shaikh's (1974, 1980) HUMBUG critiques of the neoclassical production function. These appeared in my chapter on the neoclassical response to the Cambridge capital controversies.[1] One of these responses, besides the recourse to general equilibrium theory or to temporary equilibrium, or the refusal, plain and simple, to acknowledge the existence and impact of the capital controversies, was the recourse to empirics. Several neoclassical authors at the time had insisted that the numerous empirical successes of neoclassical production functions such as the Cobb-Douglas production function or the constant-elasticity-of-substitution (CES) production function had demonstrated that the Cambridge capital controversies, with their associated rejection of the principle of relative scarcity due to the possible presence of reswitching and capital reversing, were of no practical significance.

There is the well-known *cri-du-coeur* of Charles Ferguson, who wrote that the validity of neoclassical theory was depending on econometric or empirical arguments rather than theoretical ones and hence that he had faith (Carter 2011). But others made the same pledge. Here I mention just a few additional examples, from authors that were well-known at the time of the Cambridge capital controversies. Sato (1974, p. 383) thought that neoclassical theory could be vindicated by having recourse to empirical analysis, in particular the estimation of CES functions. Jorgenson (1974) also argued that the CES production function seemed to be most appropriate when its constant elasticity was close to unity, that is, when it was no different or little different from the Cobb-Douglas production function. Similarly Bronfenbrenner (1971, p. 474) argued that the theory of distribution based on marginal productivity was vindicated by the numerous empirical successes in all fields of the Cobb-Douglas production function.

Thus, Shaikh's demonstration, that the empirical successes and good statistical fits of the aggregate Cobb-Douglas production function were due to the fact that the algebraic representation of such a production function was no different from a mathematical transformation of the national accounts identity, should have given a halt to this instrumental defence of neoclassical theory. But of course, retrospectively, we know that it did not. In Lavoie (1987, p. 116), I also recalled that Herbert Simon (1979), in his acceptance speech for the Bank of Sweden prize in economic sciences in memory of Alfred Nobel, had himself underlined this tight link between neoclassical production functions and the accounting identities, demonstrating this in particular for the case of the CES function. I had also noted the curious fact that Simon (1979) had not cited Shaikh (1974) who had written on precisely this topic, despite the fact that Robert Solow had been identified as one of the readers of his draft (Lavoie 1987, p. 196). Obviously, Solow knew about Shaikh's (1974) paper since he had written a rather nasty (and misleading) comment on it, as Shaikh (1980) later demonstrated, and to which he was not allowed to reply.[2]

As recalled by Felipe and McCombie (2011–12, p. 276), in response to a letter that I had sent to Simon, pointing to the similarities between his and Shaikh's argument, Simon replied that, being less connected with economics, he had to rely on friends and colleagues to keep track of the literature, thus implying that Solow had omitted to inform him about Shaikh's work. It is probable that Solow did not want to give Simon further arguments against the marginal productivity theory of factor pricing. In a conversation with me, Shaikh observed that Solow was still angry about his HUMBUG article since, more than 30 years after its publication, Solow steadfastly refused to shake hands with him during a ceremony in honour of Modigliani which was held at the New School in 2007.

Shaikh (1990) published a third version of the HUMBUG production function which contains an interesting graph. It illustrates the case of an economy based on a Leontief production function with fixed technical coefficients, which is subjected to technical progress of the Harrod-neutral sort, with a constant capital to output ratio. Shaikh (op. cit) shows the straight upward-sloping line that links capital per head on the horizontal axis and output per head on the vertical axis. Neoclassical theory will assume instead the presence of a standard neoclassical production function with diminishing returns, and hence a relationship between capital per head and output per head which has the usual curvature. Neoclassical authors would thus interpret the data of this economy by claiming that there has been a move along the non-linear neoclassical production function accompanied by a shift of the entire production function. Thus, even if the technology is of the Leontief type, neoclassical economists running standard regressions would pretend that they have demonstrated the existence of a well-behaved neoclassical production function. But this was precisely the point that Nicholas Kaldor (1957) had made several years before, when claiming that "any sharp or clear-cut distinction between the movement along a 'production function' with a given state of knowledge, and a shift in the 'production function' caused by a change in the state of knowledge is arbitrary and artificial" (p. 596).

2.2 Reductio Ad Absurdum Proofs

Shaikh (2005) produced a 'reductio ad absurdum' proof of this. He generated data of a fictitious economy subjected to a Goodwin cycle, where technology, as above, is of the Leontief type with Harrod-neutral technical progress and with mark-up pricing. Still, despite all this, once technical progress is assessed in an appropriate way, Shaikh shows that the data can appear to have a high fit with a Cobb-Douglas production function, as he obtains a regression with a very high R^2 and an estimated output elasticity of capital which is very close to the actual profit share, just as neoclassical theory would have it. This is so despite the fact that, by construction, the data has nothing to do with neoclassical theory and violates all of its usual assumptions.

This kind of 'reductio ad absurdum' proof is the best proof that can be offered to demonstrate that the apparent empirical successes of the neoclassical production functions have nothing to do with reality conforming to neoclassical theory.[3] The clearest and most astute such proof was offered by McCombie (2001) himself. He generates microeconomic data, where the output elasticity of labour is 0.25 while the output elasticity of capital is 0.75, so that the sum of the two elasticities is equal to unity, thus assuming constant returns to scale. Inputs and outputs are assumed to be homogeneous, so as to avoid any problem of aggregation. Running a regression over this hypothetical economy, with some random fluctuations, and with physical output being a function of labour and the physical value of machines, the estimates of the output elasticities turn out to be equal to those that were assumed by construction. However, things are quite different when regressions are conducted on the basis of deflated monetary values, that is, constant-price values. McCombie assumes that firms set prices on the basis of a mark-up procedure, with the wage share being 75 per cent while the profit share is only 25 per cent. Running a regression over the same hypothetical economy, but using the deflated values, he finds that the apparent estimate of the output elasticity of labour is now 0.75, instead of the 0.25 true output elasticity of labour that was assumed by construction in the data.

McCombie (2001) thus provides an undeniable proof that regressions over deflated values, the only ones that economists can run in the case of macroeconomic studies or even at the industrial level, will necessarily provide an estimate of the wage share in the economy instead of an estimate of the output elasticity of labour. Thus, as I conclude in Lavoie (2014) "even if the technology is from Mars and Martians manage to produce output independently of inputs, provided Martian firms follow some form of cost-plus pricing, the regressions over deflated values will tell us that the Martians use Cobb-Douglas production technology with diminishing returns, constant returns to scale, and factor pricing following the principles of marginalism" (p. 60).

As recalled by Lavoie (2008, 2014, pp. 60–62) and Felipe and McCombie (2013, pp. 302–306), my former co-author Wynne Godley also engaged into this kind of 'reductio ad absurdum' proof in a paper that was unjustly neglected (Anyadike-Danes and Godley 1989). They constructed a hypothetical economy where nominal wages, employment and output are growing independently of each other, with some random fluctuations and with prices once again being set by some cost-plus procedure. Still, a regression equation similar to those ran by the likes of Layard et al. (1991) produced a statistical and significant relationship between real wages and employment, while employment did not seem to depend either on current output or lagged employment, whereas by construction there was no relationship between real wages and employment while current employment was heavily dependent on previous employment.

As noted in an appendix by Felipe and McCombie (2013, pp. 308–310), I also myself provided a kind of 'reductio ad absurdum' proof when criticizing the wage-setting/price-setting model of Layard, Nickell and Jackman, or what became to be known as the WS-PS model of the NAIRU (Lavoie 2000, 2008). The 2000 paper was written as a reaction to an article by Cotis et al. (1998) which claimed to explain the evolution of the NAIRU from an econometric estimation of this WS-PS model and that claimed that the model was not contradicted by the data. The authors marvelled at the fact that their empirical estimate of the PS curve was fully compatible with the first-order conditions of a well-behaved neoclassical production function, with diminishing marginal product of labour, perfect competition and factor pricing at the value of the marginal product. Their regression had uncovered a positive relationship between the log of real wages and the rate of unemployment. The first author, Jean-Philippe Cotis, had been the chief economist at the OECD and had just then been named head of the French statistical agency—the INSEE—so this was not the work of some innocent bystander. Once again, it was shown that both their so-called medium-run and long-run equilibrium unemployment rates could be derived from the income side of the national income and product accounts and not from some behavioural equation tied to the neoclassical theory of labour demand.[4]

2.3 General Consequences

All these proofs demonstrate that there is no empirical support for neoclassical production and distribution theory. Orthodox authors decorate their theories, they do not actually attempt to verify them, let alone falsify them. They calibrate their production models; they do not actually demonstrate that their behavioural equations based on neoclassical theory are the proper description of the way the world works. As Kaldor (1972) said a long time ago, "the role of empirical estimation is to 'illustrate', or to 'decorate' the theory, not to provide support to the basic hypothesis (as for example, in the case of numerous studies purporting to estimate the coefficients of production functions)" (p. 1239).

It is sometimes objected, because critics have paid so much attention to the Cobb-Douglas production function, that only this production function is subjected to the threat of reproducing the identities of the national accounts. But as mentioned earlier, the CES production function was already under attack from Simon. Furthermore, another contribution of McCombie is to have proven that indeed the CES production function and the translog production function were subjected to the very same criticisms (McCombie and Dixon 1991; Felipe and McCombie 2001). This is an important contribution because Kaldor's stylized fact of a constant wage share has been undermined over the last three decades, thus generating better fits with the CES production functions than with the Cobb-Douglas function, and thus leading several economists to adopt the CES function (while also rejecting Leontief production functions). It is thus important to underline the fact that CES functions seem to perform better now because the wage share has been decreasing over the last 30 years or so, not because they are a better representation of the real production process.

The studies of Shaikh, McCombie, Felipe and others show that the econometric estimates of neoclassical production functions based on deflated monetary values, where direct physical data cannot be used, yield pure *artefacts*, that is, purely imaginary results. This affects all of neoclassical applied aggregate work that relies in some way on well-behaved production functions and profit-maximizing conditions: NAIRU measures, labour demand functions and wage elasticities (Felipe and McCombie 2009); investment theory; measures of multifactor productivity or total factor productivity growth (Felipe and McCombie 2007); estimates of endogenous growth; theories of economic development; theories of income distribution; estimates of cost functions; measures of potential output; estimates of the impact of changes in the minimum wage, social programmes or in tax rates.

Even when setting aside problems of aggregation, neoclassical economists are claiming to measure something, whereas they are really measuring something entirely different. One may wonder, however, whether the critique could also affect elements of post-Keynesian theory.

3 The Kaldor-Verdoorn Law

3.1 A First Look at the Similarities with the National Accounts Identity

McCombie has been an early advocate of the Kaldor-Verdoorn law, and he has written an extended survey of the studies that have been devoted to it (McCombie 2002). In the introduction of this chapter, I mentioned that when meeting John McCombie in 1983, I asked him if the Kaldor-Verdoorn law could be subjected to the same problems as the neoclassical production function. The discussion did not produce any conclusion, but in my 1992 book, I made the effort to at least reconsider the issue (Lavoie 1992, pp. 322–324).

It has been known for a long time that Kaldor's (1957) technical progress function, which can be considered as a theoretical version of the empirical Kaldor-Verdoorn law, could be rewritten under the form of a dynamic Cobb-Douglas production function, and hence that it could be rewritten under the form of the national accounts identity. The technical progress function, in its linear form, has been formalized as:

$$\lambda = \lambda_0 + \lambda_k \hat{k} \tag{12.1}$$

where λ is the rate of technical progress (the growth rate of output per unit of labour), while $\hat{k}$ is the growth rate of the capital to labour ratio.

A Cobb-Douglas production function of the sort $q = e^{\mu t}K^{\alpha}L^{1-\alpha}$ can be rewritten as output per unit of labour, hence as $y = e^{\mu t}k^{\alpha}$, with y the output per labour, so that in growth terms, we have a relationship which appears to be no different from the preceding one:

$$\hat{y} = \mu + \alpha \hat{k} \tag{12.2}$$

As to the national accounts, it can be shown that their income side can be rewritten in a dynamic form which resembles the above, as we get:

$$g = (1-\pi)\hat{\omega} + \pi \hat{r} + \pi \hat{K} + (1-\pi)\hat{L} \tag{12.3}$$

where g is the growth rate of output, $\hat{K}$ is the growth rate of the capital stock, $\hat{L}$ is the growth rate of the labour force, π is the profit share, $\hat{\omega}$ is the growth rate of real wages and $\hat{r}$ is the growth rate of the rate of profit.

By recalling that $\hat{k}$ is the growth rate of the capital to labour ratio, Eq. (12.3) can be rewritten as:

$$g - \hat{L} = \hat{y} = \tau + \pi \hat{k} \tag{12.3A}$$

with $\tau = (1-\pi)\hat{\omega} + \pi \hat{r}$

From an elementary point of view, these three Eqs. (12.1), (12.2) and (12.3A) look quite alike. Both Kaldor's technical progress function and the Cobb-Douglas production function could be brought back to the national accounts identity. What about the Kaldor-Verdoorn law? Does it suffer from the same fate? Could it also be a statistical artefact? Although the Kaldor-Verdoorn law says that the rate of technical progress, that is, the growth rate of labour productivity, is a function of the growth rate of output in the manufacturing industry, it is often written as a function of the growth rate of GDP. With g standing once again for the growth rate of overall economic activity, the Kaldor-Verdoorn relationship may be written as:

$$\lambda = \lambda_0 + \lambda_g g \tag{12.4}$$

Combining the Kaldor-Verdoorn relation (Eq. 12.4) with Kaldor's technical progress function (Eq. 12.1), we obtain what Michl (1985) calls the augmented technical progress function:

$$\lambda = \lambda_0 + \lambda_g g + \lambda_k \hat{k} \tag{12.5}$$

Writing the extended form of the national accounts dynamic identity, that is, combining Eq. (12.3A) with the value of its τ component, we obtain something that does look highly similar to Eq. (12.5):

$$\hat{y} = \pi \hat{r} + (1 - \pi)\hat{\omega} + \pi \hat{k} \tag{12.6}$$

When running his regression on the augmented technical progress function, Michl (1985) finds an estimate of the λ_k coefficient, which is very close to the share of profits in manufacturing, that is, around 0.38 and 0.40, thus corresponding to the π value in the national accounts identity. Furthermore, since $\hat{r}$ at the time was close to zero, with no trend in the rate of profit, one would expect the λ_0 parameter to be not significantly different from zero, which is also what Michl (op. cit) obtains. So far, the estimates are in line with the national accounts identity. However, the λ_g coefficient in Eq. (12.5) provides us with a piece of information which is not present in the national accounts identity given by relation (Eq. 12.6). It says that faster rates of growth of output (g) are associated with faster rates of growth of real wages ($\hat{\omega}$). The national accounts do not yield such a prediction. The Kaldor-Verdoorn law says that there is a relationship between the growth rate of output and the growth rate of output per labour and hence by extension that there might be a relationship between the growth rate of real wages and the growth rate of *output*. The national accounts by contrast tell us that there is a relationship between the growth rate of real wages and the growth rate of *output per labour*, not the growth rate of output.

Note, however, that the latter relationship may put in jeopardy another behavioural equation often found in post-Keynesian economics, that is, the dynamic Webb effect, also called the Marx or the Hicks effect. This causal relationship going from the growth in real wages to the growth in labour productivity is emphasized, for instance, by Hein and Tarassow (2010) and Storm and Naastepad (2012) in their discussion of productivity regimes. However, when calculating their productivity regimes, they suppose that the rate of technical change is influenced by both the Kaldor-Verdoorn effect and the Webb effect. Thus these authors include simultaneously the g and $\hat{\omega}$ variables in their estimates of their effect on the growth rate of labour productivity, so that they have an equation that does not correspond to the national accounts identity. This equation, in the same notations, is given by:

$$\lambda = \lambda_0 + \lambda_g g + \lambda_w \hat{\omega} \tag{12.7}$$

3.2 Another Reductio Ad Absurdum Proof

McCombie himself has recently tackled the possible relationship between the Kaldor-Verdoorn law and the national accounts identity. This is done in the paper of McCombie and Spreafico (2016). The authors start by noting what we just said above, that is, the technical progress function in its linear form can be brought back to a Cobb-Douglas production function and hence to the dynamic version of the national accounts. Hence the technical progress function "also suffers from the criticisms that Kaldor made of the neoclassical production function" (McCombie and Spreafico 2016, p. 1118). But what about the Kaldor-Verdoorn law? McCombie and Spreafico (2016) show that "Verdoorn's law could be regarded as a specification of the linear technical progress function allowing for the possibility of increasing returns to scale" (p. 1134). Does it mean that the Kaldor-Verdoorn effects are just as spurious as those of the neoclassical production function?

To convince us that they are not, McCombie and Spreafico once more resort to the use of the constructed data of a hypothetical economy. This is a highly useful method, for we know the true data that underlies the estimates that are being calculated. They test the Kaldor-Verdoorn relation given by Eq. (12.4) on 15 hypothetical regions, over ten years, thus testing the following equation, with the subscript i representing each region:

$$\lambda_i = \lambda_{0i} + \lambda_g g_i \tag{12.8}$$

For expositional ease, as will be shown below, they assume "Kaldor's stylized fact that the growth rate of the capital stock equals the growth rate of output (i.e., the capital-output ratio is constant). As a consequence of also assuming that factor shares are constant, this implies that the growth in the rate of profit is zero" (McCombie and Spreafico 2016, pp. 1127–1128). This means that in Eq. (12.3) of the national accounts, they assume for simplification that $g = \hat{K}$ (so that $\hat{k} = 0$) and $\hat{r} = 0$, so that by construction they have:

$$g(1-\pi) = (1-\pi)\hat{\omega} + (1-\pi)\hat{L} \tag{12.9}$$

Hence, the national accounts under the above restrictions become:

$$\hat{y} = g - \hat{L} = \hat{\omega} + 0.g \tag{12.10}$$

In this simplified case, on the basis of the national accounts, it is obvious that we ought to find no relationship whatsoever between the growth rate of output g and the growth rate of labour productivity $\hat{y}$.

What happens when regressions are run? McCombie and Spreafico construct a series of variables that give rise to g, $\hat{y}$ and $\hat{w}$ for their 15 hypothetical regions over ten years. In the first experiment, they assume by construction the existence of a Kaldor-Verdoorn effect, that is, they assume that there is a positive relationship between the growth rate of output and the growth rate of labour productivity. Running a regression based on Eq. (12.8), and assuming that the λ_{0i} parameter is allowed to vary for each region, the regression captures the special national accounting identity of Eq. (12.10), as the estimate of λ_{0i} is captured by the growth rate of real wages $\hat{\omega}$, while the estimate of λ_g is indeed statistically no different from zero. By contrast, when the λ_{0i} parameter is assumed to be the same for all regions, that is, when it is assumed that the exogenous constant of technical progress is the same for all regions, as is usually done in this kind of study, the Kaldor-Verdoorn effect is captured. Indeed the estimate of λ_g is statistically different from zero and around 0.45, which is close to the value with which the data was generated by construction.

McCombie and Spreafico (2016) conduct a second experiment, constructing the data in a similar manner, "with the exception that for any given productivity growth rates of a particular region, the output growth rates were random" (p. 1130). In other words, the Kaldor-Verdoorn effect is absent by construction. Once again, assuming that the λ_{0i} parameter is allowed to vary for each region, the regression has a near perfect fit (a R^2 close to unity) as it captures the national accounting identity of Eq. (12.10), with the estimate of λ_{0i} being captured by the growth rate of real wages $\hat{\omega}$, while the estimate of λ_g is not significant. By contrast, when the λ_{0i} parameter is assumed to be the same for all regions, the regression has a very poor fit, with the R^2 being close to zero, and the estimate of λ_g is not statistically different from zero, as it should be since the Kaldor-Verdoorn effect had been excluded by construction.

3.3 General Considerations

From these experiments, as well as from the arguments offered when comparing Eqs. (12.6) and (12.7), I believe it is safe to conclude that the Kaldor-Verdoorn effects are not an artefact. They do not arise from a specification that reproduces the national accounts identity. This is an important conclusion because the Kaldor-Verdoorn effects provide a possible explanation of the super-hysteresis effects that were empirically measured by León-Ledesma and Thirwall (2002) and that have been rediscovered by Blanchard et al. (2015). Super-hysteresis effects mean that a slowdown in the actual rate of growth of the economy, due, for instance, to a restrictive monetary policy, will have long-ranging effects, not only on the potential level of output but also on the growth rate of potential output. This can be explained, at least in part, by the

Kaldor-Verdoorn effect, as the slow growth in actual output is said to generate a slowdown in the growth rate of labour productivity. The effect can also occur on the upside, although neoclassical authors, seduced by downward hysteresis or super-hysteresis, seem dubious of a possible upward hysteresis effect.

As a conclusion on this section, it may be worth recalling that the Kaldor-Verdoorn effects became a hot topic during the recent primaries of the Democratic Party in the USA, when a controversy erupted between Gerald Friedman (2016a, b)—an economist from the University of Massachusetts in Amherst—and Christina and David Romer (2016), two economists who had held important positions in the federal administration. The controversy arose with regard to Friedman's estimates of the impact of the economic programme of the 2016 Democrat presidential candidate Bernie Sanders. To his dismay, Friedman discovered that the Kaldor-Verdoorn hysteresis effects that he had assumed in his estimates were not part of standard modelling, that is, the kind of models which are used by the Council of Economic Advisers. While Friedman may have overestimated the effects, at least the debate propelled the Kaldor-Verdoorn law into the sight of part of the layman public.

4 Methodological Considerations

4.1 Instrumentalism in Mainstream Economics

At the end of their book, Felipe and McCombie (2013, Chap. 12) wonder why their criticisms (and those of Shaikh, Simon and many others) of the aggregate production function have generally been ignored. In fact, McCombie's first foray in methodology was his 1998 paper on 'paradigms, rhetoric and the relevance of the aggregate production function', where he was already asking a similar question. Why are these demonstrations, and in particular the 'reductio ad absurdum' proofs, unable to convince neoclassical economists? One answer, offered by Solow, as recalled by Felipe and McCombie (2013, Chap. 5 and pp. 320–321), is that 'we knew it all beforehand'. This is dubious, because Solow (1957) in his own early work marvelled 'that the fit is remarkably tight'. But even if we grant the benefit of the doubt, this kind of answer contradicts the fact that neoclassical economists keep using standard aggregate production functions even today. It must be quite frustrating to realize that a quite straightforward argument just does not seem to have any impact on the profession when the consequences of the argument are so profound for mainstream theory.

I had my own experience when I presented a paper devoted to this issue in the economics department of the University of Ottawa (Lavoie 2008). I followed this up by sending by email two Shaikh and McCombie papers to those of my colleagues that I considered to be most open to dialogue. My colleagues listened politely during the presentation, and there was no contestation. Amazingly, the overall response was an apparent inability to understand the implications of the presentation and of the papers that I had sent. I found that a confusing feature for neoclassical economists is that their theory predicts that with perfect competition and factors paid at the value of their marginal product, the output elasticities will equate the factor shares. This is what the regressions yield when technical progress is properly taken into account. Mainstream economists don't get the point that, because of the identity, the estimate of the output elasticities will always turn out to be equal to factor shares.

The most genuine answer came from a member of the department that had been involved in the government and in advising developing countries: he told me that without the estimates of the output elasticities of the factors of production, there was nothing that he could advise about any more. As a consequence, he had to rely on the elasticity estimates derived from the regressions over deflated values, whatever their true significance. In other words, as Paul Davidson (1984) once put it when describing mainstream economics, he would prefer "to be precisely wrong rather than roughly right" (p. 572).

The reactions of neoclassical or mainstream economists to the findings regarding the tight links between the aggregate neoclassical production function and the national accounts identity, when these links are known, are thus strongly reminiscent of the instrumentalist position held by the majority of these economists. This is pointed out by Felipe and McCombie (2013, p. 314), as they recall that instrumentalism in economics is usually attributed to the (only?) methodological essay of Milton Friedman

(1953), according to whom the realism of assumptions is totally irrelevant and can even be a drawback. Robert Lucas (1981) has no doubt pursued instrumentalism to the hilt, when he claimed that “insistence on the ‘realism’ of an economic model subverts its potential usefulness in thinking about reality”, adding that good models had to “necessarily be artificial, abstract, patently unreal” (p. 270).

Indeed, McCombie and Negru (2014) remark that the New Consensus model, also called the New Neoclassical Synthesis, is based on an instrumentalist approach. They very correctly point out that “the criterion of success is the successful empirical implementation through calibration, rather than econometric testing” (p. 61). They add that “the accuracy of the assumptions, *per se*, is irrelevant. Primacy is given to the construction of artificial models that closely mimic the observed path of the economy (Lucas 1978). Indeed, at times, it seems as if econometric testing is irrelevant. What matters is that there should be a fully-articulated model, based on paradigmatic pseudo-assumptions, that has been shown to be capable of replicating the path of the economy” (*ibid*, p. 62).

While this critique may seem rather harsh, Paul Romer (2016), the new chief economist at the World Bank, has addressed an even more ruthless critique to New Classical economists and their dynamic stochastic general equilibrium (DSGE) models based on real business cycle theory. He complained that these authors were calibrating their models so that they could fit a number of stylized facts, without ever being able, however, to demonstrate that the assumed mechanisms—imaginary shocks to technology or utility functions—had any relationship with reality. Indeed, with the large number of parameters of these DSGE models, with enough patience, it is nearly always possible to provide a fair fit. As a result, Romer (2016) concludes that

> in the last three decades, the methods and conclusions ofmacroeconomics have deteriorated to the point that much of the work in this area no longer qualifies as scientific research. The treatment of identification in macroeconomic models is no more credible than in the first generation large Keynesian models, and is worse because it is far more opaque….The larger concern is that macroeconomic pseudoscience is undermining the norms of science throughout economics (p. 1).

While Romer focuses his attention to the likes of Lucas, Prescott and Sargent, his critique extends to the New Keynesian version of the New Consensus, as he also questions the way sticky prices are introduced into DSGE models. This is obvious if one recognizes, as do McCombie and Negru (2014), that "the difference between the New Keynesians (but not the post-Keynesians) and the New Classical economists are now a matter of degree, rather than of a fundamental nature" (p. 60). The benchmark model in the New Consensus is the real business cycle model of the New Classical economists; the New Keynesians add lots of rigidities and frictions to this benchmark model, but the logic of their revised DSGE model is no different from that of the benchmark model. New Keynesians integrate some degree of realisticness through the incorporation of auxiliary hypotheses—asymmetric information, credit rationing, liquidity-constrained households and sticky prices.

The main assumptions, however, based on an all-knowledgeable agent, attempting to maximize some utility function for eternity, defy common sense. The question, then, is whether it is possible to arrive at a model that describes the real world adequately by adding auxiliary realistic characteristics. Nicholas Kaldor (1966), for one, thought it was not possible: in an attempt to relieve the programme of its unrealistic foundations, the whole edifice would crumble. As he put it, removing the scaffolding "is sufficient to cause the whole structure to collapse like a pack of cards" (p. 310). Indeed, Kaldor (1972) used the same argument six years later, saying that "the scaffolding gets thicker and more impenetrable with every successive reformulation of the theory, with a growing uncertainty as to whether there is a solid building underneath" (p. 1239). There is no doubt that the scaffolding has taken gigantic dimensions with the advent of Lucasian economics as well as that of the New Consensus and its DSGE models, which are at the heart of mainstream macroeconomics.

Coming back to the issue of why the critique of the aggregate production function does not seem to have made a dent in the armour of neoclassical macroeconomics, McCombie and Pike (2013, p. 503) recall that econometric results rarely did have an impact on the beliefs of the profession. They give as an example the damning econometric critique by Hendry and Ericsson (1991) of the claims made by Friedman and the monetarists, which came out in 1985 but only got accepted for publication in 1991: most likely it was not the cause of the downfall of monetarism. McCombie and Negru (2014, p. 60) also mention the empirical proofs, too numerous to be recorded, showing that the uncovered interest parity equation just does not hold in the real world, have had no effect on the theoretical models constructed by both orthodox and heterodox economists. In this regard, the Sonnenschein-Mantel-Debreu theorem which questions from within the principle of scarcity and the Walrasian general equilibrium model has had no impact either, as pointed out again by McCombie and Negru (2014, p. 61) and as I have myself called attention to in Lavoie (1992, pp. 36–41) and Lavoie (2014, pp. 50–53).

4.2 Meta-Regression Analysis

An interesting counter-example is the work of Card and Krueger (1995), who provoked an economic earthquake when they contended, based on their own work and a rudimentary meta-regression analysis that raising the minimum wage had virtually no negative effect on employment and that previous research was flawed by publication bias. This counter-example is interesting on several grounds. First, from the sociological standpoint, the authors were considered to be traitors to the (neoclassical) cause as they came from Ivy League universities and were rejecting what was considered until then as one of the best established facts of neoclassical theory, a fact which had found proud of place in most introductory textbooks. Second, surprisingly, despite their work having been subjected to intense criticisms, a number of US economists seem to be less convinced by the negative impact that an increase in the minimum wage is likely to have on the employment of youths.[5] The only explanation that I can find for this proposition is that the lack of a negative effect is only incompatible with the pure competition version of neoclassical theory; within a labour-market model based on the confrontation between a monopolist and a monopsonist, anything goes when real wages are raised. Third, the book of Card and Krueger seems to have given a boost to meta-regression analysis in economics. This type of empirical analysis has found room in many different journals and particularly in the *Journal of Economic Surveys*.

I have tried to provide an introduction to meta-regression analysis in my book on post-Keynesian economics as I believe that meta-regression analysis offers an additional tool to overcome the publication bias so common in economics and to dismiss the belief that empirical studies provide support for most if not all of the standard claims of neoclassical textbooks (Lavoie 2014, pp. 64–70). Thus, in a sense, meta-regression analysis is a companion to the works of Shaikh as well as those of Felipe and McCombie, which have dismissed the bogus empirical support for neoclassical production functions. It helps to provide an antidote to the claim that there is no alternative (TINA) and it helps to convince students that there is room for alternatives in economic theory.

At the heart of the identification of publication bias is the notion that investigators who rely on smaller samples, with fewer degrees of freedom, are prone to larger standard errors. This implies that estimates of a parameter are likely to be less precise. In order to obtain statistically significant effects (say t ratios above 1.6), they will need to find large effects since the t statistic is the size of the coefficient divided by the standard error. This may require several tries, with different specifications. By contrast, with large samples, estimates are likely to be more precise, standard errors will be smaller, and hence a statistically significant result can be achieved despite smaller values of the estimated parameter. Thus, an adept of meta-regression analysis needs two things from each past regression: the size of the estimated parameter e, usually some elasticity measure, and a proxy of the precision of the estimate, ideally measured by the inverse of the standard error (SE). The meta-regression will thus be the following:

$$e_i = \beta_1 + \beta_0 SE_i + \varepsilon_i \tag{12.11}$$

β_1 represents the estimated *true* value of the parameter, for if the standard error SE is zero, then the estimate e will be equal to β_1.[6] We can then proceed to standard tests and check whether the null hypothesis $H_0 : \beta_1 = 0$ can be rejected or not. A fancier meta-regression analysis, based on a multi-variate approach, can also be pursued. In the case of research on the effect of the minimum wage, Doucouliagos and Stanley

(2009) find that the true effect is −0.009 and that it is statistically significant (the t ratio is 3.15). The true effect however is not economically significant: a 50 per cent increase in the real wage would lead to less than a ½ per cent decrease in teenage employment. If one were to take the average elasticity of the 1474 regressions analysed on this topic, the effect would be 20 times bigger! This shows the importance to rely on meta-regression analysis, as Doucouliagos and Stanley (2009) also find evidence of publication bias in studies devoted to the minimum wage.[7]

McCombie and Negru (2014, p. 62) note that it is not easy for neoclassical economists to reject the natural rate of unemployment hypothesis (or for that matter the NAIRU hypothesis), even when the evidence seems to be unfavourable to the hypothesis. Ray Fair (2012) has long been arguing that "the dynamics behind NAIRU equations are not supported by the data" (p. 9). Meta-regression analysis also concludes that there is no support for the concept of the natural rate of unemployment or the NAIRU. Tom Stanley has done two meta-regression analysis studies that pertain to the natural rate of unemployment story. In Stanley (2004), he looks at the persistence coefficient and finds that the true value appears to be very close to unity, thus implying that one cannot reject the hypothesis of unemployment hysteresis: the natural rate follows the actual rate of unemployment. This result is in line with the more recent work on hysteresis conducted in the conventional manner by Blanchard et al. (2015). In Stanley (2005), the other side of the NAIRU hypothesis is being explored: he looks at the relation between inflation and unemployment. He concludes from it and from his previous study that "the natural rate hypothesis may now be regarded as empirically falsified" (Stanley 2005, p. 626).

Despite all this, Fuller and Geide-Stevenson (2014, p. 135) report that the percentage of US economists that approve or approve with provisos the statement that "there is a natural rate of unemployment to which the economy tends in the long run" has not changed between 1990 and 2010: that percentage remains around 75 per cent. There is indeed resistance, except among heterodox economists and perhaps orthodox dissenters, to the dismissal of the natural rate hypothesis or that of the NAIRU concept, despite all their empirical failings, just as there is resistance to the dismissal of the neoclassical aggregate production function.

There are many more neoclassical key constructs, which have recently been questioned by the results of meta-regression analyses. This is particularly the case in microeconomics, which is usually considered to be the *forte* of neoclassical economics. Doucouliagos and Stanley (2013) look at 87 areas of economic research. They conclude that approximately 60 per cent of these suffer from severe or substantial publication bias. In microeconomics, the price elasticities of demand for residential water, tobacco, beer, spirits and alcohol all suffer from substantial or severe selectivity problems. They also show that these elasticities are very much overestimated, all of them being much below unity, so that the strength of substitution effects, which is at the heart of orthodox economics, is much weaker than usually described. In particular, the true value of the elasticity of CEO pay relative to performance seems to be zero!

Does meta-regression analysis have to say anything on some of the cherished beliefs of heterodox or post-Keynesian economics? Krassoi-Peach and Stanley (2009) look at what they call the efficiency-wage hypothesis, that is, in our own terms, what I have called the Webb effect. They conclude that whereas "most previous studies report mixed or ambiguous support for the efficiency-wage hypothesis, we find clear and robust evidence of a positive efficiency-wage effect on production" (p. 267). Indeed, their true estimate of the Webb effect elasticity is around 0.30, a number which is similar to what is found by Storm and Naastepad (2012, p. 103).

And what about the Kaldor-Verdoorn effect? Ludwig List, a PhD student at the University of Paris 13, has just conducted a meta-regression analysis on this effect. On the basis of nearly 120 estimates, List (2017) finds the true value of the Kaldor-Verdoorn effect to be 0.42, with no evidence of publication bias. This finding is consistent with the estimates claimed by post-Keynesians: McCombie (2002, p. 106) argued for robust estimates between 0.30 and 0.60; Hein and Tarassow (2010, pp. 748–749) found estimates between 0.27 and 0.86; and Storm and Naastepad (2012, p. 103) arrived at estimates ranging between 0.39 and 0.47. We may thus conclude from this meta-regression analysis that the Kaldor-Verdoorn effect is genuine and within its usually estimated range.

5 Summary and Conclusions

John McCombie is part of the Cantabrigian school of economics, which has provided an alternative vision of what macroeconomics is all about. He is one of the few who has managed to find a position and keep alive this tradition within the confines of the University of Cambridge, inspired by his predecessors such as Nicholas Kaldor or Bob Rowthorn. He has maintained the use of econometrics in post-Keynesian economics, at a time when abstract considerations were in fashion, and he made several important contributions to the Kaldorian strand of post-Keynesian economics.

In this chapter, I have reappraised his contribution to the Cambridge capital controversies by underlining the main arguments justifying the claim that empirics cannot provide support to the aggregate neoclassical production functions. I have shown that the Kaldor-Verdoorn law is not subjected to this critique. And I have argued that meta-regression analysis can provide additional elements in support of post-Keynesian economics.

Notes

1. It also appeared in my book on the foundations of post-Keynesian economics (Lavoie 1992, pp. 33–36).
2. Shaikh (1980, p. 2005) shows that labour productivity is highly nonlinear, nearly sinusoid, and hence cannot be represented by a linear trend. Regressions of the neoclassical production functions in time series will provide bad or even absurd results when technical progress is mishandled (for instance, by assuming a linear trend). A possible remedy is to include the rate of capacity utilization as an additional variable in the regression. This was indeed my personal experience when a graduate student of mine ran production regressions on Canadian data and was getting desperate until he added the rate of utilization in his regressions.
3. They are also discussed in my 2014 book (Lavoie 2014, ch. 1).
4. The authors declined to respond to my critique.

5. Fuller and Geide-Stevenson (2014) report that when surveyed members of the American Economic Association were confronted with the statement "A minimum wage increases unemployment among the young and unskilled workers", in 1990, 62 per cent agreed, 20 per cent agreed with provisos and 18 per cent disagreed; in 2010, 40 per cent agreed, 34 per cent agreed with provisos and 25 per cent disagreed.
6. In reality, to correct for possible heteroskedasticity, meta-regressions are based on the following equation, with β_1 and β_0 keeping their previous meaning: $(ei/SEi) = ti = \beta_1(1/SEi) + \beta_0 + \varepsilon_2$.
7. The absolute value of the parameter β_0 is a measure of publication bias and the authors find that its t ratio is above 10.

References

Anyadike-Danes, M., & Godley, W. (1989). Real wages and employment: A sceptical view of some recent econometric work. *Manchester School, 57*(2), 172–187.

Blanchard, O., Cerutti, E., & Summers, L. (2015). *Inflation and activity: Two explorations and their monetary implications* (NBER working paper 21726). http://www.nber.org/papers/w21726.pdf

Bronfenbrenner, M. (1971). La théorie néo-classique de la répartition du revenu en macro-économie. In J. Marchal & B. Ducros (Eds.), *Le partage du revenu national*. Paris: Cujas.

Card, D., & Krueger, A. B. (1995). *Myth and measurement: The new economics of the minimum wage*. Princeton: Princeton University Press.

Carter, S. (2011). C.E. Ferguson and the neoclassical theory of capital: A matter of faith. *Review of Political Economy, 23*(3), 339–356.

Cotis, J. P., Renaud, M., & Sobczak, N. (1998). Le chômage d'équilibre en France. *Revue économique, 49*(3), 921–935.

Davidson, P. (1984). Reviving Keynes's revolution. *Journal of Post Keynesian Economics, 6*(4), 561–575.

Doucouliagos, H., & Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations, 47*(2), 406–428.

Doucouliagos, H., & Stanley, T. D. (2013). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys, 27*(2), 316–339.

Fair, R. (2012). Has macro progressed? *Journal of Macroeconomics, 34*(1), 2–10.
Felipe, J., & McCombie, J. S. L. (2001). The CES production function, the accounting identity, and Occam's razor. *Applied Economics, 33*(10), 1221–1232.
Felipe, J., & McCombie, J. S. L. (2007). Is a theory of total factor productivity really needed? *Metroeconomica, 58*(1), 195–229.
Felipe, J., & McCombie, J. S. L. (2009). Are estimates of labour demand functions mere statistical artefacts? *International Review of Applied Economics, 23*(2), 147–168.
Felipe, J., & McCombie, J. S. L. (2011–12). On Herbert Simon's criticisms of the Cobb-Douglas and the CES production functions. *Journal of Post Keynesian Economics*, *34*(2), 275–293.
Felipe, J., & McCombie, J. S. L. (2013). *Aggregate production function and the measurement of technical change: Not even wrong*. Cheltenham: Edward Elgar.
Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 153–184). Chicago: Chicago University Press.
Friedman, G. (2016a). *What would Sanders do? Estimating the economic impact of Sanders' programs*. http://www.dollarsandsense.org/What-would-Sanders-do-013016.pdf
Friedman, G. (2016b). *Response to the Romers*. http://dollarsandsense.org/Friedman-Response-to-the-Romers.pdf
Fuller, D., & Geide-Stevenson, D. (2014). Consensus among economists—An update. *Journal of Economic Education, 45*(2), 131–146.
Hein, E., & Lavoie, M. (2015). Interview with John McCombie: I think there's absolutely no way out for them: An aggregate production function does not make any sense at all! *European Journal of Economics and Economic Policies, 12*(1), 1–6.
Hein, E., & Tarassow, A. (2010). Distribution, aggregate demand and productivity growth: Theory and empirical results for six OECD countries based on a post-Kaleckian model. *Cambridge Journal of Economics, 34*(4), 727–754.
Hendry, D., & Ericsson, N. R. (1991). An econometric analysis of U.K. money demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz. *American Economic Review, 81*(1), 8–38.
Jorgenson, D. W. (1974). Investment and production: A review. In M. D. Intriligator & D. A. Kendryck (Eds.), *Frontiers of quantitative economics* (Vol. 2). Amsterdam: North Holland.

Kaldor, N. (1957). A model of economic growth. *Economic Journal, 67*(268), 591–624.
Kaldor, N. (1966). Marginal productivity and the macro-economic theories of distribution. *Review of Economic Studies, 33*(4), 309–319.
Kaldor, N. (1972). The irrelevance of equilibrium economics. *Economic Journal, 82*(328), 1237–1252.
Krassoi-Peach, E., & Stanley, T. D. (2009). Efficiency wages, productivity and simultaneity: A meta-regression analysis. *Journal of Labor Research, 30*(3), 262–268.
Lavoie, M. (1987). *Macroéconomie: Théorie et controverses postkeynésiennes*. Paris: Dunod.
Lavoie, M. (1992). *Foundations of post-Keynesian economic analysis*. Aldershot: Edward Elgar.
Lavoie, M. (2000). Le chômage d'équilibre: réalité ou artefact statistique. *Revue Économique, 51*(6), 1477–1484.
Lavoie, M. (2003). A fully coherent post-Keynesian model of the euro zone. In P. Arestis, M. Baddeley, & J. McCombie (Eds.), *Globalization, regionalism and economic activity* (pp. 98–126). Cheltenham: Edward Elgar.
Lavoie, M. (2008). Neoclassical empirical evidence on employment and production laws as artefact. *Rivista Economía Informa, 351*, 9–36.
Lavoie, M. (2014). *Post-Keynesian economics: New foundations*. Cheltenham: Edward Elgar.
Layard, R., Nickell, S., & Jackman, R. (1991). *Unemployment: Economic performance and the labour market*. Oxford: Oxford University Press.
León-Ledesma, M. A., & Thirwall, A. P. (2002). The endogeneity of the natural rate of growth. *Cambridge Journal of Economics, 26*(4), 441–459.
List, L. (2017). *Does output influence productivity? – A meta-regression analysis* (Working paper). University of Paris 13.
Lucas, R. E. (1978). Unemployment policy. *American Economic Review, 68*(2), 353–357.
Lucas, R. E. (1981). *Studies in business cycle theory*. Cambridge, MA: MIT Press.
McCombie, J. S. L. (1987). Does the aggregate production function imply anything about the laws of production? A note on the Simon and Shaikh critiques. *Applied Economics, 19*(8), 1121–1136.
McCombie, J. S. L. (2001). What does the aggregate function show? Further thoughts on Solow's second thoughts on growth theory. *Journal of Post Keynesian Economics, 23*(4), 589–616.

McCombie, J. (2002). Increasing returns and the Verdoorn law from a Kaldorian perspective. In J. McCombie, M. Pugno, & B. Soro (Eds.), *Productivity growth and economic performance: Essays on Verdoorn's law* (pp. 64–114). Basingstoke: Palgrave Macmillan.
McCombie, J. (2015). Book review: Lavoie, M., post-Keynesian economics: New foundations. *European Journal of Economics and Economic Policies, 12*(2), 243–248.
McCombie, J. S. L., & Dixon, R. (1991). Estimating technical change in aggregate production functions: A critique. *International Review of Applied Economics, 5*(1), 24–46.
McCombie, J. S. L., & Negru, I. (2014). On economic paradigms, rhetoric and the micro-foundations of macroeconomics. *European Journal of Economics and Economic Policies, 11*(1), 53–66.
McCombie, J. S. L., & Pike, M. (2013). No end to the consensus in macroeconomic theory? A methodological enquiry. *American Journal of Economics and Sociology, 72*(2), 497–528.
McCombie, J. S. L., & Spreafico, M. R. M. (2016). Kaldor's technical progress function and Verdoorn's law revisited. *Cambridge Journal of Economics, 40*(4), 1117–1136.
Michl, T. R. (1985). International comparisons of productivity growth: Verdoorn's law revisited. *Journal of Post Keynesian Economics, 7*(4), 474–492.
Romer, P. (2016). The trouble with macroeconomics. http://ccl.yale.edu/sites/default/files/files/The%20Trouble%20with%20Macroeconomics.pdf
Romer, C. D., & Romer, D. H. (2016). *Senator Sanders's proposed policies and economic growth*. https://ineteconomics.org/uploads/general/romer-and-romer-evaluation-of-friedman1.pdf
Sato, K. (1974). The neoclassical postulate and the technology frontier in capital theory. *Quarterly Journal of Economics, 88*(3), 353–384.
Shaikh, A. (1974). Laws of production and laws of algebra. The humbug production function. *Review of Economics and Statistics, 56*(1), 115–120.
Shaikh, A. (1980). Laws of production and laws of algebra: Humbug II. In J. Nell (Ed.), *Growth, profits, & property: Essays in the revival of political economy* (pp. 80–95). Cambridge: Cambridge University Press.
Shaikh, A. (1990). Humbug production function. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Capital theory: The New Palgrave* (pp. 191–194). London: Macmillan.
Shaikh, A. (2005). Non-linear dynamics and pseudo production functions. *Eastern Economic Journal, 31*(3), 347–366.

Simon, H. A. (1979). On parsimonious explanations of production relations. *Scandinavian Journal of Economics, 81*(4), 459–474.

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics, 39*(2), 312–320.

Stanley, T. D. (2004). Does unemployment hysteresis falsify the natural rate hypothesis? A meta-regression analysis. *Journal of Economic Surveys, 18*(4), 589–612.

Stanley, T. D. (2005). Integrating the empirical tests of the natural rate hypothesis: A meta-regression analysis. *Kyklos, 58*(4), 611–634.

Storm, S., & Naastepad, C. W. M. (2012). *Macroeconomics beyond the NAIRU*. Cambridge, MA: Harvard University Press.

13

Is the Balance of Payments Constrained Growth Rate Time-Varying? Exchange Rate Over Valuation, Policy-Induced Recessions, Deindustrialization, and Long Run Growth

Mark Setterfield and Selen Ozcelik

1 Introduction

At least since Singh (1977), macroeconomists have expressed concern with 'premature deindustrialization'—a decline in the manufacturing share of economic activity in advance of that associated with established secular trends towards tertiarization in capitalist economies.[1] Concern with premature deindustrialization in advanced capitalist economies has long since extended to the notion that such failure may not be endemic to the manufacturing sector itself, but may instead be policy induced. Persistent exchange rate over valuation, for example, may reduce the cost-competitiveness of the manufacturing sector in international trade, resulting in the sector's contraction (Blecker 2003; Hersh 2003; Palley

M. Setterfield (✉) • S. Ozcelik
Department of Economics, New School for Social Research,
New York, NY, USA

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8_13

2003, 2016).[2] Prolonged recessions, meanwhile, which may be policy induced, are understood to disproportionately affect the cyclically sensitive manufacturing sector, reducing the share of manufacturing in overall activity (see Rowthorn and Wells 1987; Uemura and Tahara 2015). Recessions may also retard induced productivity growth in manufacturing (via Verdoorn's law), reducing either the price or non-price competitiveness of manufactured tradables, and so inducing further shrinkage of the manufacturing sector.

This chapter integrates concern over premature deindustrialization into balance of payments constrained growth (BPCG) analysis, one of the major contemporary currents in Kaldorian growth theory, to which John McCombie has been such an important contributor for over 35 years. The basis for what follows is the multi-sector BPCG model of Araujo and Lima (2007). We build on this contribution to develop a theory of a time-varying balance of payments constrained growth (TV-BPCG) rate, in which the standard multi-sector BPCG rate is rendered time-varying by virtue of temporary but persistent currency over valuations and/or policy-induced recessions that retard the actual rate of growth in the short run. These developments also induce structural changes, specifically, premature deindustrialization, which affect the value of the long-run equilibrium BPCG rate.[3] Since the BPCG rate thus varies with the short-run growth rate (by virtue of the sensitivity of both to currency over valuations and/or policy-induced recessions), the TV-BPCG rate so-derived can be interpreted as *quasi path dependent*. The results of these theoretical extensions to the canonical multi-sector BPCG model are also shown to give rise to a new empirical agenda for BPCG analysis.

The remainder of the chapter is organized as follows. Section 2 establishes the link between exchange rate fluctuations, recessions, and premature deindustrialization. In Sect. 3 we develop an extended BPCG model that demonstrates how the events discussed in Sect. 2 can affect long-run growth. Section 4 reflects on the implications of the resulting TV-BPCG model, and Sect. 5 summarizes and concludes.

2 Currency Over Valuation, Recessions, Premature Deindustrialization, and BPCG Analysis

In the literature on Thirlwall's law, the role of relative prices and hence price competitiveness in determining the long-run BPCG rate is generally disregarded, since the price elasticities of imports and exports are assumed to be very small (so-called elasticity pessimism). On that account, variations in the exchange rate, the domestic price of foreign currency, are considered to have no effect on long-run growth (Thirlwall 1979). Nevertheless, if an economy is considered within a multi-sectoral framework, variations in the exchange rate may have an impact on the income elasticity of exports if exchange rate variations disproportionately affect the manufacturing and non-manufacturing sectors which, in turn, account for different shares of total tradable goods and services. In this way, variations in the exchange rate, together with any other causes of structural change that alter the relative shares in activity of the manufacturing and non-manufacturing sectors, may have long run consequences for domestic macroeconomic performance. Put differently, deindustrialization and its causes appear as important phenomena to investigate in BPCG analysis.

Deindustrialization is commonly defined as a decline in the relative share of employment in the manufacturing sector.[4] There are known to be different paths towards deindustrialization, however. The most common path is positive deindustrialization due to the maturity effect (Rowthorn and Wells 1987). This occurs when advanced countries experience a decline in the relative share of manufacturing employment as a consequence of high productivity growth in the manufacturing sector, so that they no longer need to maintain the manufacturing share of employment to keep up the same share of manufacturing in total output. In this case, unemployment need not occur because job opportunities in slow productivity growth (and therefore labour absorbing) non-manufacturing industries allow workers to shift between sectors. This type of deindustrialization is symptomatic of economic success.[5]

Nonetheless, countries may experience premature deindustrialization because of currency over valuations and/or policy-induced recessions. Recessions that cause a reduction in investment spending and productivity growth in the manufacturing sector or persistent currency over valuations that diminish the competitiveness of manufacturing industries can reduce the manufacturing share of output and employment and also harm aggregate economic performance in the short run. But the detrimental effects of such events may not remain limited to the short run. Even temporary variations in domestic demand and/or the exchange rate may have lasting consequences for domestic macroeconomic performance via their impact on the manufacturing sector.

According to Kaldor's first law, the manufacturing sector is the engine of growth performance in an economy, since this sector has higher productivity growth than the non-manufacturing sector (Kaldor 1960).[6] Higher productivity growth ensures higher non-price competitiveness and hence a higher income elasticity of demand for exports, a parameter that is central to the determination of the BPCG rate. Moreover, through spillover effects and complementarity, the success of the manufacturing sector feeds into improved non-manufacturing sector performance in terms of employment and productivity. When the share of the manufacturing sector declines as a result of exchange rate appreciations and/or policy-induced recessions, the result is premature or negative deindustrialization that is not caused by the maturity effect and economic success but by economic (including policy) failure.

As experienced in the UK during the 1970s and in the USA during the 1990s, recessionary conditions and/or currency over valuation may result in the loss of manufacturing capacity (as opposed to just a fall in the capacity utilization rate) due, in the first instance, to declining domestic and international demand. This process may be self-perpetuating due to Verdoorn's law, which describes productivity growth as depending on the manufacturing growth rate and highlights the important influence of demand on supply conditions. As productivity growth slows, a country's manufacturing sector loses both its price and non-price competitiveness compared with its rivals in international trade inducing a further worsening of macroeconomic performance. Once the manufacturing sector is damaged, moreover, restoring the sector to health is far from a simple task. Blecker (2003) summarizes the cost of dollar over valuation beginning in the late 1990s for the US manufacturing sector. According to Blecker (op. cit.), increasing the value of the dollar relative to the euro and other major currencies of countries that have strong trade relations with the USA (such as Japan and China) affected domestic manufacturing sector through employment, profit, and investment spending channels. When the US dollar appreciated, export growth slowed resulting in a negative net effect on employment, with 740,000 jobs lost in the manufacturing sector by 2002. Moreover, loss of trade resulted in lower profit for manufacturing firms, reducing their willingness to invest and depriving them of the necessary financial resources to fund planned investment spending. As a result, the US manufacturing sector lost $100 billion in annual profits and experienced a decline of over $40 billion annually in investment spending (corresponding to 25% of total US manufacturing investment) through 2002 (Blecker 2003). Meanwhile, according to Rowthorn and Wells (1987), when Britain went through recession during the early 1970s, followed by the sudden sterling appreciation after 1977, and then experienced the Thatcher government's deflationary policies designed to control inflation following the second oil crisis, the British manufacturing sector was severely harmed. Even though the combination of recessionary conditions and over valuation came to an end after 1982, complete recovery of the manufacturing sector could not be achieved (Rowthorn and Wells 1987, pp. 136–137).

3 The Model

In this section, we use the insights discussed in the previous section to develop an extended BPCG model that shows how temporary but persistent recessions and/or exchange rate over valuations can induce premature deindustrialization and so reduce the long-run equilibrium BPCG rate.

3.1 A Multi-sector BPCG Model

The canonical BPCG model is a demand-led, one-sector aggregate structural model based on a single (tradable) commodity. Kaldorian growth theory, however, suggests that economic growth depends (in part) on the sectoral composition of output. This is because of the dynamic properties of the manufacturing sector, which is seen as the engine of growth.[7] It is not surprising to find, therefore, that multi-sector variants of BPCG theory have emerged that, while continuing to emphasize the importance of demand, also emphasize economic structure as a fundamental driver of growth (Nell 2003; Araujo and Lima 2007; Razmi 2011; Nishi 2016).

The essence of the multi-sector approach to BPCG theory can be represented using a simplified version of the Araujo and Lima (2007) model as follows. Consider an economy that imports a single commodity (denoted M) but exports a mixture of manufacturing (mf) and non-manufacturing (nmf) goods in proportion to its domestic industrial structure. The total value of exports can be described as:

$$PX = P_{mf}X_{mf} + P_{nmf}X_{nmf} \tag{13.1}$$

where P and X can be thought of as composite indices of the domestic price and volume of exported goods, respectively; P_{mf} and P_{nmf} are the domestic prices of manufactures and non-manufactures, respectively; and X_{mf} and X_{nmf} denote the volume of manufactured and non-manufactured exports, respectively. This expression can be re-written as:

$$X = \frac{P_{mf}}{P}X_{mf} + \frac{P_{nmf}}{P}X_{nmf} \tag{13.2}$$

Assuming that the relative prices of exports remain constant in the long run,[8] it follows that:

$$x = \omega_{mf}x_{mf} + \left(1-\omega_{mf}\right)x_{nmf} \tag{13.3}$$

where lower case variables represent proportional rates of growth and ω_{mf} is the share of manufacturing activity in the domestic economy (and hence, by assumption, in total exports).

Suppose now that import and export demand can be written as:

$$M = \left(\frac{P_f E}{P} \right)^{\psi} Y^{\pi} \tag{13.4}$$

where P_f is the foreign price of the imported commodity, M is the volume of imports, E is the nominal exchange rate, and ψ and π are the price and income elasticities of demand for imports, respectively, and:

$$X_i = \left(\frac{P_i}{P_f E} \right)^{\phi_i} Z^{\rho_i}, i = mf, nmf \tag{13.5}$$

where ϕ_i and ρ_i are the price and income elasticities of exports, respectively. Assuming constancy of relative prices in a common currency, it follows that:

$$m = \pi y \tag{13.6}$$

and:

$$x_i = \rho_i z, i = mf, nmf \tag{13.7}$$

from which (by substituting Eqs. (13.7) into (13.3)) we arrive at:

$$x = \left[\omega_{mf} \rho_{mf} + \left(1 - \omega_{mf}\right) \rho_{nmf} \right] z \tag{13.8}$$

Finally, we can write the aggregate balance of payments constraint on growth as:

$$P_f M E = P X \tag{13.9}$$

Once again assuming constancy of relative prices in a common currency (in this case, a constant real exchange rate, so that $p_f + e - p = 0$) we get:

$$m = x \tag{13.10}$$

Substituting Eqs. (13.6) and (13.8) into Eq. (13.10) and solving for y, we arrive at:

$$y_B = \frac{\left[\omega_{mf}\rho_{mf} + \left(1 - \omega_{mf}\right)\rho_{nmf}\right]z}{\pi} \tag{13.11}$$

In Eq. (13.11), y_B is the long-run (balance-of-payments-constrained) equilibrium growth rate. Note that, from Eq. (13.11), it follows that:

$$\frac{dy_B}{d\omega_{mf}} = \frac{\left(\rho_{mf} - \rho_{nmf}\right)z}{\pi} > 0 \tag{13.12}$$

if, as is assumed here, $\rho_{mf} - \rho_{nmf} > 0$. This latter inequality implies that world income growth is disproportionately directed towards growth in demand for manufactures. Its impact on the sign of the derivative in Eq. (13.12) is in keeping with the Kaldorian disposition to regard the manufacturing sector as the engine of growth (Kaldor's first law). Hence according to Eq. (13.12), the greater the manufacturing content of domestic output (and hence exports), the higher the long run equilibrium BPCG rate y_B.

3.2 Policy-Induced Recessions, Exchange Rate Over Valuation, and Growth: A Baseline Model

Suppose initially that we take ω_{mf} as given, so that we can write:

$$y_B = \frac{\left[\bar{\omega}_{mf}\rho_{mf} + \left(1 - \bar{\omega}_{mf}\right)\rho_{nmf}\right]z}{\pi} = \bar{y}_B \tag{13.13}$$

Now consider an extension of the model developed in the previous sub-section, such that:

$$y = y_B + \delta\left(E - E^*\right) \tag{13.14}$$

$$E = E^* + \epsilon \tag{13.15}$$

$$\epsilon = \alpha\epsilon_{-1} + \eta \tag{13.16}$$

where $\eta \sim \left(0,\sigma_\eta^2\right), y$ is the short-run rate of growth, and E^* is a constant short-run value of the nominal exchange rate that, within a particular short period, is consistent with the constant real exchange rate used to derive Eq. (13.11).[9] Equation (13.14) states that the short-run rate of growth can depart from the long run equilibrium BPCG rate if the nominal exchange rate is over- or under-valued ($E \neq E^*$). Equations (13.15) and (13.16), meanwhile, describe such misalignments as arising from persistent shocks. Note that the analytical framework in Eqs. (13.13, 13.14, 13.15, and 13.16) permits reinterpretation of E^* as the value of policy-sensitive domestic components of aggregate demand that is consistent with the long-run equilibrium BPCG rate y_B, so that $E \neq E^*$ can also be thought of as representing policy-induced, domestic-demand-led booms or recessions that cause the short-run rate of growth to depart from its long-run equilibrium value. In this way, under the same analytical conditions ($\eta < 0 \Rightarrow \epsilon < 0 \Rightarrow E < E^*$), Eqs. (13.13, 13.14, 13.15, and 13.16) can be interpreted as representing the effects on short-run growth (relative to its long-run equilibrium value) of *either* a persistent currency over valuation *or* a persistent policy-induced recession.[10]

Combining Eqs. (13.13, 13.14, 13.15, and 13.16) yields:

$$y = \bar{y}_B + \delta\alpha^{t-1}\eta \tag{13.17}$$

Suppose now that with $y = \bar{y}_B$ initially, $\eta < 0$ in some period $t = 1$. The resulting trajectories of y (in accordance with Eq. (13.17)) and y_B (in accordance with Eq. (13.13)) are illustrated in Fig. 13.1. As might be expected, y departs from the constant $\bar{y}_B$ in accordance with the initial

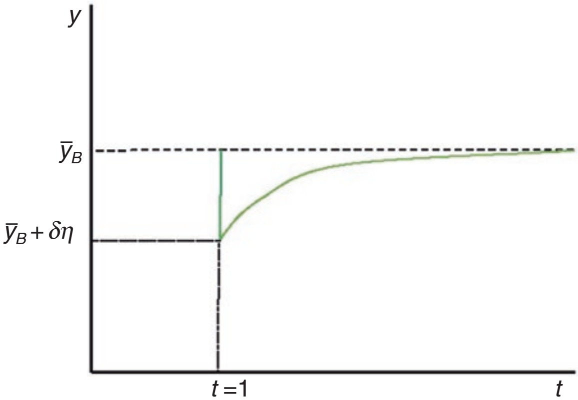


Fig. 13.1 Response of growth to a persistent shock in the baseline model

shock, subsequently converging back to $y = \bar{y}_B$ at a rate determined by α (which captures the degree of persistence in the initial shock). Although the short-run rate of growth is adversely affected (which adverse effect is increasing in the size of α), the long-run equilibrium rate of growth is unaffected.

3.3 Policy-Induced Recessions, Exchange Rate Over Valuation, and Growth: An Extended Model with Premature Deindustrialization

Now suppose that we retain Eqs. (13.14, 13.15, and 13.16), but relax the assumption of a constant ω_{mf} and, in keeping with the results in Eqs. (13.11) and (13.12), replace Eq. (13.13) with:

$$y_B = f(\omega_m), f' > 0 \tag{13.18}$$

and:

$$\dot{\omega}_m = \beta(E - E^*), \beta > 0 \tag{13.19}$$

Equation (13.18) is merely an implicit version of Eq. (13.11), while Eq. (13.19) endogenizes the manufacturing share of activity. Specifically, Eq. (13.19) suggests that the manufacturing share varies over time in accordance with $E \neq E^*$. With $\beta > 0$, this captures the possibility that either exchange over valuation or policy-induced recessions ($E < E^*$) can induce premature deindustrialization.

The effects of Eq. (13.19) on the behaviour of the system can be described as follows. First, note that it follows from Eq. (13.18) that $\dot{y}_B = f'\dot{\omega}_m$. Combining this last expression with Eq. (13.19), we arrive at:

$$\dot{y}_B = f'\beta\left(E - E^*\right) \tag{13.20}$$

Combining Eq. (13.20) with Eqs. (13.15) and (13.16), re-writing Eq. (13.17) to take into account time variation in the BPCG rate, and treating $\dot{x} \sim x_t - x_{t-1}$ for any x, we arrive at:

$$y_{Bt} = y_{Bt-1} + f'\beta\alpha^{t-1}\eta \tag{13.21}$$

$$y_t = y_{Bt} + \delta\alpha^{t-1}\eta \tag{13.22}$$

Now assume that $f'\beta < \delta$ – in other words, any $\eta < 0$ initially will cause a larger drop in y than in y_B. This is because while it is reasonable to assume that the actual rate of growth will respond immediately to exchange rate appreciation or domestic demand deflation, the sensitivity of the long-run equilibrium BPCG rate to these same conditions will take time to gradually become manifest, working as it does through channels that involve structural change in the economy.

Suppose now that $\eta < 0$ in some initial period $t = 1$. The resulting trajectories of y and y_B consistent with Eqs. (13.22) and (13.21), respectively, are illustrated in Fig. 13.2. Once again, Fig. 13.2 depicts the immediate (negative) impact of $E < E^*$ on y, following which the actual rate of growth begins a process of convergence back towards its long-run equilibrium value at a rate determined by the magnitude of the persis-

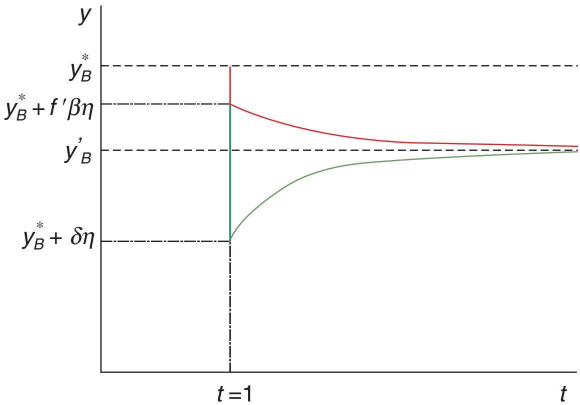


Fig. 13.2 Response of growth to a persistent shock in the extended model

tence of the initial shock (the size of α). This time, however, $E < E^*$ also causes an initial drop in the value of the long-run equilibrium BPCG rate itself, the latter continuing to decay as the shock persists. In other words, the persistence of the shock causes the long-run equilibrium growth rate to fall continuously, even as the actual rate of growth converges back towards its long-run equilibrium value. The traverse path so-described results (in the limit) in the economy settling into a new long run equilibrium BPCG rate at y'_B in Fig. 13.2. The sensitivity of the long-run growth rate to temporary (but persistent) exchange rate over valuation and/or policy-induced recessions is a product of the deindustrialization that these same conditions create. The result, as depicted in Fig. 13.2, is variability over time in the BPCG rate, so that the long-run equilibrium rate of growth may now be described as a time-varying balance of payments constrained growth (TV-BPCG) rate.

To solve for the new long run equilibrium TV-BPCG rate depicted in Fig. 13.2, first note that it follows from Eq. (13.21) that:

$$y_{Bt} = y_{B0} + f'\beta\eta \sum_{i=1}^{t} \alpha^{t-i}$$

Substituting this expression into Eq. (13.22) we arrive at:

$$y_t = y_{B0} + \delta\alpha^{t-1}\eta + f'\beta\eta\sum_{i=1}^{t}\alpha^{t-i} \tag{13.23}$$

Note that:

$$\lim_{t\to\infty}\delta\alpha^{t-1}\eta = 0$$

while:

$$\lim_{t\to\infty} f'\beta\eta\sum_{i=1}^{t}\alpha^{t-i} = \frac{1}{1-\alpha}f'\beta\eta < 0$$

Recalling that $y_{B0} = y_B^*$, it therefore follows from (13.23) that:

$$\lim_{t\to\infty} y_t = y_B^* + \frac{1}{1-\alpha}f'\beta\eta = y_B' \tag{13.24}$$

this last value being the new equilibrium TV-BPCG rate depicted in Fig. 13.2.

3.4 Simulations

Having developed the TV-BPCG model in the previous sub-section, it is useful to show exactly how the system reacts to variations in E (representing currency over valuations or policy-induced recessions) under different parameter restrictions. As can be inferred from Eq. (13.21), in addition to the initial value of the shock η and the degree of persistence of this shock (α), how the long-run BPCG rate varies over time depends on the size of the parameters β (the sensitivity of $\dot{\omega}_m$ to $E - E^*$, which we terms the 'Palley effect') and f' (the sensitivity of $\dot{y}_B$ to $\dot{\omega}_m$, which we term the ‘Kaldor effect’). The size of these parameters may differ from country to country, subject to their precise economic structures. Keeping in mind that the exact values we assign to parameters (and to the initial values of variables) are arbitrary (the purpose of the exercise here being to assess the effect of increases or decreases in the size of parameters relative to a benchmark case, in the manner of a comparative static exercise), we can simulate different TV-BPCGs by assigning different values to the two key parameters f' and β while keeping all other parameters constant. With this in mind, our basic parameter set (together with the initial values of the variables η, y_B, and y) is as follows:

$$y_{B0} = y_0 = 5.0;$$

$\delta = 0.1$;[11]
$\alpha = 0.8$; and finally
$\eta_0 \neq 0$, where the exact value of η is determined by a random draw from a normal distribution with $\mu_\eta = -10$ and $\sigma_\eta = 0.5$.

In the first (benchmark) simulation, we assign values to our key parameters capturing the Palley and Kaldor effects of $\beta = 0.2$ and $f' = 0.06$. The response of the system to a transitory shock $\eta_0 < 0$ ($\eta_i = 0 \quad \forall\, i \neq 0$) is depicted in Fig. 13.3.

In Fig. 13.3, the blue line corresponds to the trajectory of the TV-BPCG rate y_B, and the yellow line depicts the trajectory of the actual growth rate y.[12] As can be seen, a temporary but persistent shock to E results in a continuously falling long run BPCG rate. The short-run actual growth rate, meanwhile, which deteriorates markedly more than the TV-BPCG rate in the immediate aftermath of the shock, subsequently reverts towards the declining TV-BPCG rate, the two growth rates eventually converging towards a new, lower, long run equilibrium growth rate. This benchmark simulation replicates the system dynamics depicted in Fig. 13.2, where the short-term effects of the shock $\eta \neq 0$ on growth are larger than the long-term effects, but the latter are relatively large (the TV-BPCG falling from an initial value of 5.0% to about 4.4%).

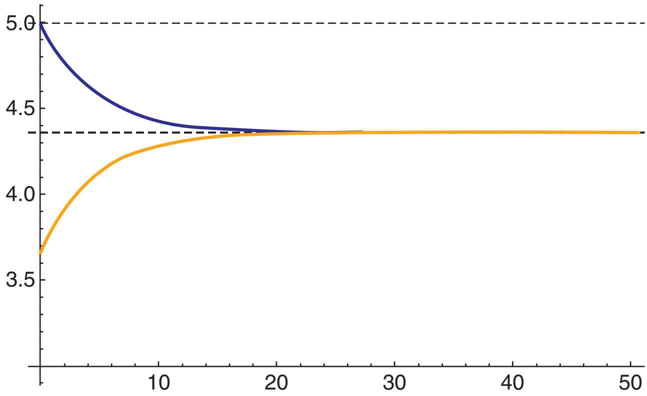


Fig. 13.3 The benchmark case

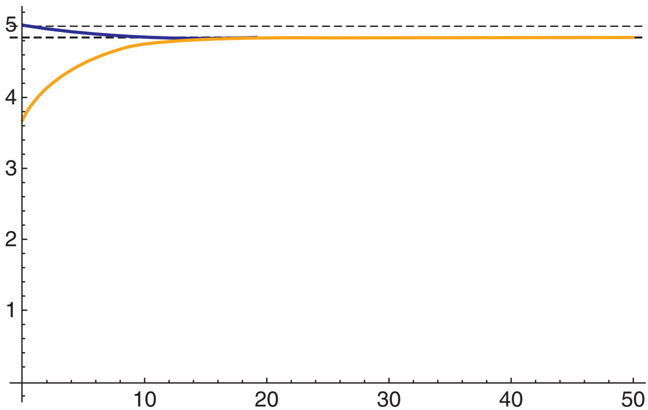


Fig. 13.4 A less shock-prone TV-BPCG rate

Our next simulation depicts a case involving smaller Palley and Kaldor effects, reflected in the smaller parameter values $\beta = 0.1$ and $f' = 0.03$. The results of this second simulation are depicted in Fig. 13.4.

Here, the system dynamics display a similar pattern to those observed in the benchmark case depicted in Fig. 13.3. But the impact on long run growth is much reduced, the new steady state being only marginally lower than its original 5% value. This is because neither the sensitivity of the manufacturing share to exchange rate appreciation/domestic demand deflation (the Palley effect) nor the sensitivity of the long run growth rate to the manufacturing share (the Kaldor effect) is large enough to create major changes in the TV-BPCG rate. This pattern may be observed in countries which have relatively higher income elasticities of demand for their manufacturing exports compared to their trade rivals, as a result of firms' efforts to increase non-price competitiveness, so that a shock to the exchange rate that has detrimental effects on price competitiveness results only in slight declines in the manufacturing sector and hence the TV-BPCG rate. Moreover, as a result of positive de-industrialization and/or industrial policy, both the income elasticity of demand for non-manufacturing exports and the share of the non-manufacturing sector in exports may be high. At the same time, spillover effects from a robust manufacturing sector to the non-manufacturing sector may result in enhanced productivity performance in this sector (at least among what Baumol et al. (1991) term 'dynamic' service industries that are less prone to Baumol's disease), as a result of which certain non-manufacturing industries share (with the manufacturing sector) the role of 'engine of growth'. In other words, with smaller Palley and Kaldor effects (smaller values of β and f', respectively), the time variance of the BPCG rate may be relatively minor, and, as a consequence, the TV-BPCG rate will be less shock-prone.

In our final simulation, we depict an extreme case involving larger Palley and Kaldor effects captured by the parameter values $\beta = 0.3$ and $f' = 0.3$ (see Fig. 13.5). In this case, the actual growth rate and the BPCG rate eventually collapse to a new steady-state growth rate of less than 1%, far below the initial equilibrium growth rate of 5%.

As seen in Fig. 13.5, either one or both of the Palley and Kaldor effects can be so large as to be overwhelming, producing extreme sensitivity of y_B to the initial shock $\eta_0 \neq 0$. In this case, the initial drop in the short-term actual rate of growth y due to $\eta_0 \neq 0$ causes a subsequent drop in y_B that, due to the Palley and Kaldor effects, is so large that the properties of y_B as an attractor subsequently pull y down further below its initial depressed value. The end result is that both y and y_B continue to fall as the initial shock persists, eventually converging to a new steady-state value that lies below $y_0 + \eta_0$ (where $\eta_0 < 0$). In other words, the long-term

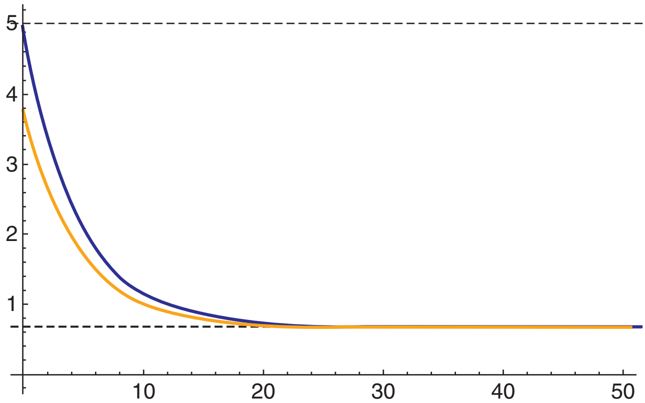


Fig. 13.5 The damaging consequences of large Palley and/or Kaldor effects

growth effects of a persistent demand shock are now larger than the short-term effects.

The simulation exercises above demonstrate the sensitivity of the TV-BPCG, in an economy subject to transitory but persistent demand shocks, to differing magnitudes of the Palley effect (the sensitivity of the manufacturing share to exchange rate over valuation and/or domestic recessions) and the Kaldor effect (the extent to which manufacturing is the engine of long run growth). In so doing, they draw attention to the way in which negative deindustrialization caused by recessions and/or exchange rate over valuation may impact economic structure in such a way as to lower the long-run equilibrium BPCG rate. Note that if $E \neq E^*$ due to the side-effects of policy (e.g., a high interest rate regime designed to curb inflation), then ultimately the outcomes reflected in the TV-BPCG rate are also policy side effects. These adverse consequences for growth are long-term and (ceteris paribus) permanent, but it cannot be argued that they result from policy lock in/out (as in Palley (2016)), since $E = E^*$ in the long run (in Palley's terms, policy can be, and eventually is, 'dialed back'). But economic structure is nevertheless affected with detrimental long-run macroeconomic performance effects, revealing that permanent damage can be wreaked by even temporary policy mistakes (such as the 'Thatcher experiment' with monetarism) that are not subject to policy lock in.

4 A Time-Varying and Quasi Path-Dependent Balance of Payments Constrained Growth Rate

The behaviour of y and y_B resulting from Eqs. (13.22) and (13.21) gives rise to a long-run equilibrium BPCG rate that is time-varying, as in Eq. (13.24). The obvious contrast is with orthodox BPCG theory, as summarized (in the case of a multi-sector economy) by Eq. (13.11). Here, the BPCG rate is understood as a single-valued steady state that, in the absence of exogenous parametric change, remains constant in the long run and unaltered by short-term adjustment dynamics that lead the actual rate of growth back towards the BPCG rate in the event of a transitory disequilibrating shock (see, e.g., McCombie and Thirlwall 1994; McCombie 2011; Setterfield 2011; Thirlwall 2011).

In fact, with $y \neq y_B$ in the short-term now resulting in permanent change in y_B in the long run, Eqs. (13.22) and (13.21) resemble a path-dependent system wherein any (on the face of it, temporary) change in the actual rate of growth results in change in the long-run equilibrium rate of growth.[13] This appearance of path dependence should not be exaggerated, however. As the analysis in Sect. 3.3 makes clear, *a common third factor*—the sensitivity of both y and y_B to $E \neq E^*$—is driving the behaviour of the system as a whole. Nevertheless, it might be said that y_B is not only time-varying, but also *quasi* path dependent.

Both the theoretical development of a TV-BPCG rate and its identification as quasi path dependent are potentially exploitable in an empirical context. Consider first the existence of a TV-BPCG rate. If we think of y_B as the low-frequency trend in an economy's real output growth time series data, then any evidence of the non-constancy of this trend would substantiate the hypothesis that there exists a TV-BPCG rate. For instance, rejection of the null hypothesis that the trend in an economy's real output growth time series data is I(0) would provide *prima facie* evidence of the existence of a TV-BPCG rate.

Consider now the value to BPCG empirics of identifying the TV-BPCG rate as quasi path dependent. Having already interpreted y_B as the low-frequency trend in an economy's real output growth time series data, it is reasonable to think of the accompanying high-frequency cyclical component of the same data as representing $y \neq y_B$ as a result of $E \neq E^*$. The quasi path dependence of the TV-BPCG rate now suggests that there is merit to studying the sensitivity of the trend to the cyclical component of real output growth data. The methodology for such analysis has already been established in the NAIRU literature and has since migrated to Post-Keynesian macrodynamics (Schoder 2012).

In sum, while this chapter has focused on developing the theory of a TV-BPCG rate, the analysis presented has clear empirical extensions. The TV-BPCG rate (interpreted as the trend component of time-series growth rate data) can be expected to vary with the short-run growth rate (the cyclical component of time-series growth rate data), as a result of its quasi path dependence. This is suggestive of a new empirical agenda in BPCG analysis that focuses on studying the time variation in, and quasi path dependence of, the BPCG rate.

5 Summary and Conclusions

Multi-sector BPCG models emphasize the importance of economic structure (as well as demand) for long run growth, in keeping with Kaldor's first law. Analysis of structural change, meanwhile, suggests that premature deindustrialization can be induced by persistent currency over valuations and/or recessions—either of which may be a by-product of macro policy. This chapter combines these insights to develop a model of time-varying balance of payments constrained growth (TV-BPCG). The TV-BPCG rate is shown to be quasi path dependent, suggesting a new empirical agenda for BPCG analysis associated with trend-cycle interactions in time series growth data. It is hoped that these insights will contribute to the further advancement of BPCG theory and Kaldorian growth theory writ large and that in so doing, they pay fitting tribute to the foundational work of John McCombie in this field.

Notes

1. See, for example, Rowthorn and Wells (1987) on the so-called maturity effect, a process of deindustrialization that appears to be an inevitable accompaniment of the growth trajectory of advanced capitalist economies.
2. This is sometimes interpreted as part of a larger set of concerns regarding 'finance versus industry' in modern capitalism (Epstein and Schor 1990; Epstein 2001).
3. Overall, then, our basic approach is not unlike that of Missio et al. (2017), who also build on the multi-sector BPCG model of Araujo and Lima (2007) to show that an exchange rate devaluation can affect the BPCG rate by (inter alia) increasing the manufacturing share of economic activity. Unlike Missio et al. (2017), however, who consider the comparative static effects of the exchange rate on the equilibrium rate of growth, the model developed here captures the permanent effects (on the BPCG rate) of a transitory exchange rate appreciation. It also shows how an explicitly *transitory* shock to domestic demand can be interpreted as having the same (long run equilibrium growth retarding) effect.
4. The manufacturing share of total output is also sometimes used to gauge deindustrialization. See, for example, Felipe and Mehta (2016) for a recent discussion of international trends in the manufacturing share of total output.
5. It may not be altogether benign, however, since stagnant service-sector productivity growth may contribute to stagnant real wage growth in this sector, and hence rising income inequality.
6. See Thirlwall (1983) on Kaldor's laws and Cantore et al. (2017) for a review of, and further contribution to, recent empirical evidence affirming the importance of Kaldor's first law.
7. See, for example, Thirlwall (2013) for a recent discussion of these basic tenets of Kaldorian growth theory in the context of BPCG theory, and in particular, the importance of both demand and economic structure in the determination of long-run growth.
8. This assumption is a counterpart to the assumption of a constant real exchange rate from which the canonical one-sector version of Thirlwall's law can be derived.
9. Long-run constancy of the real exchange rate does not require long-run constancy of the nominal exchange rate, but we assume here that, for

discrete intervals of time, the (own currency) prices of tradable goods are fixed so that there exists, in the short run, a unique constant value of E consistent with long-run constancy of the real exchange rate.

10. Recall that as defined in Eqs. (13.4) and (13.5), E is the domestic price of foreign currency, so that a *fall* in the value of E such that $E < E^*$ represents a nominal exchange rate *appreciation*.
11. Recall that δ captures the sensitivity of y to $E - E^* \neq 0$.
12. The same conventions are maintained throughout the simulations that follow.
13. Indeed, the results derived from equations (22) and (21) might even be described as hysteresis-like, in the sense that they describe a strictly transitory shock as ultimately having permanent effects on the growth path of the economy. We refer to a "hysteresis-like" result here by virtue of its having one of the salient properties of hysteresis—namely, the propensity of a temporary cause to have a permanent effect. Note, however, that it does not arise from a process of "strong" or "true" hysteresis. See Setterfield (2009) for further discussion.

References

Araujo, R. A., & Lima, G. T. (2007). A structural economic dynamics approach to balance-of-payments-constrained growth. *Cambridge Journal of Economics, 31*(5), 755–774.

Baumol, W., Blackman, S., & Wolff, E. (1991). *Productivity and American leadership: The long view*. Cambridge, MA: MIT Press.

Blecker, R. (2003). *The benefits of a lower dollar: How the high dollar has hurt U.S. manufacturing producers and why the dollar still needs to fall further* (Briefing paper). Washington, DC: Economic Policy Institute.

Cantore, N., Clara, M., Lavopa, A., & Soare, C. (2017). Manufacturing as an engine of growth: Which is the best fuel? *Structural Change and Economic Dynamics, 42*, 56–66.

Epstein, G. (2001). *Financialization, rentier interests, and central bank policy* (Technical report). Amherst: University of Massachusetts.

Epstein, G., & Schor, J. (1990). Macropolicy in the rise and fall of the golden age. In S. A. Marglin & J. B. Schor (Eds.), *The golden age of capitalism: Reinterpreting the postwar experience*. Oxford: Clarendon Press.

Felipe, J., & Mehta, A. (2016). Deindustrialization? A global perspective. *Economics Letters, 149*, 148–151.

Hersh, A. S. (2003). Is there a future for manufacturing? The visible hand of U.S. deindustrialization. In A. E. Eaton (Ed.), *Industrial relations research association: Proceedings of the 55th annual meeting*. Champaign: Industrial Relations Research Association.

Kaldor, N. (1960). *Causes of growth and stagnation in the world economy*. Cambridge: Cambridge University Press.

McCombie, J. S. L. (2011). Criticisms and defences of the balance-of-payments constrained growth model: Some old, some new. *PSL Quarterly Review, 64*(259), 353–392.

McCombie, J. S. L., & Thirlwall, A. P. (1994). *Economic growth and the balance-of-payments constraint*. London: Macmillan.

Missio, F., Araujo, R. A., & Jayme, F. G. (2017). Endogenous elasticities and the impact of the real exchange rate on structural economic dynamics. *Structural Change and Economic Dynamics, 42*, 67–75.

Nell, K. (2003). A 'generalised' version of the balance-of-payments growth model: An application to neighbouring regions. *International Review of Applied Economics, 17*(3), 249–267.

Nishi, H. (2016). A multi-sectoral balance-of-payments-constrained growth model with sectoral heterogeneity. *Structural Change and Economic Dynamics, 39*, 31–45.

Palley, T. (2003). The over-valued dollar and the US slump. In C. F. Bergsten & J. Williamson (Eds.), *Dollar overvaluation and the world economy*. Washington, DC: Institute for International Economics.

Palley, T. (2016). *A theory of economic policy lock-in and lock-out via hysteresis: Rethinking economists' approach to economic policy* (Economics discussion papers 2016–50). Kiel, Germany: Kiel Institute for the World Economy (IfW).

Razmi, A. (2011). Exploring the robustness of the balance of payments-constrained growth idea in a multiple good framework. *Cambridge Journal of Economics, 35*(3), 545–567.

Rowthorn, B., & Wells, J. (1987). *De-industrialization and foreign trade*. Cambridge: Cambridge University Press.

Schoder, C. (2012). Hysteresis in the Kaleckian growth model: A Bayesian analysis for the US manufacturing sector from 1984 to 2007. *Metroeconomica, 63*(3), 542–568.

Setterfield, M. (2009). Path dependency, hysteresis and macrodynamics. In P. Arestis & M. Sawyer (Eds.), *Path dependency and macroeconomics*. London: Palgrave Macmillan.

Setterfield, M. (2011). The remarkable durability of Thirlwall's law. *PSL Quarterly Review, 64*(259), 393–427.
Singh, A. (1977). UK industry and the world economy: A case of de-industrialisation? *Cambridge Journal of Economics, 1*(2), 113–136.
Thirlwall, A. P. (1979). The balance of payments constraint as an explanation of international growth rate differences. *Banca Nazionale del Lavoro Quarterly Review, 32*(128), 45–53.
Thirlwall, A. P. (1983). A plain man's guide to Kaldor's growth laws. *Journal of Post Keynesian Economics, 5*(3), 345–358.
Thirlwall, A. P. (2011). Balance of payments constrained growth models: History and overview. *PSL Quarterly Review, 64*(259), 307–351.
Thirlwall, A. P. (2013). *Economic growth in an open developing economy: The role of structure and demand*. Cheltenham: Edward Elgar.
Uemura, H., & Tahara, S. (2015, June 10–12). *De-industrialization in Japan and international production linkages in East Asia*. Technical report, presented at the Colloque International Recherche & Régulation, Paris.

Index[1]

[1]Note: Page number followed by 'n' refer to notes.

P. Arestis (ed.), *Alternative Approaches in Macroeconomics*,
https://doi.org/10.1007/978-3-319-69676-8